**RESEARCH ARTICLE**

**Key Points:**
- High-resolution spatiotemporal precipitation reconstruction from NDVI data
- New data set provides accurate spatiotemporal information over data sparse regions
- New data set highlights spatial variability in precipitation over central Andes

**Correspondence to:**
C. Yarleque,
cyarleque@albany.edu

# Multiscale assessment of spatial precipitation variability over complex mountain terrain using a high-resolution spatiotemporal wavelet reconstruction method

Christian Yarleque[1], Mathias Vuille[1], Douglas R. Hardy[2], Adolfo Posadas[3,4], and Roberto Quiroz[3]

[1]Department of Atmospheric and Environmental Sciences, University at Albany, State University of New York, Albany, New York, USA, [2]Department of Geosciences, University of Massachusetts Amherst, Amherst, Massachusetts, USA, [3]Crop Systems Intensification and Climate Change Program, International Potato Center, Lima, Peru, [4]World Agroforestry Centre (ICRAF), Nairobi, Kenya

**Abstract** Studying precipitation variability in the Peruvian Andes is a challenge given the high topographic variability and the scarcity of weather stations. Yet previous research has shown that a near-linear relationship exists between precipitation and vegetation in the semiarid central Andes. We exploit this relationship by developing a new, spatially highly resolved spatiotemporal precipitation reconstruction method, using daily precipitation time series from in situ weather stations, and dekadal (10 calendar days) normalized difference vegetation index (NDVI) fields. The two data sets are combined through a wavelet decomposition method. A $4° \times 4°$ region around Quelccaya ice cap (QIC), the world's largest tropical ice cap located in the central Peruvian Andes, was selected as study area, due to its importance for climatic, glaciologic, and paleoclimatic research. The reconstructed end product, a ~1 km² gridded precipitation data set at dekadal temporal resolution, was validated against independent rain gauge data and compared with the Tropical Rainfall Measuring Mission (TRMM) 3B42 version 7 product. This validation showed a better overall performance of our own reconstruction than the TRMM data. Additionally, a comparison of our precipitation product with snowfall measurements at the QIC summit (5670 m) shows a regionally coherent signal at the dekadal scale, suggesting that the precipitation falling at QIC is driven by regional- rather than local-scale convective activity. We anticipate that this methodology and the type of data generated in this study will be useful for hydrological and glaciological studies, as well as for validation of high-resolution downscaling products in mountain regions.

## 1. Introduction

The Andes mountain chain, extending along western South America from Venezuela to Chile, affects climate and the atmospheric circulation of the entire region [*Garreaud*, 2009], primarily in lower tropospheric levels. Modeling studies show that the Andes affect first and foremost the tropical precipitation distribution in the interior of the continent [*Lenters and Cook*, 1999] and the Intertropical Convergence Zone (ITCZ) north-south asymmetry [*Takahashi and Battisti*, 2007]. The Andes are home to a large number of tropical glaciers and ice caps, which are sensitive to the current climate warming [*Schauwecker et al.*, 2014; *Vuille et al.*, 2015], negatively affecting their energy and mass balance [*Vuille et al.*, 2008a; *Rabatel et al.*, 2013]. The warming has led to rapid shrinkage of many tropical glaciers [*Vuille et al.*, 2008a], affecting the regional hydrology [e.g., *Baraer et al.*, 2012]. Atmospheric warming is considered to be the main factor explaining the current retreat observed on tropical glaciers and ice caps, with higher freezing levels increasingly exposing the ablation zones to rain as opposed to snow [*L'Hote et al.*, 2005; *Bradley et al.*, 2009; *Rabatel et al.*, 2013].

While the temperature increase in the Andes is well documented [*Vuille and Bradley*, 2000; *Vuille et al.*, 2015], the role that changes in precipitation amount or in the rain/snow ratio might play in negative glacier energy and mass balance is not nearly as well understood. Lack of a dense station network, the complex topography leading to sharp precipitation gradients over short distances and the logistical difficulties in obtaining accurate measurements of snow accumulation on tropical glaciers have prevented a thorough analysis of spatiotemporal precipitation patterns at high resolution. Here we focus on quantifying precipitation and its variability, at high spatial resolution, in a region of southern Peru, surrounding the world's largest tropical ice cap, Quelccaya [*Thompson et al.*, 2006, 2013; *Hardy and Hardy*, 2008; *Hanshaw and Bookhagen*, 2014]. An automated weather station has been recording continuous in situ climatic information at the summit of the

Quelccaya Ice Cap (QIC) for over a decade [*Bradley et al.*, 2009; *Hurley et al.*, 2015]. Thus, the QIC region is an ideal natural laboratory to evaluate tropical mountain atmosphere-cryosphere interactions.

On interannual time scales, the occurrence of the El Niño–Southern Oscillation (ENSO) phenomenon affects precipitation over the southern tropical Andes [*Vuille*, 1999; *Vuille et al.*, 2000; *Garreaud and Aceituno*, 2001; *Garreaud et al.*, 2003]. In general, the amount of precipitation increases (decreases) over the southern Peruvian Andes during La Niña (El Niño) [*Vuille et al.*, 2008b]. The exact mechanism by which ENSO affects precipitation over the QIC region is not very well studied, although the main large-scale forcing factor appears to be the meridional baroclinicity and the resulting change in the strength of the upper level wester- lies over the Andes region, induced by the anomalous sea surface temperature change over the eastern Pacific [*Garreaud and Aceituno*, 2001]. The upper level winds affect the strength of the upslope flux of humid air from the lowlands to the east. Studies on the ENSO footprint over the broader Altiplano region show, how- ever, that the eastern Andean region, where the QIC is located, is not as strongly affected by this forcing as the more arid western cordillera [*Vuille et al.*, 2000].

The upper level easterlies are established on the northern side of the Bolivian High [*Lenters and Cook*, 1999] as part of the South American monsoon system during austral summer and lead to a wet season on QIC that starts around October and November, and lasts until March or April [*Hurley et al.*, 2015]. The clear spatial pre- cipitation gradient across the southern Andes, ranging from the more humid northeast to the arid southwes- tern part of southern Peru, as measured, for example, by the Tropical Rainfall Measuring Mission (TRMM) products during the wet season [*Mohr et al.*, 2014], is a clear testament to an easterly moisture source. While the moisture is provided by evaporation over the tropical Atlantic and recycled over the Amazon basin, recent studies on the synoptic conditions associated with snowfall over Quelccaya highlight the importance of cold air incursions, embedded in extratropical Rossby wave activity. Snowfall on QIC appears to be related to uplift of moist air along the leading edge of these northward propagating cold air masses to the east of the Andes [*Hurley et al.*, 2015]. Hence, the moisture is of tropical origin and associated with deep convective storms [e.g., *Bookhagen and Strecker*, 2008; *Casimiro et al.*, 2009; *Lavado et al.*, 2012, 2013], but uplift and condensation are triggered by northward moving cold fronts. In fact, more than 70% of snowfall on Quelccaya is tied to the leading edge of northward propagating cold air incursions [*Hurley et al.*, 2015].

While this larger-scale forcing is fairly well established, the interaction of easterly flow and northward pene- trating cold frontal systems with the Andean orography is not well understood, due principally to the scarcity of observational data and the complex topography affecting the precipitation distribution around QIC [*Quiroz et al.*, 2011; *Heidinger et al.*, 2012; *Espinoza et al.*, 2015]. As a result little is known regarding the small-scale variability of precipitation in this region, even though it is likely to be strongly modulated by topography. Some preliminary analyses on the spatial distribution of precipitation in this region have been performed by means of high-resolution satellite data. For instance, *Bookhagen and Strecker* [2008] used TRMM data to analyze how the Andean orography affects the precipitation distribution. *Hunink et al.* [2014] combined satel- lite data, data from local weather stations, and reanalysis products to reconstruct precipitation at high resolu- tion in the Andes of Ecuador using a multilinear model. *Quiroz et al.* [2011] reconstructed daily precipitation over the Peruvian Andes by combining the normalized difference vegetation index (NDVI) and data from a few local weather stations. Other studies based their spatial reconstructions on interpolation techniques [*Wilks*, 2011], but those techniques do not permit a correct (local) representation unless topography is impli- citly included in the interpolation algorithm [*Bookhagen and Strecker*, 2008; *Posadas et al.*, 2015]. An alterna- tive approach is to use geographically dependent land surface information from satellites, as is the case when using NDVI data [*Quiroz et al.*, 2011; *Hunink et al.*, 2014].

Here we analyze the local- to regional-scale precipitation distribution over the broader QIC region, taking into account the topography of the central Andes, and the fact that the availability of station data in this region is very limited. Building on the one-dimensional precipitation reconstruction method established by *Quiroz et al.* [2011], we develop a new method to obtain a high-resolution, spatiotemporally complete precipitation reconstruction using NDVI data and a few rain gauge stations. The resulting NDVI-derived precipitation pro- duct allows analyzing precipitation in this region at a much higher spatial and temporal resolution than what was previously possible. Section 2 presents the study area, the in situ and satellite-derived data used, and how the data were preprocessed. In section 3 the high-resolution precipitation reconstruction method is dis- cussed in detail. Section 4 includes a discussion of the results, and a validation of their representativeness

using a set of independent stations. TRMM and topographic data are used alongside the reconstruction in interpreting local conditions and the rain-snow line around QIC. Section 5 contains a discussion of the potential and caveats of our method and ends with conclusions and an outlook on further studies.

## 2. Study Area and Data

### 2.1. Quelccaya Ice Cap Region

Quelccaya Ice Cap (QIC) is the largest tropical ice cap, located in the Cordillera Vilcanota of southern Peru (13°56′S, 70°50′W; Figure 1a), with an approximate summit elevation of 5670 m above sea level (asl) [*Thompson et al.*, 2013]. The ice cap is undergoing an accelerated retreat at a rate of $0.57 \pm 0.10 \, \text{km}^2 \, \text{yr}^{-1}$ since 1980, as evidenced in a study by *Hanshaw and Bookhagen* [2014]. This retreat is consistent with the retreat occurring throughout the tropical Andes, including in the Cordillera Blanca and the Cordillera Ampato [*Rabatel et al.*, 2013; *Vuille et al.*, 2008b], to the north and south of the Cordillera Vilcanota, respectively. A land cover map (Figure 1b) presents the main land surface characteristics of the study area, which is dominated by dry grassland (puna) and shrubland.

### 2.2. Observed Precipitation Data Around QIC

Daily precipitation data from 13 weather stations were used over the period 2000 to 2009. The stations were selected from a network of 29 initial stations belonging to the Peruvian Servicio Nacional de Meteorologia e Hidrologia (see Table 1) and available from the Autoridad Nacional del Agua (ANA) official webpage (www.ana.gob.pe). The stations are labeled with an ID, following the nomenclature of ANA. The 13 precipitation records (Table 1 and Figure 1a) were selected and preprocessed as follows: (1) the stations had to be located above 3000 m and contain more than 75% complete data; (2) precipitation for each station was summed up over periods of 10 calendar days, to create precipitation dekadals, to allow a comparison with the 10 day NDVI unit calendar time scale; (3) a moving average with an 11 day window size was applied to fill gaps and obtain 100% continuous data; and (4) a strong linear relationship between precipitation and NDVI (max $R^2_{\text{NDVI,precipitation}} > 0.35$ with $p < 0.05$) was required by all stations to guarantee a physical relationship between NDVI and precipitation. The NDVI product is calculated over all land surfaces including bare rocks, ice, or lakes, where no vegetation is present. In our study area a weak NDVI-precipitation relationship is mainly due to such elements other than vegetation, which do not respond to precipitation variability in the way that vegetation does. Since we are trying to generate a spatial precipitation product based on a vegetation index, it is important to remove such negative nonvegetation effects by only including stations with a strong linear NDVI-precipitation relationship. Thus, removing stations with a low $R^2$ ensures that the NDVI-precipitation relationship really represents statistically and physically the vegetation response to precipitation. Here one station was discarded for this reason. More details and a discussion of this point are given in sections 2.3 and 5. Vegetation (the NDVI signal) responds to precipitation with a lag, but since the study region represents fairly homogeneous precipitation [*Quiroz et al.*, 2011], and land surface characteristics (Figure 1b), the lag values should be comparable across station locations. Therefore, stations with an excessive NDVI response time (≥7 dekadals) were omitted as they point toward nonclimatic factors affecting the NDVI response (three stations were discarded for this reason). Similar to criteria 4 above, nonvegetated surfaces can influence the lag calculation and vegetation can also respond to nonprecipitation effects such as irrigation, river runoff, etc. More in-depth environmental monitoring based on data from others satellite products might aid in exploring spatial variability in this lag in response to varying land surface conditions, but this aspect is beyond the scope of the present study. This issue is discussed in more detail in section 2.3.

Data from an additional weather station, Tambopata (14.22°S, 69.15°W, 1340 m asl), were used at a later stage in the validation process to highlight the fact that our method is not applicable over lower elevation forested terrain (section 4.4.2). Note that Tambopata is a low-elevation station, which does not satisfy the selection criteria discussed above and applied to the 13 precipitation stations on the Central Andes plateau.

Continuous hourly snow height change (m) data between June 2005 and December 2009, obtained from an automated weather station (AWS), located at QIC summit (5680 m asl, 13.9°S, 70.8°W) [*Hurley et al.*, 2015], were summed up to 10 day (dekadal) snow height totals. Since positive and negative snow height change represent accumulation (precipitation) and ablation, respectively, only positive dekadal snow height values were used and multiplied with the mean snow density (0.236 or 263 kg m$^{-3}$) measured near the AWS, to
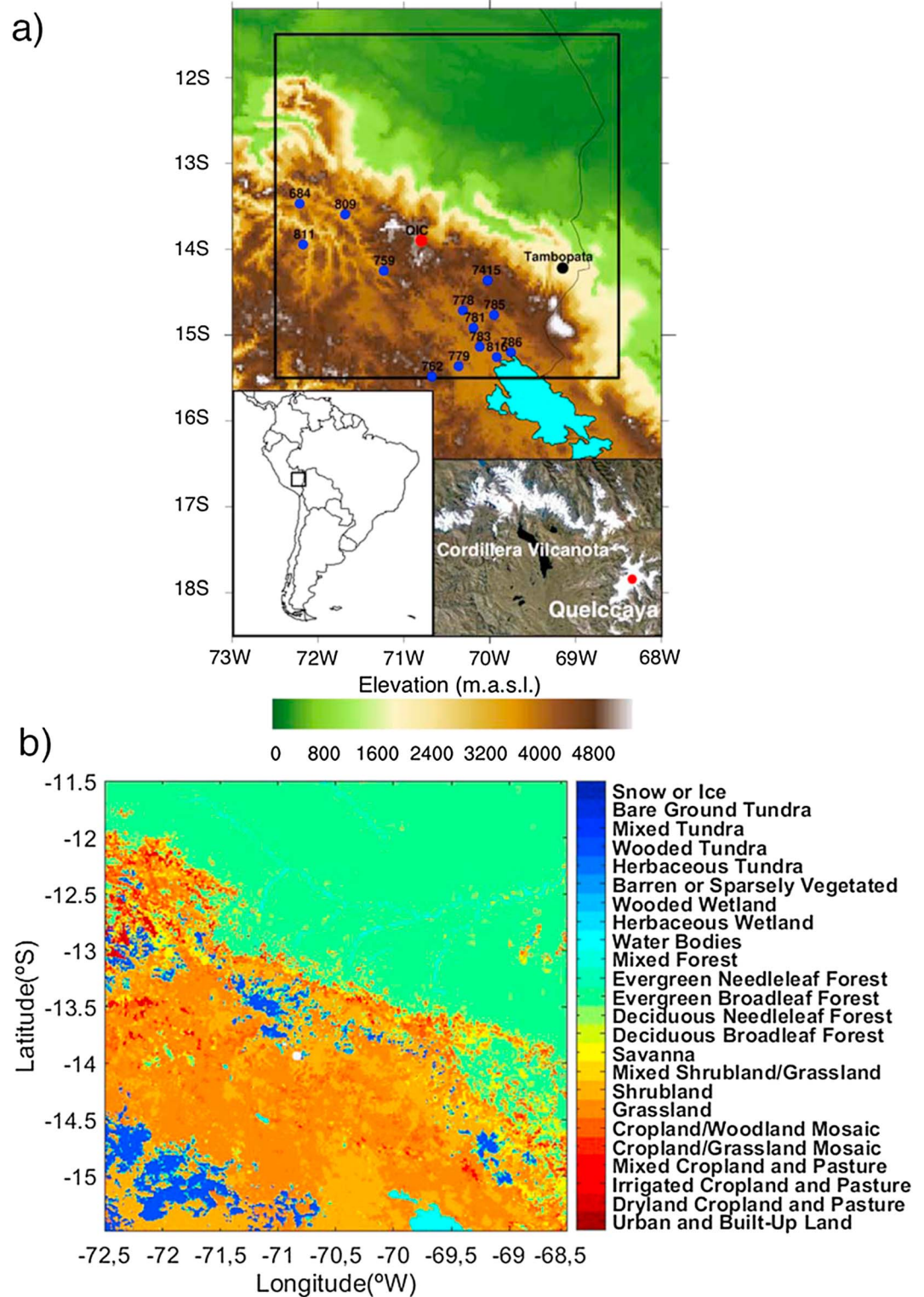
**Figure 1.** (a) The study area selected is highlighted by the black square, with coordinates 11.5°S–15.5°S and 72.5°W–68.5°W. Weather station locations are shown with blue dots, including the AWS (red dot) located on QIC where snow accumulation was measured. Color shading represents the elevation in meters above sea level (m asl). (b) USGS South America land cover version 2.0 map over the study area. The original land cover array is 446 rows and 452 columns but was resampled to a 449 × 449 array using a bicubic interpolation technique to match the spatial scale of the NDVI data (~1 km). Quelccaya ice cap location is indicated by the white dot.

**Table 1.** Weather Stations Located in the Study Area (Figure 1) Satisfying the Criteria Described in Section 2.3[a]

| Order | ID (by ANA) | Name of Stations | Elevation (m asl) | Latitude | Longitude | Lag (dekadals) | max $R^2$: Prec Versus Lagged NDVI |
|---|---|---|---|---|---|---|---|
| 1 | 684 | Anta Ancachuro | 3340 | −13.468056 | −72.215556 | 4 | 0.45 |
| 2 | 759 | Sicuani | 3574 | −14.253333 | −71.236944 | 4 | 0.66 |
| 3 | 762 | Pampahuta | 4400 | −15.483528 | −70.675778 | 5 | 0.54 |
| 4 | 778 | Progreso | 3980 | −14.715633 | −70.314492 | 4 | 0.57 |
| 5 | 779 | Lampa | 3892 | −15.363306 | −70.364769 | 5 | 0.58 |
| 6 | 781 | Azangaro | 3863 | −14.914361 | −70.19075 | 5 | 0.59 |
| 7 | 783 | Arapa | 3830 | −15.13625 | −70.118222 | 5 | 0.52 |
| 8 | 785 | Munani | 3948 | −14.766944 | −69.951806 | 5 | 0.58 |
| 9 | 786 | Huancane | 3890 | −15.2015 | −69.753556 | 5 | 0.54 |
| 10 | 809 | Caicay | 3150 | −13.594494 | −71.687153 | 5 | 0.60 |
| 11 | 811 | Tambobamba | 3275 | −13.945556 | −72.175556 | 5 | 0.68 |
| 12 | 816 | Taraco | 3820 | −15.255836 | −69.920339 | 5 | 0.46 |
| 13 | 7415 | Crucero | 4183 | −14.364194 | −70.025917 | 5 | 0.54 |

[a]The lag and the maximum coefficient of determination (max $R^2$) values correspond to the precipitation-NDVI dekadal time series comparison during the period 2000–2009.

calculate the equivalent precipitation on QIC. The same data-smoothing process used for the 13 precipitation stations (step 3) was applied to this calculated precipitation data as well.

### 2.3. NDVI as a Proxy for Precipitation

The linear relationship between grassland vegetation and precipitation is documented in recent studies over the Andes [*Hunink et al.*, 2014; *Quiroz et al.*, 2011], as well as in other mountain areas, such as Tibet [*Immerzeel et al.*, 2005]. This relationship can be observed in arid or semiarid regions such as the Altiplano, where annual precipitation ranges from 50–400 mm in the west to 600–1000 mm in the east [*Vuille and Keimig*, 2004] and vegetation is mainly dominated by grassland and shrubland, as is the case in our study area (Figure 1b), which responds to the amount of precipitation [*Quiroz et al.*, 2011], or on the Tibetan Plateau, where vegetation greening is linked with the occurrence of the monsoon [*Immerzeel et al.*, 2005]. This linear relationship can be quantified using the normalized difference vegetation index (NDVI), which is calculated as the ratio between the difference and the sum of spectral reflectance measurements in the visible and near-infrared portion of the electromagnetic spectrum. The NDVI varies between −1.0 and 1.0, with a value of 1.0 indicating maximum biomass. The NDVI dekadal and ~1 km spatial resolution synthesis data from SPOT-VEGETATION Products (VGT-S10) are used in the present analysis. The dekadals are defined according to the legal calendar: The first dekadal labeled 1 corresponds to the period from the 1st to 10th day; the second dekadal labeled 11, from the 11th to 20th day; and the third dekadal labeled 21, from the 21st day to the end of each month. Each pixel of VGT-S10 products contains the maximum NDVI value recorded over the period of 10 days, to ensure a minimal effect of cloud cover. Original NDVI digital values range between 0 and 255 but were rescaled to 0–1 values.

The present study focuses on the precipitation reconstruction process over the 2000–2009 period, but the NDVI data were selected from 1 January 2000 to 21 December 2010, because the NDVI lags precipitation [*Quiroz et al.*, 2011]. Details of this lag correction, as well as the NDVI noise correction preprocessing, are explained in the next paragraphs.

Satellite measurements of land cover are often disturbed by the intermittent optical depth in the atmosphere, sensor instability, or orbit deviations [*Immerzeel et al.*, 2005]. In this context a cloud or noise correction is the most important preprocessing step to be applied to NDVI data. There exist several methodologies to remove the noise from NDVI signals [e.g., *Hird and McDermid*, 2009; *Hunink et al.*, 2014], but the best methodology depends on the NDVI signal (location, climatic conditions, etc.) and its application. Because we are interested in using the low-frequency NDVI signal, we removed the cloud/atmospheric noise by applying a running average filter, which we applied 3 times, over the first to third forward-backward neighboring dekadal NDVI data (i.e., without taking the central value) consecutively. An example of the NDVI filter result is given in Figure 2.

The NDVI field corresponding to the present study area (section 2.1) is given by a matrix of 449 rows × 449 columns × 360 dekadals with a grid length of approximately 1 km. The linear dependence of NDVI on precipitation (Table 1) is phase lagged (i.e., vegetation responds with a delay to precipitation); therefore, the NDVI
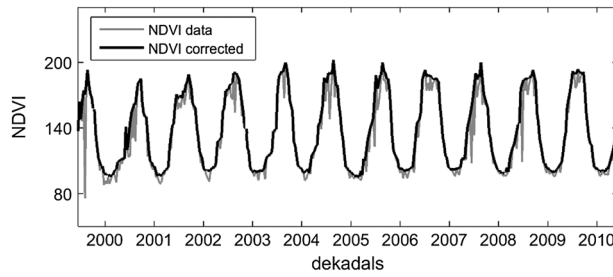
**Figure 2.** NDVI dekadal correction process using a smoothing methodology (section 2.3). The gray lines show the raw NDVI decadal data including extreme minimum values corresponding to periods of clouds and strong atmospheric scattering (noise). The bold black line corresponds to the filtered (corrected) NDVI data. Years on the x axis are centered on the third dekadal of June. The NDVI values in the y axis correspond to the original NDVI units from 0 to 255.

signal was shifted in time in order to remove the lag between both signals for each station location, thereby maximizing the $R^2$ value between both signals (Figure 3a). In order to maintain the length of the time series post lag correction, the NDVI data period was selected from 1 January 2000 to 21 December 2010. An example is given in Figure 3b for Sicuani station, showing a comparison between NDVI and precipitation, both without and after (Figure 3b) applying such a lag correction. Lags between NDVI and precipitation, however, do not in every case correspond to a delayed vegetation response to rainfall, since other surface characteristics can also affect the NDVI (reflectance) measure. At four stations (Ramis, Pomacanchi, Colquepata, and Paucartambo) out of the group of 17 initially selected, conditions 4 and 5, as indicated in section 2.2, were not met. At the station Ramis the NDVI data are uncorrelated with precipitation (max $R^2_{NDVI,precipitation} = 0.03$). This lack of
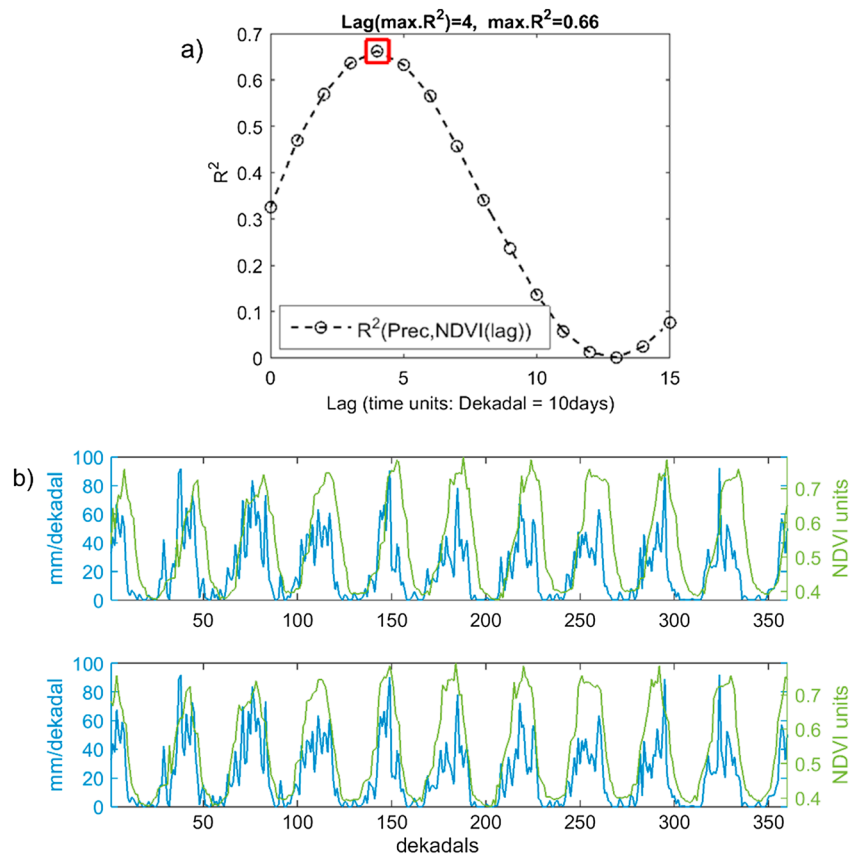


**Figure 3.** An example of the NDVI-lag determination and correction based on the station Sicuani. (a) The optimal lag value (= 4 dekadals) as indicated by the maximum $R^2 = 0.66$ (red square) based on lagged correlation between precipitation and NDVI from January 2000 to December 2009. (b) Precipitation-NDVI lag correction over the same period. Upper graph shows the dekadal precipitation (blue line) and normalized and cloud-corrected NDVI (green line) dekadal time series. Lower graph shows same time series but after shifting the NDVI data backward by 4 dekadals in order to remove the lag between precipitation and cloud-corrected NDVI data.

correlation is likely caused by the location of Ramis at the shore of Lake Titicaca, leading to a mixed spectral NDVI signal that includes lake water and bare, nonvegetated shoreline. The stations Pomacanchi, Colquepata, and Paucartambo portrayed significant relationships between NDVI and precipitation (max $R^2_{\text{NDVI,precipitation}} = 0.47$, 0.56, and 0.56, respectively), but at lags (12, 8, and 7 dekadals, respectively) that are physically implausible and much longer than the rest of the stations, all characterized by lags of either 4 or 5 dekadals (Table 1). Again the location of the stations is likely to blame for these unreasonable lags as Pomacanchi station is located along a river mouth draining into Lake Titicaca, Colquepata station is located on a ravine where the field is dominated by farms that are using a channelized rill, and Paucartambo station is located next to a river and a ravine (all verified with Google Earth). In all cases, the channelized water likely affected the spectral signal measured by the SPOT sensor. Hence, these four stations were not considered for the calibration and validation process. Some additional considerations and caveats regarding spatial variations in NDVI-precipitation lags due to varying terrain and land surface characteristics are given in section 5.

There is considerable interannual variability in the NDVI signal (e.g., Figure 2), mostly driven by the South American Summer Monsoon and ENSO, which both influence precipitation over the central and southern Peruvian Andes [*Vuille and Werner*, 2005]. While the NDVI-SASM/ENSO dependence is an interesting area to investigate in more detail, it is not the aim of the present study. Instead, here we focus on the methodological aspects of the precipitation reconstruction by means of the NDVI and on the analysis of the spatial precipitation variability in the region surrounding Quelccaya.

### 2.4. Spatiotemporal Precipitation Rate Satellite Data

Over the 2000–2009 period, the reconstruction results are compared with the Tropical Rainfall Measuring Mission (TRMM) 3B42 version 7 (v7), obtained from the Goddard Earth Science Data and Information Services Center of NASA. The TRMM Multisatellite Precipitation Analysis is a data set, which merges the microwave and infrared precipitation estimates computed every 3 h with $0.25° \times 0.25°$ latitude-longitude resolution [*Huffman et al.*, 2007]. The TRMM v7 product is preprocessed by combining a monthly multisatellite product with rain gauge data. Hence, the final TRMM v7 product is a gauge-corrected product, but we did not further correct this data set with our own station data. This version of TRMM was selected because it has shown a good rain representation over the Andes region, improving on other satellite and reanalysis products [*Zulkafli et al.*, 2014; *Blacutt et al.*, 2015; *Mantas et al.*, 2015; *Satgé et al.*, 2015; *Zubieta et al.*, 2015; *Manz et al.*, 2016; *Mourre et al.*, 2016]. It is worth noting that the precipitation rate obtained by this TRMM product includes only the liquid phase precipitation but is not taking into account the solid phase (snow).

### 2.5. Elevation Variable

To analyze the topographic characteristics of the study area (Figure 1a), the 2 min Gridded Global Relief Data (ETOPO2v2) database from the National Oceanic and Atmospheric Administration and the National Geophysical Data Center (NCAR/NGDC) was used. This database has WGS-84 horizontal projection and is in cylindrical-equidistant projection (sometimes called latitude-longitude, or geographic). The ETOPO2v2 data have a horizontal grid spacing of 2 min of latitude and longitude, and its vertical precision is 1 m. The data were rescaled to the NDVI grid resolution (1 km × 1 km, section 2.3).

## 3. Methods

### 3.1. Spatial NDVI-Precipitation Lag Calculation: Influence Zone

For locations where both NDVI and precipitation data (weather stations) exist, a correlation criterion was used to calculate the most appropriate lag to be applied to the NDVI data (section 2.3, Table 1). To apply such a lag correction over each NDVI grid cell, it first requires identifying which of the 13 base locations best represents the NDVI characteristics at each individual cell. Hence, a spatial correlation map was created for each station location, correlating its NDVI with the NDVI at each grid cell. In the end at each grid cell the station with the highest correlation coefficient was selected. This procedure allowed the creation of "influence zones" for each weather station, highlighting which precipitation station data are most pertinent to use in remote locations where weather stations do not exist. Note that the lag value calculation, which will be used in the reconstruction process with the precipitation time series (i.e., the lag obtained by comparing NDVI with precipitation at the same station), is independent from the influence zone calculation (i.e., each grid cell in the domain is

attributed to a zone by comparing its NDVI with the NDVI at the 13 stations). The lag calculation details were given in section 2.3 and will be discussed in section 5.

### 3.2. Spatial High-Resolution Wavelet Precipitation Reconstruction Method

Here the temporal reconstruction method using wavelet as described by *Quiroz et al.* [2011] and *Heidinger et al.* [2012] was modified and adapted to include a spatial dimension, with the aim of producing a spatially complete gridded precipitation field derived from NDVI data and precipitation recorded at local stations near QIC.

The reconstruction is based on the decomposition-reconstruction process of each signal (NDVI and precipitation data) in each location (of the NDVI grid field), using the dyadic Wavelet Orthonormal Multiresolution Analysis (WOMA) and symlet wavelet mother [*Quiroz et al.*, 2011]. Here we used the wavelet tool in the Matlab software. The wavelet decomposition is given as

$$\text{NDVI}(x,t) = \mathbf{A}_{\text{NDVI}}(x,t) + \mathbf{D}_{\text{NDVI}}(x,t),$$
$$\text{Prec}(x,t) = \mathbf{A}_{\text{Prec}}(x,t) + \mathbf{D}_{\text{Prec}}(x,t),$$

(1)

where $\mathbf{A}$ and $\mathbf{D}$, are understood as the "Trend" (or low frequency) and "Detail" (or high frequency) wavelet-decomposed signals, respectively, $x$ = (latitude-longitude), and $t$ is the time variable. Since NDVI and precipitation signal are linearly correlated (Table 1, $R^2$ values between precipitation and NDVI lagged), it is possible to find the state where $\mathbf{A}_{\text{NDVI}}$ and $\mathbf{A}_{\text{Prec}}$ have a good correlation as well [*Quiroz et al.*, 2011; *Heidinger et al.*, 2012]. Then the wavelet reconstruction model can be written as

$$\text{Rec}(x,t) = \mathbf{C}(x) * \mathbf{A}'_{\text{NDVI}}(x, t+\text{lag}) + \mathbf{D}_{\text{Prec}}(x, t),$$

(2)

Here the $\mathbf{A}_{\text{NDVI}}$ signal is rescaled to 0–1 values (= $\mathbf{A}'$), Rec is the reconstructed precipitation, and $\mathbf{C}$ is a scaling factor at each location $x$. Several studies have used linear models like Rec = $\mathbf{C} * \mathbf{A}_{\text{NDVI}}$, but this kind of model does not preserve the high-frequency precipitation variability ($\mathbf{D}_{\text{Prec}}$). The present study is focused on a reconstruction that maintains the precipitation variability and stochasticity properties. The value of $\mathbf{C}$ (scaling factor) is defined as

$$\mathbf{C}(x) = \overline{\mathbf{A}}_{\text{Prec}}(x) / \overline{\mathbf{A}'}_{\text{NDVI}}(x),$$

(3)

where the overbar indicates the mean value over time.

## 4. Applying the Spatiotemporal Reconstruction Method Around QIC

### 4.1. Obtaining the Influence Zone Matrix

For the cross-validation process (section 4.3) six randomly selected weather stations were withheld (Table 3), and the remaining seven weather stations were used to calculate the influence zone matrix (Figure 4a). Additionally, a second influence zone matrix using all 13 weather stations (Figure 4b) was developed. A topography mask was applied to each influence matrix masking data below 2500 m asl. This was done for several reasons as follows:1. The vegetation cover to the northeast (forested region) is completely different from the vegetation over the Andes (grassland and shrubland); hence, the linear (or otherwise) association between precipitation and vegetation is lost over the forested region (results not shown, but this can be deduced from Figure 8 and section 4.4.2), making our approach unsuitable.2. The atmospheric conditions are distinctly different in the tropical forested lowland areas when compared to the semiarid conditions in the Andes, which are required to successfully apply the reconstruction technique by *Heidinger et al.* [2012] and *Quiroz et al.* [2011].
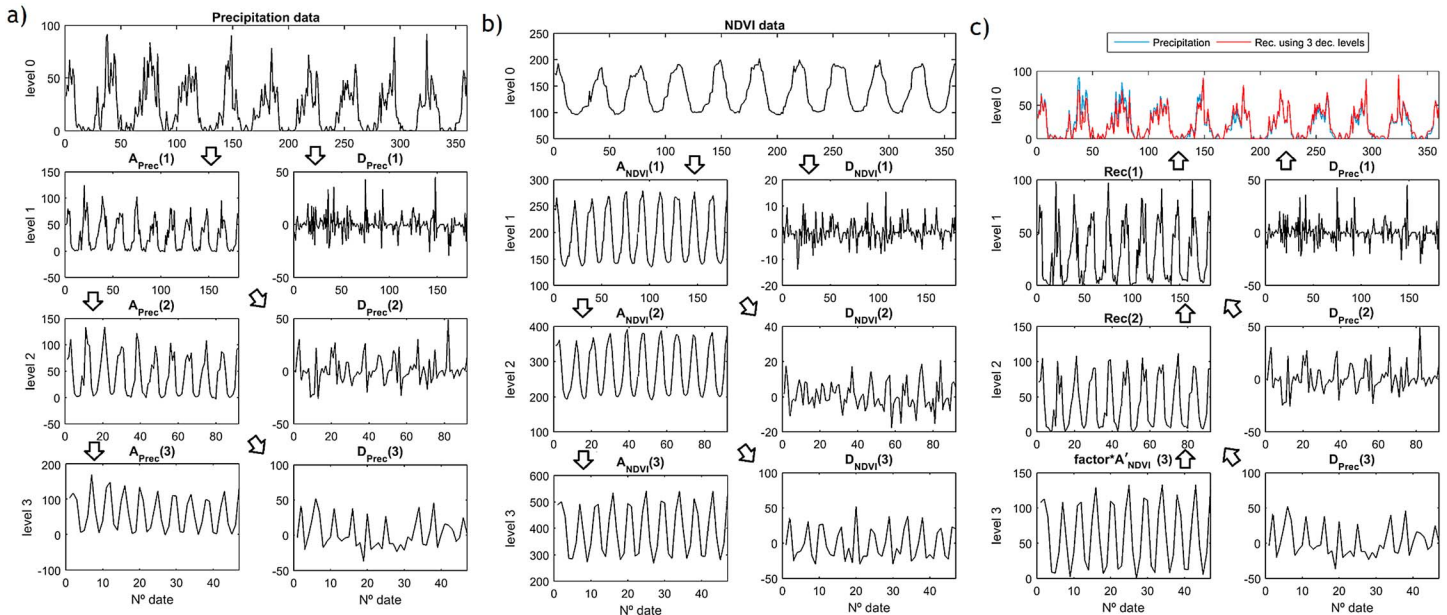
### 4.2. Selection of the Level of Reconstruction

Although the equations (1) and (2) are given for one step of decomposition (Figures 5a and 5b) and reconstruction (Figure 5c), respectively, those can be applied several times [*Quiroz et al.*, 2011; *Heidinger et al.*, 2012]. For instance, applying the decomposition (equation (1)) 5 times (or five levels of decomposition) to $\mathbf{A}_{\text{NDVI}}$ and $\mathbf{A}_{\text{Prec}}$ results in new trends and detailed signals for each application. We can also start the reconstruction process (i.e., use the equation (2)) in the decomposition level that we have selected from 5 to 1 (i.e., the process starts in reverse order). But what is the adequate level of decomposition to start the reconstruction process? To answer this question, we analyzed the decomposition results up to level 5 in
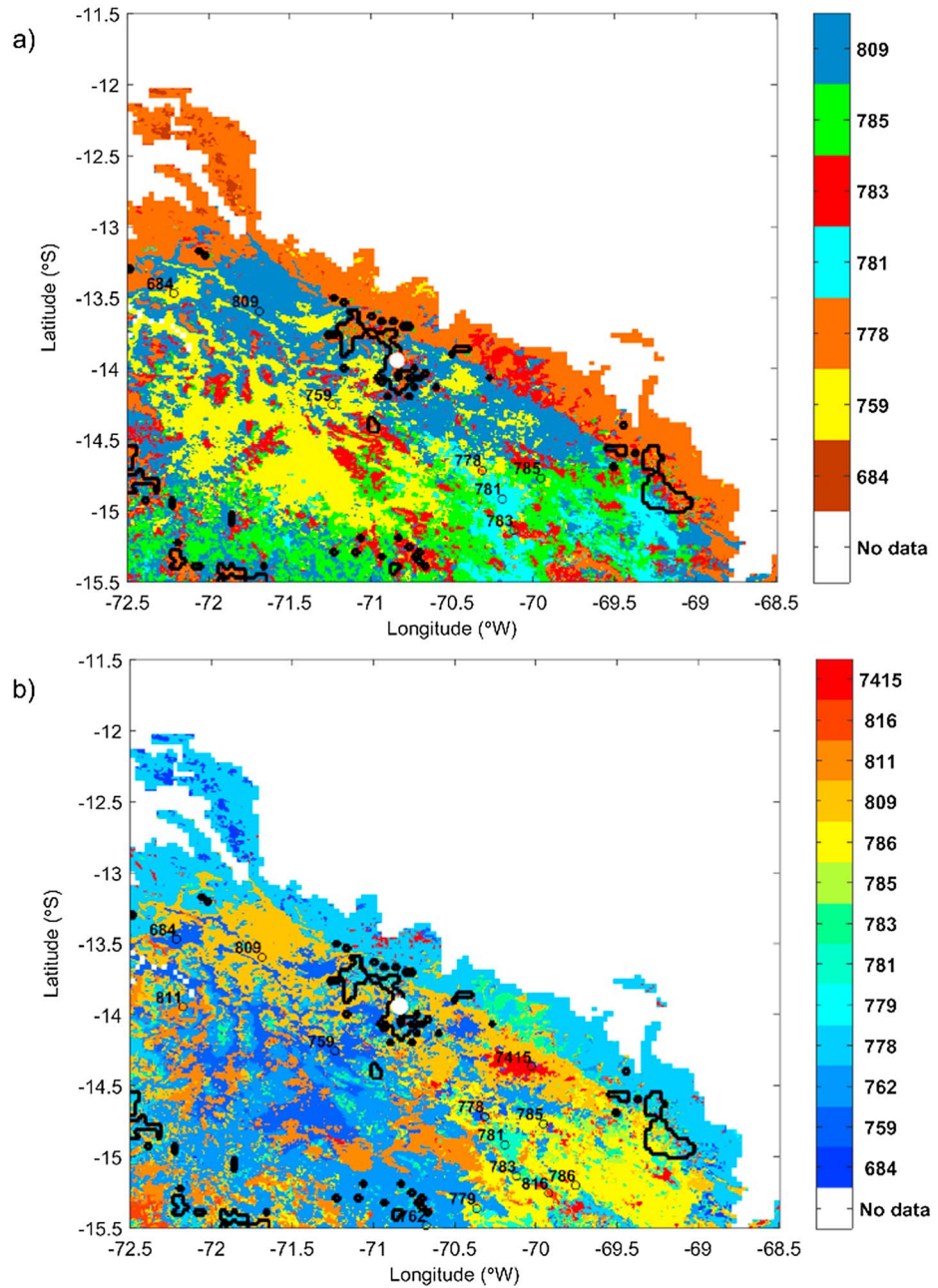
**Figure 4.** (a) Influence zone matrix derived from 7 out of 13 weather stations. The other seven stations are retained for cross-validation purposes (section 4.3). A topography mask was used to remove regions below 2500 m asl (section 3.1). Numbers on the color bar correspond to the ID of each station in Table 3. (b) Influence zone matrix at ~1 km$^2$ resolution, derived using 13 weather station locations with a $R^2 > 0.35$ and $p < 0.05$, between NDVI(ST) versus NDVI($x$), where $x$ represents locations over the study area, and ST the station locations above 2500 m asl. Numbers on the color bar correspond to the ID of each station in Table 1. Black contours in Figures 4a and 4b correspond to topography above 5000 m asl, and white dot is representing QIC location.

Table 2. Level zero corresponds to the original precipitation and NDVI corrected data, for each NDVI grid location. The precipitation time series were extracted from the group of 13 stations selected previously for each NDVI grid influence zone matrix (Figure 4b), with the corresponding lag applied to the NDVI time series at each grid location. In the first column of Table 2, levels 1–5 indicate the number of times the wavelet decomposition process was applied over Trend (A). Using the Shannon's entropy ($E$) and the coefficient of determination ($R^2$), we evaluated which level of decomposition ($i$) should be used to start the reconstruction process. Table 2 shows

**Figure 5.** Decomposition-Reconstruction Process based on the station Sicuani. Here the level 3 of reconstruction is shown (section 4.2). (a) Precipitation decomposition (downward arrows) using symlet-2 wavelet, until the level 3 is reached. (b) As in Figure 5a but for the NDVI cloud (Figure 2) and lag-corrected data. (c) Reconstruction (upward arrows) from NDVI trend and precipitation detail signals at level 3, using a scaling factor (equation (3)), and $A'$ means that $A$ is previously normalized to 0–1 values. Precipitation details and reconstruction data are used for upper levels to obtain the final reconstruction at level 0 (i.e., at dekadal time scale).

a natural behavior of entropy (based on the mean across all grid locations), with an exponential decay (column 6) across the levels i, without any clear breaking point, unlike in *Quiroz et al.* [2011] or *Heidinger et al.* [2012]. $R^2$ values between NDVI and precipitation trends portray similar values at all levels (column 7), increasing in concordance with the decomposition levels. Additionally, observed and reconstructed precipitation based on several decomposition levels were compared (column 8), using the adjusted $R^2$ as metric (since the reconstruction's degrees of freedom are increasing at each level). The $R^2$ suggests that a better precipitation reconstruction can be achieved when a higher level of decomposition is used. But in general the entropy, (ordinary) $R^2$, and adjusted $R^2$ are showing that further gains after the third level of decomposition are much smaller than at lower levels. Thus, hereafter, we will work with the reconstruction process starting at level 3.

### 4.3. Precipitation Reconstruction and Cross Validation

An example of the reconstruction process is presented in Figure 5, adjusting the space dimensions to the Sicuani station location. Here one level of decomposition was applied. This model was applied to each NDVI grid cell, using the appropriate influence zone matrix. The resulting reconstructed precipitation data

**Table 2.** Shannon's Entropy ($E$), Entropy Difference ($\Delta E$), and the Coefficient of Determination ($R^2$) for NDVI and Precipitation (Prec) for Different Wavelet Decomposition Levels (i)[a]

| Level (i) | Time Scale | $E_{NDVI}$ Trend i | $E_{Prec}$ Trend i | $\|\Delta E\| = \|E_{NDVI}$ Trend i $- E_{Prec}$ Trend i$\|$ | $\|\Delta E/max(\Delta E)\| \times 100\%$ | $R^2$ From NDVI Trend$_i$ Versus Prec Trend$_i$ | Adjusted $R^2$ From Rec$_i$ Versus Prec |
|---|---|---|---|---|---|---|---|
| 0 | 1 dekadal | 3.12 | 2.85 | 0.27 | 100 | 0.57 | --- |
| 1 | 2 dekadals | 3.13 | 2.88 | 0.25 | 94 | 0.67 | 0.69 |
| 2 | 4 dekadals | 2.86 | 2.75 | 0.11 | 41 | 0.76 | 0.79 |
| 3 | 8 dekadals | 2.52 | 2.51 | 0.01 | 4 | 0.85 | 0.87 |
| 4 | 16 dekadals | 2.15 | 2.15 | 0.00 | 1 | 0.80 | 0.91 |
| 5 | 32 dekadals | 1.75 | 1.78 | 0.03 | 10 | 0.53 | 0.92 |

[a]The precipitation time series is selected using the influence matrix to define which of the 13 station time series performed best at each location. The NDVI time series is corrected by applying a lag when compared with the precipitation time series. The entropy is calculated using a normalization unit for the corrected NDVI and precipitation. The absolute value ($\|E\|$) of the entropy difference between both signals is used for each level. The $R^2$ is calculated for each trend of the wavelet decomposition process applied to the corrected NDVI and precipitation of each dekadal time series, but only the mean across the 13 station locations is shown for each decomposition level. An adjusted $R^2$, between reconstructed precipitation started in level i (Rec$_i$) and measured Prec, is calculated per station location, but only the mean across the 13 station locations is presented.

**Table 3.** Cross Validation of Reconstruction Model and Statistical Comparison Using TRMM Data[a]

| | | Rec Versus Obs Dekadal | | | | TRMM Versus Obs Dekadal | | | |
|-----|--------------|------|------------|------|------|------|------------|------|------|
| ID | Name Station | r | Bias Ratio | MAE | RMSE | r | Bias Ratio | MAE | RMSE |
| 762 | Pampahuta | 0.68 | 0.86 | 13.6 | 22.3 | 0.80 | 0.94 | 11.0 | 17.6 |
| 779 | Lampa | 0.79 | 0.84 | 9.3 | 13.9 | 0.82 | 0.90 | 9.3 | 14.0 |
| 786 | Huancane | 0.79 | 0.79 | 11.1 | 18.9 | 0.87 | 0.97 | 7.3 | 10.9 |
| 811 | Tambobamba | 0.82 | 1.19 | 8.7 | 12.7 | 0.80 | 0.80 | 11.8 | 18.6 |
| 816 | Taraco | 0.83 | 0.95 | 9.0 | 13.6 | 0.82 | 1.33 | 8.3 | 12.7 |
| 7415 | Crucero | 0.76 | 1.01 | 8.9 | 13.6 | 0.76 | 1.29 | 8.1 | 12.3 |

[a]A cross validation was applied to reconstructed dekadal precipitation (Rec) using six randomly selected weather stations not included in the reconstruction process (Obs) starting at the third level of decomposition, over the entire period 2000–2009. The indices used are the linear correlation (*r*), bias (ratio = mean(Rec)/mean(Obs)), mean absolute error (MAE), and root-mean-square error (RMSE). Indices definitions were taken from *Dinku et al.* [2008].

array has dimensions corresponding to 449 rows × 449 columns × 360 dekadals, including the 1 January 2000 to 21 December 2009 dekadals.

A cross validation is applied to test the reconstruction results. Here seven stations were randomly chosen to generate the influence zone matrix and its respective precipitation reconstruction, while the remaining six stations were withheld to validate the results. This random separation is appropriate as the study region has a homogenous precipitation behavior. As was demonstrated in *Quiroz et al.* [2011], using a different station subset would not yield significantly different results. Afterward the reconstructed precipitation was compared with the observed precipitation at the location of the six weather stations that had been withheld from the reconstruction process. The statistics of this comparison are presented in Table 3 and were calculated using the definitions given in *Dinku et al.* [2008]. A comparison of the reconstructed precipitation results against the six independent precipitation records from the stations not included in the reconstruction process (Table 3, Rec versus Obs dekadal) shows high linear correlation coefficients (*r* > 0.68): bias values between 1.19 and 0.84, indicating a moderate overestimation and underestimation, and a low mean absolute error (MAE) with values between 8 and 13 mm/dekadal. This comparison indicates that the reconstruction model produces an acceptable precipitation reconstruction over this area on the dekadal scale. Similarly, Table 3 shows the statistics of the comparison between TRMM and precipitation at the six independent stations. The correlation, MAE, and root-mean-square error (RMSE) over the dekadal time scale show similar performance for TRMM as for the reconstructed precipitation. This indicates that the reconstruction using just seven stations produces an acceptable spatial representation of precipitation, but with a higher resolution.

Additionally, a comparison between the reconstruction using 7 and 13 weather stations was developed to show the advantage in the use of more stations for the spatial reconstruction, taking as reference the observed TRMM dekadal satellite precipitation field. Figure 6 shows a representative example for the first dekadal in January 2000. The reconstruction (rec) using 13 weather stations presents a more homogeneous behavior than the one using only 7 stations. The reconstruction based on 13 stations displays intense precipitation to the west of Lake Titicaca [e.g., *Mohr et al.*, 2014] with more intense precipitation at lower elevation regions across our study area, while the reconstruction using 7 stations is characterized by more intense localized precipitation to the southwest of the QIC and the northwest of Lake Titicaca. When compared with the corresponding TRMM dekadal field, both reconstructions fail to reproduce the intense precipitation over the Andean slope, although the known large bias of TRMM over the Andean slope prevents a detailed comparison. Thus, we infer that the reconstruction based on all 13 stations contains an even more detailed and reliable reconstruction.

### 4.4. Assessing the Spatial Precipitation Distribution
### 4.4.1. Spatial Aspects of the Reconstructed Precipitation
The seasonal cycle with enhanced precipitation over the October–April months (wet season) and less precipitation between May and September (dry period) is well known for the central Andes [*Vuille*, 1999; *Vuille et al.*, 2000; *Garreaud et al.*, 2003]. *Bookhagen and Strecker* [2008] show how the topographic relief plays an important role associated with orographic precipitation, but their analysis focused mainly on the eastern Andean slope, with tropical forest cover. Our precipitation reconstruction at approximately 1 km resolution allows for a more detailed local analysis of precipitation variability around QIC.
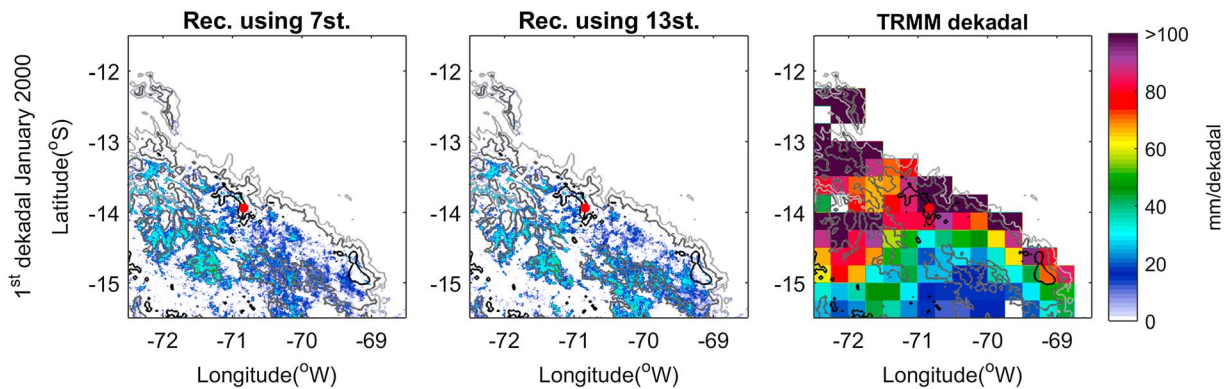
**Figure 6.** Reconstructed dekadal precipitation (Rec) for first dekadal January 2000, based on (left) 7 and (middle) 13 stations, respectively, and the corresponding (right) TRMM dekadal precipitation obtained from the raw TRMM v7 daily product. Gray, dark gray, and black contour lines indicate 3, 4, and 5 km elevation isolines, respectively. The red dot indicates the QIC location. A Lake Titicaca mask was applied.

Figures 7a and 7b show the reconstructed precipitation and TRMM climatologies, respectively. The reconstructed climatology (Figure 7a) presents a homogeneous but much more detailed precipitation pattern during December-January-February-March (austral summer) and June-July-August-September (austral winter), and a more heterogeneous pattern in April–May and October–November, which correspond to the transition seasons. During the wet periods the lower elevations receive more precipitation (<140 mm/month). An unrealistic pattern with reduced wet season precipitation, when compared with the Altiplano region, was found over the eastern Andean slope (see section 4.4.2).

The TRMM data in Figure 7b portray a homogeneous pattern except to the north and northeast of QIC and along the eastern Andean slope, where precipitation is enhanced. In fact, precipitation is overestimated by a factor of 1.5 or more by TRMM over regions with high elevation, such as QIC and the eastern Andean ridge and slope areas. This bias may be due to stratiform clouds associated with deep convection, producing ice bands at high tropospheric levels that can perturb the satellite measurements [*Mohr et al.*, 2014]. In the case of the overestimation in the neighborhood of Lake Titicaca by TRMM (for December, January and February in Figure 7b, and wet and annual in Figure 8), the lake itself could have an impact on precipitation formation [*Giovannettone and Barros*, 2009]. This overestimation is consistent with the observations at Taraco station, located close to Lake Titicaca, which features a bias value of 1.33 (TRMM versus Obs dekadal, Table 3). The problem along the eastern Andean slope is assessed separately in section 4.4.2. The fact that TRMM overestimates precipitation over the Peruvian Andes has been documented in previous studies, and several (linear and nonlinear) techniques have been proposed to correct for this bias [e.g., *Condom et al.*, 2011; *Heidinger et al.*, 2012; *Satgé et al.*, 2015]. Here we do not apply such a postprocessing to TRMM as such corrections could also be applied to the reconstructed precipitation data. We are more interested in the direct comparison between our reconstruction method and raw TRMM data, as applied in most tropical precipitation studies.

Figure 8 shows the ~1 km dry, wet, and annual reconstructed precipitation distribution over the study area. The reconstructed precipitation during the dry period shows a homogeneous behavior, with enhanced precipitation over the northeastern slopes and around some of the localized peaks with more than 5 km height, such as QIC. The wet season precipitation portrays a more heterogeneous behavior with increased precipitation at lower elevations, principally west of Lake Titicaca. The reconstruction, however, also produces reduced precipitation along the eastern Andean slopes when compared to the Altiplano itself, which is unrealistic. When the data are aggregated to annual precipitation totals, the reconstruction presents relatively homogeneous precipitation amounts, oscillating around 600 to 800 mm, but with increased precipitation at lower elevations, west of Lake Titicaca and to the southwest of QIC in particular.

The more homogeneous precipitation characteristics during the dry period may be indicative of the influence of large-scale to mesoscale circulation systems which may play a more dominant role [*Vuille and Ammann*, 1997] than local- or regional-scale convective processes, dominating during the wet season summer months and being more strongly modulated by the complex topography [*Garreaud et al.*, 2003].

Figure 8 also displays the dry, wet, and annual TRMM distributional analysis. The TRMM v7 data were not corrected with the in situ station data, as this product already includes a correction process based on local
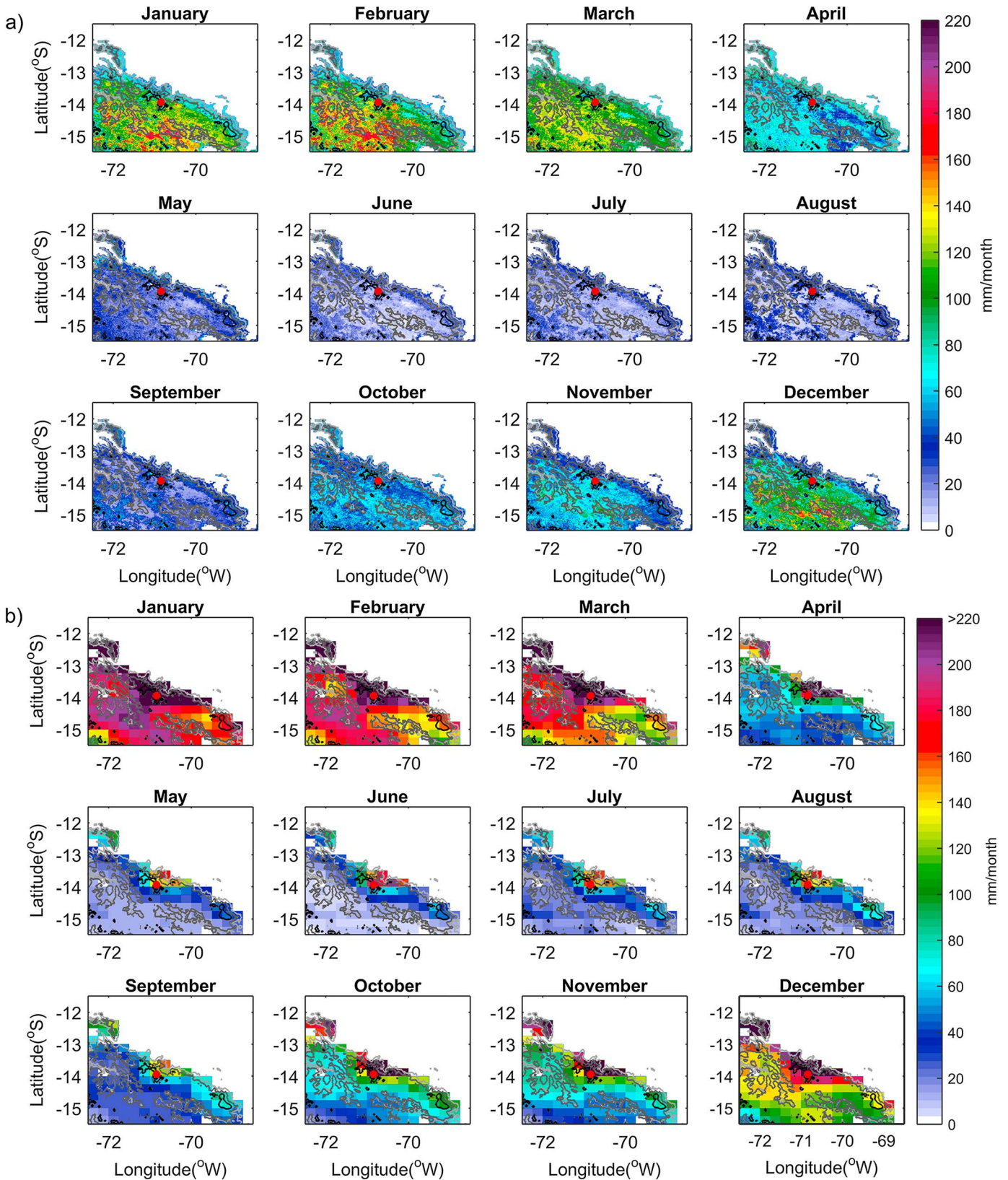
**Figure 7.** (a) Reconstructed monthly precipitation climatology. Gray, dark gray, and black contour lines indicate 3, 4, and 5 km elevation isolines, respectively. The red dot indicates the QIC location. (b) As in Figure 7a but for TRMM monthly precipitation climatology obtained from TRMM v7 daily product without any postcorrection process (raw TRMM v7 product). A Lake Titicaca mask was applied.
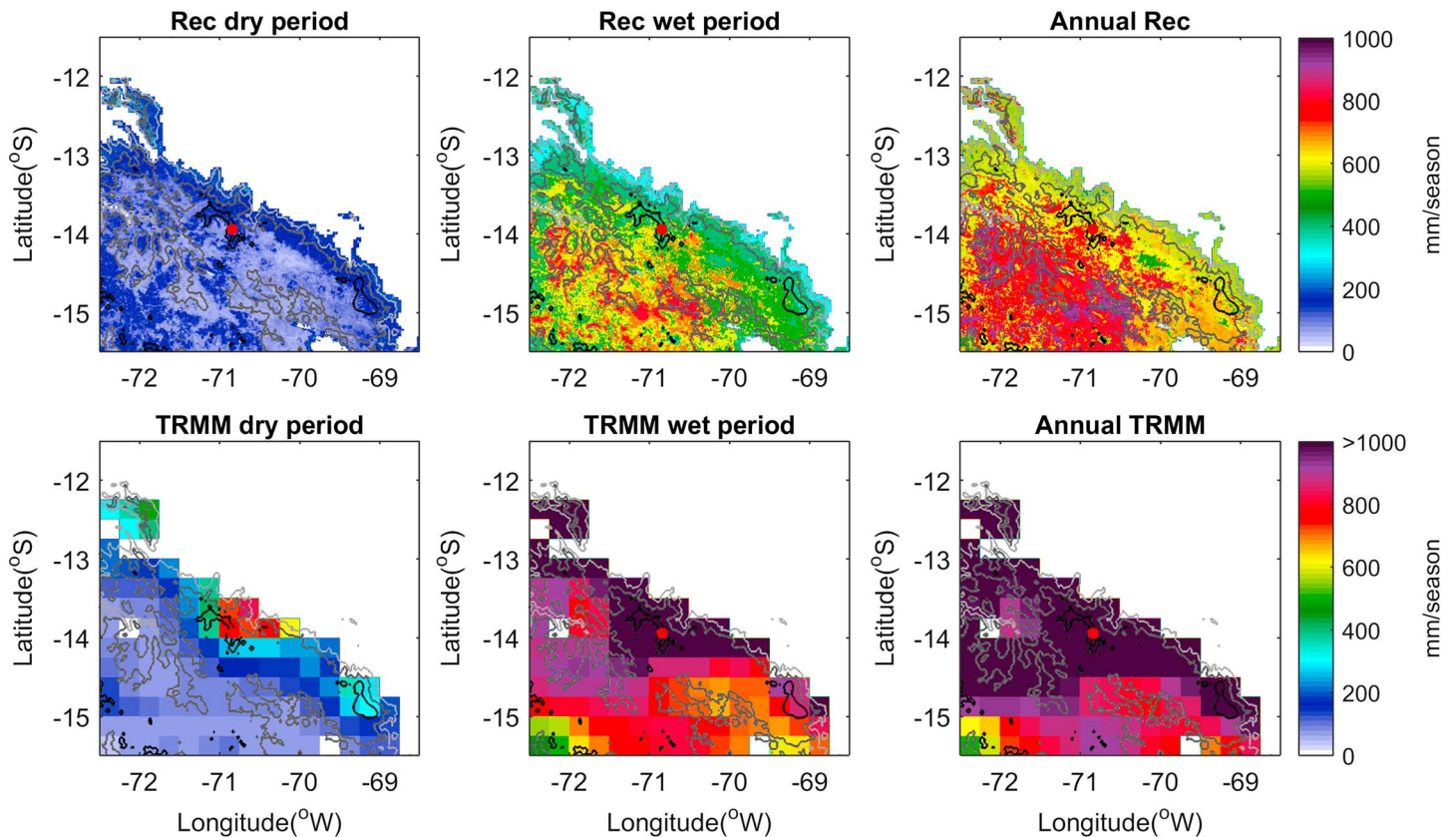
**Figure 8.** (top row) Reconstructed precipitation and (bottom row) TRMM for (left column) dry, (middle column) wet seasons, and (right column) annual total. Gray, dark gray, and black contour lines indicate 3, 4, and 5 km elevation isolines, respectively. The red dot indicates the QIC location. A Lake Titicaca mask was applied.

station data. Studies by *Scheel et al.* [2011] and *Heidinger et al.* [2012] show a large overestimation of precipitation by TRMM products over high-elevation locations, and details of the TRMM versions 6 and 7 bias are discussed in *Zulkafli et al.* [2014].

The wet and annual reconstruction patterns are comparable to the TRMM cases (Figure 8). Across the eastern Andean ridge, however, our reconstructed precipitation pattern is contrary to the well-known northeast-southwest decreasing precipitation trend [e.g., *Mohr et al.*, 2014, Figure 3; *Satgé et al.*, 2015, Figure 2] with the precipitation gradient being reversed. This problem is assessed in more detail in the following section.

### 4.4.2. Assessing the Reconstruction Over the Eastern Andean Slope

The reconstruction represents precipitation over the Andes quite well, when compared to the TRMM data (Figures 7 and 8). However, the reconstruction shows an unrealistic behavior over the eastern Andean slope during the wet season. The eastern Andes region is a highly complicated region to obtain accurate precipitation information, be it from satellite, reanalysis, or weather station measurements [*Blacutt et al.*, 2015; *Espinoza et al.*, 2015]. During December, for instance, the reconstructed precipitation over this slope is about 30 mm/month, much lower than the amount on the Altiplano, which ranges between 80 to 120 mm/month. To assess this problem over the Andean slope, we used data from an additional independent station, Tambopata (~1340 m asl), not used in the reconstruction and located at a much lower elevation on the Andean slope (Figure 1a).

We selected four locations for this additional analysis: Sicuani, Azangaro, Crucero, and Tambopata. Sicuani and Azangaro are located on the Altiplano, while Crucero is located close the eastern Andean ridge (~3042 m asl), and Tambopata, as discussed, at a much lower elevation on the eastern slope. The seasonal cycle of precipitation (Figure 9a) is similar at all stations, but with much higher values in all months at Tambopata. Moreover, intense precipitation events are not uncommon in Tambopata, as occurred during the July 2009 dry season, when more than 250 mm of precipitation fell within a week. This amount is similar
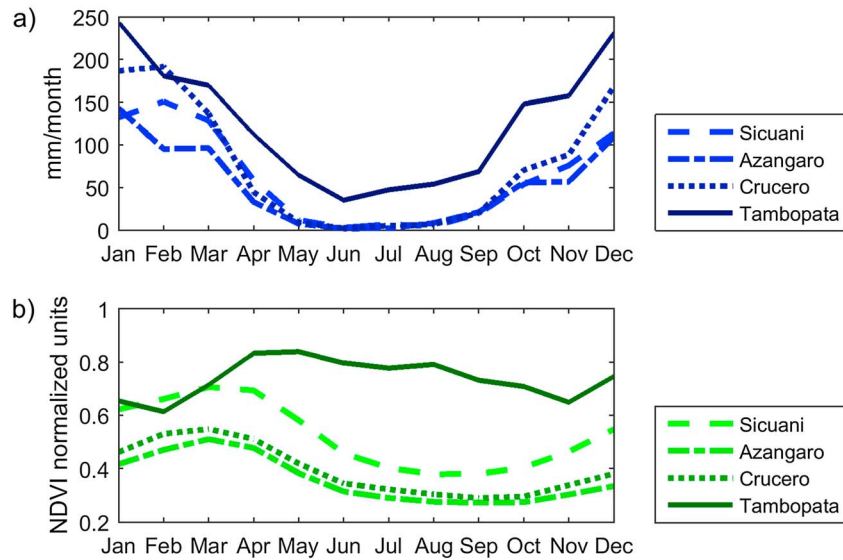
**Figure 9.** Comparison of (a) observed precipitation and (b) NDVI seasonal cycle. Sicuani, Azangaro, Crucero, and Tambopata are located to the west, south, north, and east of QIC, respectively. Sicuani, Azangaro, and Crucero belong to the QIC system, while Tambopata is located on the eastern Andean slope (1340 m asl), much below the Altiplano elevation.

to the maximum monthly precipitation over the Central Andes region during the wet season. In the case of NDVI (Figure 9b), the Andean precipitation seasonality is well reproduced by Sicuani, Azangaro, and Crucero, but completely lost in the case of Tambopata, likely because of orographic precipitation enhancement along the Andean slope throughout the year (Figure 9a). In addition, the lack of clear days during the wet season produces extremely perturbed NDVI measurements around Tambopata. As a result the precipitation-NDVI relationship as derived from Andean stations is not representative over the Andean slope region.

### 4.4.3. Comparison of Reconstructed Precipitation With Snow Accumulation at QIC

In Figure 10, dry and wet season average reconstructed precipitation data are shown over a smaller domain centered on QIC. Precipitation around QIC is distributed rather uniformly, but there is a clear enhancement of precipitation over the higher topography of QIC, especially on its eastern flank. While this is consistent with
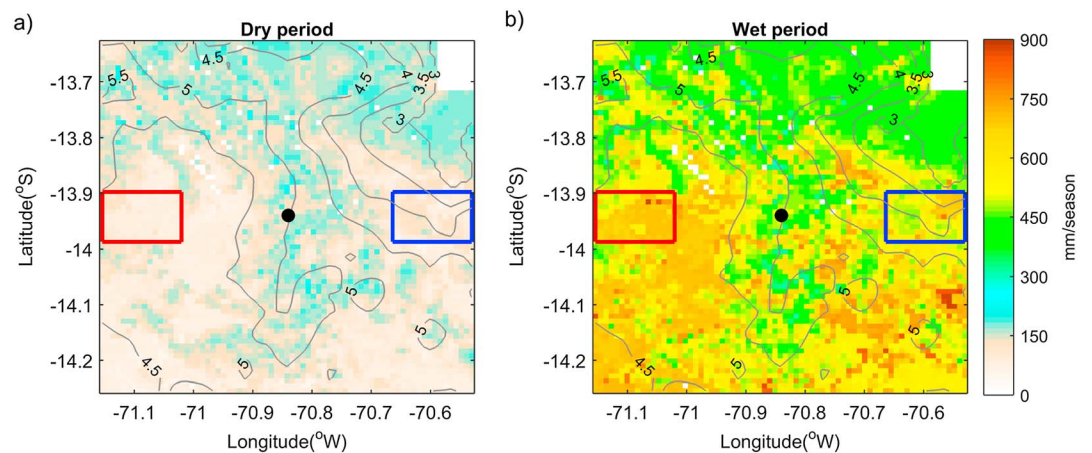


**Figure 10.** Local analysis of reconstructed precipitation surrounding QIC, with a 0.0089° (or ~0.9805 km) resolution, for (a) dry and (b) wet periods. The black dot indicates the QIC location. The western and eastern QIC rectangles (10.79 km × 15.69 km or 0.0981° latitude × 0.1426° longitude or 11 rows × 16 columns), in red and blue, respectively, indicate the two regions selected over which precipitation was averaged to create time series from 2000 to 2009. Gray contour correspond to 3 to 5.5 km elevation with a 0.5 km contour interval. White areas lack data as they are located below 2500 m asl.

the notion of easterly moisture advection during the wet season, dry season precipitation on QIC itself is significantly overestimated by our reconstruction in Figure 10a.

At the QIC summit (5670 m asl) snowfall has been routinely measured at hourly resolution since 2004 [*Hurley et al.*, 2015]. Hence, it is worth investigating the spatial coherence between these measurements with our reconstructed precipitation using NDVI. However, our reconstruction lacks physical representation or definition of objects where vegetation is absent, as is the case on QIC and its surroundings, which are covered by ice and bare rocks. Thus, the precipitation values produced by our reconstruction on the QIC itself are unrealistic, with dry season values of approximately 200–300 mm (Figure 10a), when in reality snowfall during the dry season is limited in magnitude [*Hurley et al.*, 2015]. We can therefore not directly compare calculated precipitation (from snowfall totals measured by the AWS, section 2.2) on QIC with estimated values from our precipitation reconstruction.

Since we cannot compare AWS snowfall totals with our reconstructed precipitation totals on QIC itself, we instead average the reconstructed precipitation over the eastern and western sides of QIC (blue and red rectangles, respectively, in Figure 10; 10.79 km × 15.69 km or 0.0981° latitude × 0.1426° longitude each). The two dekadal precipitation time series obtained in this way are highly correlated with $r = 0.95$ ($p < 0.0001$). Hence, the temporal precipitation variability is almost identical on both sides of QIC. If we compare the observed snowfall totals at QIC (converted into water equivalent precipitation) against reconstructed precipitation to the west and east of QIC at the dekadal scale between 2005 and 2009, we obtain correlation coefficients of 0.85 and 0.79, respectively, in both cases with $p < 0.0001$. This suggests that precipitation at QIC is regionally homogeneous and that snowfall at Quelccaya is spatially representative, in terms of amount and in terms of variability, of a larger surrounding region. On the other hand, this result also indicates that snowfall variability at dekadal scale at QIC summit may be estimated from nearby precipitation at lower levels.

## 5. Discussion and Conclusions

A spatiotemporal precipitation reconstruction model using NDVI data was built, validated with an independent set of data, and used to describe the climatology and spatiotemporal variability of precipitation in the Cordillera Vilcanota region, surrounding the QIC at decadal scale. This model is based on stochastic variability inheritance from daily precipitation data of local stations and a vegetation index, applying the WOMA and symlet wavelet mother, which maintain the local variability and internal information (entropy) of the precipitation data from neighboring locations.

The reconstruction method uses the low-frequency NDVI data, incorporated as a proxy for low-frequency precipitation, removing cloud-related errors and high-frequency local noise caused by various different sources (i.e., nonvegetation elements such as bare soil or snow). Vegetation responds to precipitation with a lag, which varies depending on ground water recharge and discharge, soil properties, land cover, topography, and others [*Quiroz et al.*, 2011]. A detailed quantification of this lag is beyond the scope of this study, since it would require more detailed in situ measurements at and below the ground level. However, based on the empirically established lags (Table 1), we can infer that the soil properties are inducing a delay on the order of 4–5 dekadals between precipitation and the corresponding vegetation response. This is consistent with the semiarid climatic conditions of the study area. In the Andes of Ecuador, for example, where conditions are humid all year, the NDVI-precipitation lag is between 1 and 2 dekadals [*Hunink et al.*, 2014]. Our lag values are also consistent with the delay found in *Quiroz et al.* [2011] over a similar study region located nearby. *Immerzeel et al.* [2005] showed that the NDVI-precipitation response on the Tibetan plateau, an ecosystem that is comparable to the Altiplano in terms of altitude, vegetation, and climate, is also delayed by approximately 5 dekadals [see *Immerzeel et al.*, 2005, Figure 6].

Our reconstruction method implicitly accounts for the stochastic nature of precipitation by incorporating the high-frequency precipitation signal, which is locally unique. High-frequency precipitation variability has a complex (chaotic) behavior and is retained in our reconstruction, unlike any method that would rely on a spatial interpolation technique. As such our reconstruction method is divided into two parts: the low-frequency precipitation, modeled by means of the low-frequency NDVI signal, and the high-frequency, chaotic precipitation variability. Moreover, since NDVI is taken as precipitation proxy, it carries some constraints about the local precipitation behavior. For example, the lag between NDVI and precipitation portrays a quasi-linearity, without consideration of nonlinear factors in the reconstruction model. These constraints can be relaxed

when a higher decomposition level is used, allowing the use of more information from high-frequency precipitation (detail signals) than is available from NDVI alone. Here we use the third decomposition level to start the reconstruction process, based on the entropy comparison between different decomposition levels.

We have calculated the explained variance that is simply due to the annual cycle by correlating the reconstructed mean annual cycle (climatology) versus the observed precipitation at all 13 stations ($R^2_{annual}$) and then repeated this analysis with the intraseasonal and interannual variability included ($R^2_{interannual}$). The mean $R^2_{annual} = 0.61$ is quite high, indicating the strong role played by precipitation seasonality. When using the actual reconstruction, including variability at intraseasonal and interannual time scales, $R^2_{interannual} = 0.87$ (as it was presented in Table 2 for the third level of reconstruction), highlighting the added skill of our model. Indeed, at 11 out of the 13 stations the $R^2_{interannual} \geq 0.90$.

It is important to note that the NDVI data were filtered to obtain a low-frequency (more periodic) NDVI signal. This filtering effectively removes noise from clouds and also from several other biophysical mechanisms, such as photosynthesis, transpiration, soil moisture, or objects that lack vegetation (open water, bare soil and rocks, roads, etc.). Thus, the preprocessing of the NDVI signal depends on the kind of information (frequency) that one wishes to retain in the final results. For our case, since we only assessed the reconstruction method over a period of 10 years, long-term NDVI variability and trends cannot be assessed (for instance, the El Niño impacts on NDVI). In the present study we focused on removing a major part of the high-frequency NDVI variability. Hence, the reconstructed precipitation carries more weight from the high-frequency component observed in precipitation at in situ stations. Future research will be oriented on the use of low-frequency NDVI models, such as a statistical NDVI model generation, or a functional composition model based on several environmental variables. Moreover, other environmental data from satellites, besides NDVI, might be used to further improve the precipitation reconstructions. By changing the satellite product, the same method can be used for other process studies, as was done, for example, by *Heidinger et al.* [2012] who used the one-dimensional reconstruction method of *Quiroz et al.* [2011] to correct the TRMM product over the Central Andes.

One of the advantages of our method is that spatial lags associated with spatial precipitation variability are inherently included and accounted for in our reconstruction method as it forms part of the NDVI response. For example, a mesoscale disturbance moving across the region over the course of several days will lead to rainfall on different days at different points of its trajectory. An in situ network may not detect this spatial lag if it is not sufficiently dense. The vegetation and hence the NDVI, however, will respond accordingly, showing an increasing lag along the disturbance's trajectory when compared to the region where the disturbance first emerged, in accordance with the duration of the storm travel time.

The spatial variability of the reconstruction was compared against TRMM 3B42 v7 data, which confirmed the robustness of the reconstruction model to generate realistic spatiotemporal characteristics, although the comparison with TRMM was hampered by TRMM overestimation in this region. In general, TRMM shows a positive precipitation bias around high-elevation areas (more than 5000 m asl), but over lower elevation Andean regions (between 3000 and 5000 m asl) the two products showed similar results.

Several important conceptual assumptions have to be considered to fully appreciate the high-resolution precipitation results obtained in this study: 1. The reconstruction is based on surface observations, i.e., in contrast to TRMM products, which are derived from infrared, thermal, and microwave spectral exitance from clouds, the NDVI-precipitation measurements are obtained from sensing dynamical land cover on the ground. 2. The NDVI field contains topographic information and variability, and this property is inherited to the precipitation reconstruction. 3. The precipitation variability used in the reconstruction process incorporates in situ weather station data. 4. The spatial resolution is constrained by the spatial scale of the NDVI data. 5. The 10 day temporal (dekadal) scale was selected here to maintain the same time scale as the NDVI data; and therefore, we did not take into consideration higher-frequency events in the analysis of our results. But since the NDVI data are only used as a "trend," the high-frequency information is retained by the local precipitation measurements from the weather stations (the "details" in equation (2) of the reconstruction process). Hence, it would be possible to extend the reconstruction process to a daily [*Quiroz et al.*, 2011; *Heidinger et al.*, 2012] or even higher temporal resolution, as long as precipitation data were available, for example, at hourly resolution. 6. A limitation of our reconstruction results is lack of a physical representation or definition of objects where vegetation is absent, as is the case on the QIC and other glaciated areas, lakes, bare ground, rocks, etc. Similarly, precipitation over the eastern Andean ridge and slope is not well represented by the

reconstruction. This limitation was to be expected since station data over the eastern Andean slope were not used in the reconstruction process, and since the vegetation-precipitation relationship breaks down at lower elevations over forested terrain. Here other types of (vegetation) indices and/or satellite data may be used in the future to obtain a more reliable representation. 7. As visualized in Table 3, the reconstruction does not show superior skill when compared to TRMM at capturing temporal variance. This is to be expected, given that the reconstruction is based on a vegetation-derived proxy, while TRMM includes information on rain rate from satellite and rain gauge measurements.

The spatial reconstruction model presented here may be extended to correct satellite products over other similar mountain regions where TRMM products may be questionable. In addition, our NDVI-derived product significantly improves the spatial resolution when compared to the much coarser spatial scale of many satellite products such as TRMM (about 0.25° or 27 km in the case of TRMM v7).

Our high-resolution precipitation product may also permit a better validation of numerical model results obtained from global and regional climate models, reanalysis, and satellite data. Such data sets may also prove valuable for forecasting and prediction studies or for developing climate change scenarios over mountain regions at a much finer resolution.

The main goal of this study was to generate high-quality regional precipitation information over a mountain region with a low density of weather stations. We acknowledge that several caveats exist, which have to be considered in more detail in future studies. In particular, the model could benefit from the inclusion of a denser surface station network or from a modification that would allow taking into account topography, land cover, soil moisture and type, and other important surface characteristics. Moreover, additional tests evaluating the reconstruction sensitivity to varying environmental boundary conditions (e.g., ENSO) would be desirable. Nonetheless, the high spatial resolution of our spatiotemporal reconstruction model makes our results appealing to a variety of applications over semiarid mountain regions, especially given is good performance compared to other spatial precipitation products, such as TRMM 3B42 v7.

**References**

Baraer, M., B. G. Mark, J. M. McKenzie, T. Condom, J. Bury, K. Huh, C. Portocarrero, J. Gomez, and S. Rathay (2012), Glacier recession and water resources in Peru's Cordillera Blanca, *J. Glaciol.*, *58*(207), 134–150, doi:10.3189/2012JoG11J186.

Blacutt, L. A., D. L. Herdies, L. G. De Concalves, D. A. Vila, and M. Andrade (2015), Precipitation comparison for the CFSR, MERRA, TRMM3B42 and combined scheme datasets in Bolivia, *Atmos. Res.*, *15*, 117–131, doi:10.1016/j.atmosres.2015.02.002.

Bookhagen, B., and M. R. Strecker (2008), Orographic barriers, high-resolution TRMM rainfall, and relief variations along the eastern Andes, *Geophys. Res. Lett.*, *35*, L06403, doi:10.1029/2007GL032011.

Bradley, R. S., F. T. Keimig, H. F. Diaz, and D. R. Hardy (2009), Recent changes in freezing level heights in the tropics with implications for the deglacierization of high mountain regions, *Geophys. Res. Lett.*, *36*, L17701, doi:10.1029/2009GL037712.

Casimiro, W. S. L., D. Labat, J.-L. Guyot, J. Ronchail, and J. J. Ordonez (2009), TRMM rainfall data estimation over the Peruvian Amazon-Andes basin and its assimilation into a monthly water balance model, Paper presented at New Approaches to Hydrological Prediction in Data-sparse Regions, Proceedings of Symposium HS.2 at the join IAHS & IAH Convention, Hyderabad, India.

Condom, T., P. Rau, and J. C. Espinoza (2011), Correction of TRMM 3B43 monthly precipitation data over the mountainous areas of Peru during the period 1998–2000, *Hydrol. Process.*, *25*, 1924–1933, doi:10.1002/hyp.7949.

Dinku, T., S. Chidzambwa, P. Ceccato, S. J. Connor, and C. F. Ropelewski (2008), Validation of high-resolution satellite rainfall products over complex terrain, *Int. J. Remote Sens.*, *29*(14), 4097–4110, doi:10.1080/01431160701772526.

Espinoza, J. C., S. Chavez, J. Ronchail, C. Junquas, K. Takahashi, and W. Lavado (2015), Rainfall hotspots over the southern tropical Andes: Spatial distribution, rainfall intensity, and relations with large-scale atmospheric circulation, *Water Resour. Res.*, *51*, 3459–3475, doi:10.1002/2014WR016273.

Garreaud, R. D. (2009), The Andes climate and weather, *Adv. Geosci.*, *22*, 3–11, doi:10.5194/adgeo-22-3-2009.

Garreaud, R. D., and P. Aceituno (2001), Interannual rainfall variability over the South American Altiplano, *J. Clim.*, *14*, 2779–2789.

Garreaud, R., M. Vuille, and A. Clement (2003), The climate of the Altiplano: Observed current conditions and mechanisms of past changes, *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, *194*, 5–22, doi:10.1016/S0031-0182(03)00269-4.

Giovannettone, J. P., and A. P. Barros (2009), Probing regional orographic controls of precipitation and cloudiness in the central Andes using satellite data, *J. Hydrometeorol.*, *10*, 167–182, doi:10.1175/2008JHM973.1.

Hanshaw, M. N., and B. Bookhagen (2014), Glacial areas, lake areas, and snow lines from 1975 to 2012: Status of the Cordillera Vilcanota, including the Quelccaya Ice Cap, northern central Andes, Peru, *Cryosphere*, *8*, 1–18, doi:10.5194/tc-8-359-2014.

Hardy, D. R., and S. P. Hardy (2008), White-winged Diuca Finch (*Diuca speculifera*) nesting on Quelccaya Ice Cap, Perú, *Wilson J. Ornithol.*, *120*(3), 613–617, doi:10.1676/06-165.1.

Heidinger, H., C. Yarleque, A. Posadas, and R. Quiroz (2012), TRMM rainfall correction over the Andean Plateau using wavelet multi-resolution analysis, *Int. J. Remote Sen.*, *33*(14), 4583–4602, doi:10.1080/01431161.2011.652315.

Hird, J. N., and G. J. McDermid (2009), Noise reduction of NDVI time series: An empirical comparison of selected techniques, *Remote Sens. Environ.*, *113*, 248–258, doi:10.1016/j.rse.2008.09.003.

Huffman, G. J., R. F. Adler, D. T. Bolvin, G. Gu, E. J. Nelkin, K. P. Bowman, Y. Hong, E. Stocker, and D. B. Wolff (2007), The TRMM multisatellite precipitation analysis: Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, *J. Hydrometeorol.*, *8*(1), 38–55, doi:10.1175/JHM560.1.

Hunink, J., W. Immerzeel, and P. Droogers (2014), A High-resolution Precipitation 2-step mapping Procedure (HiP2P): Development and application to a tropical mountainous area, *Remote Sens. Environ.*, *140*, 179–188, doi:10.1016/j.rse.2013.08.036.

Hurley, J. V., M. Vuille, D. R. Hardy, S. Burns, and L. G. Thompson (2015), Cold air incursions, $\delta^{18}$O variability and monsoon dynamics associated with snow days at Quelccaya Ice Cap, Peru, *J. Geophys. Res. Atmos.*, *120*, 7467–7487, doi:10.1029/2015JD023323.

Immerzeel, W. W., R. A. Quiroz, and S. M. De Jong (2005), Understanding precipitation patterns and land use interaction in Tibet using harmonic analysis of SPOT VGT-S10 NDVI time series, *Int. J. Remote Sens.*, *26*(11), 2281–2296, doi:10.1080/01431160512331326611.

L'Hote, Y., P. Chevalier, A. Coudrain, Y. Lejeune, and P. Etchevers (2005), Relationship between precipitation phase and air temperature: Comparison between the Bolivian Andes and the Swiss Alps, *Hydrol. Sci. J.*, *50*(6), 989–997, doi:10.1623/hysj.2005.50.6.989.

Lavado, C. W. S., J. Ronchail, D. Labat, J. C. Espinoza, and J. L. Guyot (2012), Basin-scale analysis of rainfall and runoff in Peru (1969–2004): Pacific, Titicaca and Amazonas drainages, *Hydrol. Sci. J.*, *57*(4), 1–18, doi:10.1080/02626667.2012.672985.

Lavado, C. W. S., D. Labat, J. Ronchail, J. C. Espinoza, and J. L. Guyot (2013), Trends in rainfall and temperature in the Peruvian Amazon-Andes basin over the last 40 years (1965–2007), *Hydrol. Process.*, *41*, 2944–2957, doi:10.1002/hyp.9418.

Lenters, J. D., and K. H. Cook (1999), Summertime precipitation variability over South America: Role of the large-scale circulation, *Mon. Weather Rev.*, *127*, 409–431, doi:10.1175/1520-0493(1999)127<0409:SPVOSA>2.0.CO;2.

Mantas, V. M., Z. Liu, C. Caro, and A. J. S. C. Pereira (2015), Validation of TRMM multi-satellite precipitation analysis (TMPA) products in the Peruvian Andes, *Atmos. Res.*, *163*, 132–145.

Manz, B., W. Buytaert, Z. Zulkafli, W. Lavado, B. Willems, L. A. Robles, and J.-P. Rodriguez-Sanchez (2016), High-resolution satellite-gauge merged precipitation climatologies of the tropical Andes, *J. Geophys. Res. Atmos.*, *121*, 1190–1207, doi:10.1002/2015JD023788.

Mohr, K. I., D. Slayback, and K. Yager (2014), Characteristics of precipitation features and annual rainfall during the TRMM era in the central Andes, *J. Clim.*, *27*, 3982–4001, doi:10.1175/JCLI-D-13-00592.1.

Mourre, L., T. Condom, C. Junquas, T. Lebel, J. E. Sicart, R. Figueroa, and A. Cochachin (2016), Spatio-temporal assessment of WRF, TRMM and in situ precipitation data in a tropical mountain environment (Cordillera Blanca, Peru), *Hydrol. Earth Syst. Sci.*, *20*, 125–141.

Posadas, A., E. L. A. Duffaut, C. Yarleque, M. Carbajal, H. Heidinger, L. Carvalho, C. Jones, and R. Quiroz (2015), Spatial random downscaling of rainfall signals in Andean heterogeneous terrain, *Nonlinear Process. Geophys.*, *22*, 383–402, doi:10.5194/npg-22-383-2015.

Quiroz, R., C. Yarleque, A. Posadas, V. Mares, and W. W. Immerzeel (2011), Improving daily rainfall estimation from NDVI using wavelet transform, *Environ. Modell. Software*, *26*(2), 201–209, doi:10.1016/j.envsoft.2010.07.006.

Rabatel, A., et al. (2013), Current state of glaciers in the tropical Andes: A multi-century perspective on glacier evolution and climate change, *Cryosphere*, *7*, 81–102, doi:10.5194/tc-7-81-2013.

Satgé, F., M.-P. Bonnet, M. Gosset, J. Molina, W. Hernan Yuque Lima, R. Pillco Zolá, F. Timouk, and J. Garnier (2015), Assessment of satellite rainfall products over the Andean plateau, *Atmos. Res.*, *167*, 1–14, doi:10.1016/j.atmosres.2015.07.012.

Schauwecker, S., et al. (2014), Climate trends and glacier retreat in the Cordillera Blanca, Peru, revisited, *Global Planet. Change*, *119*, 85–97, doi:10.1016/j.gloplacha.2014.05.005.

Scheel, M. L. M., M. Rohrer, C. Huggel, D. Santos Villar, E. Silvestre, and G. J. Huffman (2011), Evaluation of TRMM multi-satellite precipitation analysis (TMPA) performance in the Central Andes region and its dependency on spatial and temporal resolution, *Hydrol. Earth Syst. Sci.*, *15*, 2649–2663, doi:10.5194/hess-15-2649-2011.

Takahashi, K., and D. S. Battisti (2007), Processes controlling the mean tropical Pacific precipitation pattern. Part I: The Andes and the eastern Pacific ITCZ, *J. Clim.*, *20*, 3434–3451, doi:10.1175/JCLI4198.1.

Thompson, L. G., E. Mosley-Thompson, H. Brecher, M. Davis, B. Leon, D. Les, P. N. Lin, T. Mashiotta, and K. Mountain (2006), Abrupt tropical climate change: Past and present, *Proc. Natl. Acad. Sci. U.S.A.*, *103*(28), 10,536–10,543, doi:10.1073/pnas.0603900103.

Thompson, L. G., E. Mosley-Thompson, M. E. Davis, V. S. Zagorodnov, I. M. Howat, V. N. Mikhalenko, and P. N. Lin (2013), Annually resolved ice core records of tropical climate variability over the past ~1800 years, *Science*, *340*(6135), 945–950, doi:10.1126/science.1234210.

Vuille, M. (1999), Atmospheric circulation over the Bolivian Altiplano during dry and wet periods and extreme phases of the Southern Oscillation, *Int. J. Climatol.*, *19*, 1579–1600, doi:10.1002/(SICI)1097-0088(19991130)19:14<1579::AID-JOC441>3.0.CO;2-N.

Vuille, M., and C. Ammann (1997), Regional snowfall patterns in the high arid Andes, *Clim. Change*, *36*, 413–423, doi:10.1023/A:1005330802974.

Vuille, M., and R. S. Bradley (2000), Mean annual temperature trends and their vertical structure in the tropical Andes, *Geophys. Res. Lett.*, *27*, 3885–3888, doi:10.1029/2000GL011871.

Vuille, M., and F. Keimig (2004), Interannual variability of summertime convective cloudiness and precipitation in the central Andes derived from ISCCP-B3 data, *J. Clim.*, *17*, 3334–3348, doi:10.1175/1520-0442(2004)017<3334:IVOSCC>2.0.CO;2.

Vuille, M., and M. Werner (2005), Stable isotopes in precipitation recording South American summer monsoon and ENSO variability—Observations and model results, *Clim. Dyn.*, *25*, 401–413, doi:10.1007/s00382-005-0049-9.

Vuille, M., R. S. Bradley, and F. Keimig (2000), Interannual climate variability in the Central Andes and its relation to tropical Pacific and Atlantic forcing, *J. Geophys. Res.*, *105*, 12,447–12,460, doi:10.1029/2000JD900134.

Vuille, M., B. Francou, P. Wagnon, I. Juen, G. Kaser, B. G. Mark, and R. S. Bradley (2008a), Climate change and tropical Andean glaciers—Past, present and future, *Earth Sci. Rev.*, *89*, 79–96, doi:10.1016/j.earscirev.2008.04.002.

Vuille, M., G. Kaser, and I. Juen (2008b), Glacier mass balance variability in the Cordillera Blanca, Peru and its relationship with climate and the large-scale circulation, *Global Planet. Change*, *62*(1–2), 14–28, doi:10.1016/j.gloplacha.2007.11.003.

Vuille, M., E. Franquist, R. Garreaud, W. Lavado, and B. Caceres (2015), Impact of the global warming hiatus on Andean temperature, *J. Geophys. Res. Atmos.*, *120*, 3745–3757, doi:10.1002/2015JD023126.

Wilks, D. S. (2011), *Statistical Methods in the Atmospheric Sciences*, *Int. Geophysics Series*, vol. 100, Academic Press, San Diego, Calif.

Zubieta, R., A. Getirana, J. C. Espinoza, and W. Lavado (2015), Impacts of satellite-based precipitation datasets on rainfall-runoff modeling of the western Amazon basin of Peru and Ecuador, *J. Hydrol.*, *528*, 599–612.

Zulkafli, Z., W. Buytaert, C. Onof, B. Manz, E. Tarnavsky, W. Lavado, and J.-L. Guyot (2014), A comparative performance analysis of TRMM 3B42 (TMPA) versions 6 and 7 for hydrological applications over Andean–Amazon river basins, *J. Hydrometeorol.*, *15*, 581–592, doi:10.1175/JHM-D-13-094.1.