Spatiotemporal patterns and driving factors of lake size variability since 1985, Central Andean Plateau, South America

Master Thesis

in partial fulfillment of the requirements for the

degree of Master of Science (M.Sc.) in Geosciences

Nicolas Werner

Matriculation Number: 753274

University of Potsdam

Faculty of Science

Institute of Earth and Environmental Science

Advisor:	Prof. Dr. Bodo Bookhagen
Co-Advisor:	Dr. Aljoscha Rheinwalt

Table of Contents

List of Abbreviations

AMO	Atlantic Multidecadal Oscillation
BEST	Bivariate ENSO-Index
CA	Central Andes
CAP	Central Andean Plateau
CFSR	Climate Forecast System Reanalysis
DJF	December-January-February
ENSO	El Nino Southern Oscillation
HCA	Hierarchical Clustering Analysis
ITCZ	Intertropical Convergence Zone
JJA	June-July-August
MAM	March-April-May
MEI	Multivariate ENSO-index
MERRA	Modern-Era Retrospective Analysis for Research and Applications
MNDWI	Modified Normalized Difference Water Index
NCEP	National Center for Environmental Prediction
NCAR	National Center for Atmospheric Research
NDWI	Normalized Difference Water Index
NE	North-East
NOAA	National Oceanic and Atmospheric Administration
ONI	Oceanic Nino Index
PDO	Pacific Decadal Oscillation
SACZ	South Atlantic Convergence Zone / South American Converge Zone
SALLJ	South American Low Level Jet
SE	South-East
SI	Supplementary Information
SON	September-October-November
SRTM	Shuttle Radar Topography Mission
SW	South-West
SWIR	Short-Wave Infrared
TM	Thematic Mapper
TNA	Tropical Northern Atlantic
TOA	Top-of-Atmosphere
TRMM	Tropical Rainfall Measurement Mission
TSA	Tropical Southern Atlantic
ZAP	Plateau der Zentralanden

Zusammenfassung

Veränderungen von Seegrößen geben Auskunft über Klimavariationen in abgelegenen Hochgebirgsregionen, von denen es nur wenige meteorologische Daten gibt. Seegrößenvariationen indizieren saisonal und jährliche bis dekadische Klimadynamiken, indem sie Niederschlagsereignisse über Einzugsgebiete integrieren. In dieser Arbeit haben wir Zeitreihen von 1058 Seen des Plateaus der Zentralanden (ZAP) in Südamerika untersucht. Wir nutzten die hierarchische Clusteranalyse um Zeitreihen ähnlicher Dynamik zu gruppieren und repräsentative Zeitreihen für die Gruppen abzuleiten. Weiterhin verglichen wir die erzeugten Cluster mit Klimaindizes, die die Klimavariabilität im Atlantik und im Pazifik widerspiegeln.

Unsere Ergebnisse zeigen, dass dekadische Klimadynamiken auf ZAP vielseitig sind. Trotz ähnlicher Höhe und Größe der Einzugsgebiete existieren mehrere eindeutige Gruppen gleichen Verhaltens. Die hierarchische Clusteranalyse unterteilte das ZAP in 7 eindeutige Regionen gleicher klimatischer Variabilität: Das nördliche, nordwestliche und das südliche nördliche ZAP (Altiplano), die südwestliche und das nordöstliche südliche ZAP (Puna) und zwei Gebiete am nordöstlichen Rand des Altiplano. Wir fanden keinen eindeutigen Einfluss des Atlantiks auf die Variabilität der Seegrößen des ZAP, wobei es Hinweise auf einen dekadischen Einfluss gibt. Stattdessen sind Seegrößenvariationen auf jährlichen Skalen von 3-5 Jahren deutlich von Veränderungen im Pazifik geprägt mit einem deutlichen Einfluss der ENSO-Zyklen. Dabei zeigen sich feuchte (trockene) Jahre vorwiegend zu La Nina (El Nino) Episoden, verursacht durch vorherrschende Ostwinde (Westwinde) während La Nina (El Nino). Vorwiegend negative dekadische Trends von Seegrößen, aber auch positive Trends, die durch Gletscherschmelzen verursacht werden, zeigen einen deutlich Einfluss des Klimawandels auf das ZAP in den letzten Dekaden.

Spatiotemporal dynamics of lake areas on the Central Andean Plateau since 1985

Autors: N. Werner, B. Bookhagen, A. Rheinwalt

Abstract

Changes in surface-lake areas have large potential to provide information on climatic variability in remote and mountainous areas, where meteorological station data are sparse. Lake-area changes record seasonal and inter-annual to decadal climatic dynamics by integrating daily rainfall events over catchment scales. In this study we analyzed time series of 1058 lakes on the Central Andean Plateau (CAP) in South America. In order to generate representative time series and to group lakes with synchronous behavior, we used hierarchical clustering and compared the lake-area clusters to climate indices characterizing climate variability of the Atlantic or the Pacific.

Our results indicate that decadal climate dynamics on the CAP are diverse and that there exist specific groups of similar behavior despite similar altitude and catchment area of lakes. The hierarchical clustering analysis subdivided the CAP into 7 distinct zones of similar climatic variability that represent the northern, northwestern and southern northern Andean Plateau (Altiplano), the southwestern and northeastern southern Andean Plateau (Puna) and two zones on the northeastern boundary of the Altiplano. We did not observe a clear relationship of lake-size variability on the CAP and the climate dynamics of the Atlantic Ocean, although there is evidence for a decadal-scale influence. Instead, we observe that changes in lake areas on time scales of 3-5 years are controlled by climatic indices representing the variability of the Pacific Ocean with a dominant influence of the ENSO cycles as wet years show a tendency to coincide with La Nina and dry years with El Nino due to easterly (westerly) moisture transport during La Nina (El Nino). Prevalent negative decadal trends of lake-areas, but also positive trends that seem to be caused by glacial melting highlight a pronounced impact of climatic changes on the CAP within the last decades.

1. Introduction

The assessment of climate change in remote and mountainous areas is complicated, as there exists limited ground-control data and numerical modeling studies often rely on weakly constrained boundary conditions. However it has been argued that high-elevation regions experience rapid climate changes and mountain regions have been coined the sentinel of changes (e.g. Urrutia and Vuille, 2008; Rangwala and Miller, 2012; Diaz et al., 2014; Mountain Research Initiative EDW Working Group, 2015; Vuille et al., 2015; Ragettli et al., 2016). Our study focuses on the Central Andes (CA) of South America with the orogenic Central Andean Plateau (CAP) that have been argued to experience rapid climate changes during the past decades.

Several authors noted an increase in temperatures in the CAP within the last century (e.g. Bradley et al., 2006; Vuille et al., 2003; Bradley et al, 2006; Casimiro et al., 2013; Seiler et al., 2013; Vuille et al., 2015; López-Moreno et al., 2016) that was accompanied by a less pronounced decrease in precipitation, especially in southern Peru and northern Bolivia (Vuille, 2003). Climate models predict a further warming within the next century (Christensen et al., 2007; Urrutia and Vuille, 2009; Thibeault et al., 2010), with temperatures rising up to 5 degrees (Vuille et al., 2008). The effects of a projected warming are severe on the population (Bush et al., 2010) due to aridification and a consecutive water scarcity (Vergara, 2007; Rangecroft et al., 2013; Seiler et al., 2013). Precipitation is assumed to occur less frequently, but more heavily and despite the uncertainty in precipitation prediction, any increases in total rainfall will not counteract the rise in temperatures (Thibeault et al., 2010; Thibeault et al., 2012). However, the assessment and the prediction of climate change in the CA rely on sparsely distributed meteorological stations (e.g. López-Moreno et al., 2016), as well as low resolution satellite and reanalysis data, such as NCEP/NCAR, MERRA and CFSR (e.g. Garreaud et al., 2003; Carvalho et al., 2012; Eichler and Londoño, 2013; Blacutt et al., 2015). Due to the heterogeneous topography and wide gaps within the station density, the interpolation of ground observation data is prone to errors. The assessment of climatic variables from reanalysis data, especially from precipitation, is also subject to biases. MERRA and CFSR products for example overestimate rainfall in the Altiplano (Carvalho et al., 2012; Blacutt et al., 2015). Within topographically complex terrains such as the Andes there are always difficulties related to reanalysis data (Eichler and Londoño, 2013) or sparse station data (Bennett et al., 2016). Datasets from the Tropical Rainfall Measurement Mission (TRMM) have been used for precipitation estimates in the CA (e.g. Bookhagen and Strecker, 2008; Bookhagen and Strecker, 2012; Romantschke and Houze, 2013; Mohr et al., 2014; Castino et al., 2016). While they are able to describe the large-scale rainfall

picture (Boers et al., 2015), they may underestimate rainfall over the CA (Scheel et al., 2011; Espinoza et al., 2015; Blacutt et al., 2015). Furthermore the resolution of reanalysis and TRMM products cannot resolve regional differences in precipitation.

The monitoring of surficial manifestations of climate variability, such as glaciers and lakes, can complement sparse ground observations. Both features are indicative of changes in temperature or precipitation and overcome problems related to point-wise measurements as they integrate over catchments. They are sensible enough to resolve intra-seasonal variations, yet stable enough to sustain single precipitation events. Although glaciers and their changes in the CA have been studied extensively (e.g. Rabatel et al., 2013; Hanshaw and Bookhagen, 2014, Schauwecker et al., 2014, Veettil et al., 2016), they are almost absent in the interior of the CAP (Bookhagen, 2016). In contrast lakes can be found throughout the CAP and provide for CAP-wide information on climatic changes. While larger lakes such as Lake Titicaca in Bolivia or Lake Poopó in Bolivia have been studied previously (e.g. Garreaud and Aceituno, 2001; Abarca-Del-Rio et al., 2012) and are under constant observation, there is no study of all lakes in the CAP, despite their fundamental importance as a source of drinking water, for agriculture and the mining industry (e.g. Vergara, 2007; Michelutti et al., 2015). The effects of climate change on the central Andean lakes are still uncertain (Michelutti et al., 2015). The dramatic shifts of lake extents are best illustrated by the decline of Lake Poopó, that used to be Bolivia's second largest lake with an area of more than 3500 km² in 1986 but completely vanished in the beginning of 2016.

Satellite sensors provide images of the Earth's surface for several decades and allow for the continuous monitoring of lakes. Numerous methods exist to classify water in satellite images, the most prominent being band ratios such as the Normalized Differential Water Index (NDWI - McFeeters, 1996) or the Modified Normalized Differential Water Index (MNDWI - Xu, 2006). Both use the ratio of the green spectral range and the near-infrared or short-infrared range, respectively. In contrast to other land classes, water reflects the green portion and strongly absorbs the infrared portion of the sun's radiation and therefore can be differentiated with these ratios.

In order to investigate the temporal variations of lake sizes and the spatial differences within the CAP we rely on multi-spectral Landsat imagery from 1985 to 2011 to classify all lakes of the CAP and to derive their corresponding time series of areal changes. Weighted linear regressions were used to decipher general trends of every lake and we grouped the data to subdivide the CAP into regions with coherent lake-area time series characterizing different climatic regions. We further analyzed the spatiotemporal patterns of the lake area variability and the influence of large scale oscillations as determinants of lake changes.

2. Geographic and climatic setting

The Central Andean Plateau (CAP) is encompassed by the eastern and western Cordilleran ridges of the Central Andes and is separated into the northern CAP, referred to as the Altiplano and the southern CAP, the Puna de Atacama, referred to as the Puna. With its extension from 14°S to 29°S (Fig. 1) it forms the second largest orographic plateau after the Himalaya (Allmendinger et al., 1997). The climate of the CAP is semi-arid to arid determined not only by its mean elevation of about 4000m but also by the prevention of moisture influx by the circumjacent Cordilleran ridges. Wet episodes occur mainly during austral summer, when moisture is transported from the continental lowlands by easterly winds, with dominantly dry phases prevailing throughout the rest of the year (e.g. Garreaud, 1999; Marwan et al., 2003; Garreaud et al., 2003; Rohrmann et al., 2014; Bennett et al., 2016). Those phases are linked to the South American Monsoon System (e.g. Liebmann and Mechoso, 2011; Marengo et al., 2012; Silva and Kousky, 2012), which is one of the main features of the South American climate. Intensified easterly trade winds from the tropical Atlantic (Zhou and Lau, 1998) converge in front of the Andes and the southward component, also referred to as the South American low-level jet (SALLJ), brings moisture into the subtropics along the Andean mountain range (e.g. Ferreira et al., 2003; Vera et al., 2006; Liebmann and Mechoso, 2011). A southward shift of the Inter Tropical Convergence Zone (ITCZ) during austral summer (e.g. Garreaud et al., 2008; Silva and Kousky, 2012) results in widespread deep convection over the Amazon basin that extends towards southeastern South America as an elongated band along the South American Convergence Zone (SACZ; e.g. Carvalho et al., 2004). As a dynamical response to enhanced convection over Amazonia an upper tropospheric anticyclone, known as the Bolivian High, forms, with its strength being mainly determined by precipitation in the CA (Lenters and Cook, 1997; Lenters and Cook, 1999). The intensity and the position of the Bolivian High strongly regulate moisture inflow from the continental lowlands into the CAP with an easterly/wet, westerly/dry pattern (Lenters and Cook, 1999; Vuille, 1999; Garreaud et al., 2003; Neukom et al., 2015). Intraseasonal variabilities of the SAMS mainly result from Rossby wave trains from the Southern Pacific. The El-Nino Southern Oscillation (ENSO) plays a major role on the inter-annual climate variabilities, especially in the CAP (e.g. Vuille, 1999; Vuille et al., 2000; Garreaud et al., 2003; Placzek et al., 2009; Morales et al., 2015; Bennett et al., 2016). The warming of the Eastern Pacific during positive ENSO phases (El Nino) weakens the Pacific trade winds and leads to anomalous uprising over the Pacific and anomalous subsidence over Amazonia through Walker cell connections. This subsidence hinders deep convection over the South American tropics and thereby weakens and

shifts the Bolivian High northward. The associated warming of the troposphere over the Pacific also induces higher temperatures over South America, reducing the meridional temperature gradient whereby westerlies are dominant over most of South America. Those westerlies at lower and upper levels, as well as the weakening and the shift of the Bolivian High strongly restrain moisture transport into the CAP during El Nino periods. During La Nina this pattern reverses and easterly flow and therefore wetter conditions over the CAP persist (e.g. Garreaud and Aceituno, 2001; Garreaud et al., 2003; Marwan et al., 2003; Strecker et al., 2007; Garreaud, 2009; Bookhagen and Strecker, 2010). The enhanced heating of the CAP during El Nino episodes initiates a monsoon return flow that strengthens the southern Atlantic trade winds within the following summer, thus triggering increased precipitation in the year after an El Nino (Vuille, 1999; Vuille et al., 2000). Temperatures vary up to 2°C between both ENSO phases, with a warming over the CAP during El Nino and a cooling during La Nina (e.g. Garreaud, 2009)

3. Datasets

3.1 Multi-spectral satellite imagery and elevation data

We used a total of 3365 level-1G processed images of the Thematic Mapper (TM) multi-spectral instrument provided by the Landsat 4 and Landsat 5 satellites from 1985 to 2011. The study region is subdivided into 22 scenes, with an average of 153 images per scene (Fig. S1 in the SI material). Due to geometric distortions and georeferencing errors of the level-1G Landsat products, some of the images of the initial image stack had to be deleted. All Landsat images have been transformed to top-of-atmosphere (TOA) reflectance (Chander and Markham, 2003). This conversion compensates for varying satellite sensor parameters, e.g. the sensor-specific gain or the angle of the solar irradiation, which cause intensity deviations of the same target in multiple images (Nelson, 1985). Additionally we used a 30m Digital Elevation Model from the Shuttle Radar Topography Mission (SRTM), acquired within February 2000, to compute catchment area and slope data.

4. Methods

4.1. Lake classification

We used the first shortwave-infrared (SWIR) channel of the Landsat TM (band 5) instead of a band ratio in the classification process. The NDWI is strongly influenced by sediment, algae and solubles within the water and the MNDWI is prone to misclassifications in the direct vicinity of the lakes, e.g. wet sediment. Every SWIR-image is turned into a binary image with water and non-water regions by using an empirically chosen threshold (see SI material for additional information). A slope map was used additionally to account and automatically remove false classifications of mostly cloud and mountain shadows that cover the same spectral range as water.

We perform multiple iterations to achieve the best classification result: First, we loop trough all images to classify water bodies from every image and cumulatively add them to an initial water mask. Second, this initial water mask, along with multi-spectral satellite images, is used to manually select lakes of interest and to create a reference lake mask. In this step, small ponds, transient water bodies, river segments and falsely classified mountain and cloud shadows are removed. Some lakes have been manually edited, where nearby hillslope shadows or glaciers and snow have been falsely classified as part of the respective lakes. The reference lake mask gives an estimation of where lakes have been present within the last 30 years, but does not give insights on the exact lake extents at every date covered by the Landsat images. Third, the classification of water bodies in the image stack is repeated but for areas outlined by the reference lake mask. Every classified object is labeled according to a corresponding index in the reference water mask. The number of pixels of every object is converted to an area for every time step.

Because the Landsat images covering the entire CAP are constructed from 22 path/row combinations (see SI material), the dates of image acquisition differ. We therefore resampled all lake-area time series to a common grid and binned all values into 3-month periods that are based on seasonality (DJF: December, January, February; MAM: March, April, May; JJA: June, July, August; SON: September, October, November). We assigned the mean value for every bin and no value if the bin does no contain data values.

4.2. Hierarchical clustering analysis (HCA)

The time series have been clustered to aggregate time series with a similar coevolution. This process also works as a quality control by indicating outliers and errors within the dataset. Using clusters in the climatic analysis furthermore gives a more regional picture in the climatological context. We used HCA (e.g. Jain et al., 1999; Manning et al., 2008; Rokach and Maimon, 2010; Murtagh and Contreras, 2012) to combine lake time series with a similar coevolution. Our methodology follows the approach described in Rheinwalt et al. (2015). In our study similarity is expressed as the pairwise correlation of the lake area time series. As a distance (d) metric (e.g. Murtagh and Contreras, 2012) the arcos of the Pearson correlation coefficient r is used. The more different (similar) time series are the higher (lower) is the angle in vector space and the distance between measurements. Correlation of the time series has been calculated pairwise, whereby rows of a pair with no value have been ignored. Additionally lakes have been deleted, where more than 80 percent of the 3-month smoothing window are zero or have no value. The average linkage method was chosen for the clustering which uses the average distance of two clusters as the distance metric in the clustering process. The complete linkage criterion is more sensitive to outliers and uncertainties regarding the classification and previous smoothing. The single linkage criterion is not suitable as it shows chaining effects and creates very large clusters (Manning et al., 2008). As the average linkage criterion used the average distance between clusters to assign them, it is not guaranteed that all lakes within one cluster correlate well with each other lake from the same cluster. There is a chance of outlier lakes that fit best into a certain cluster, but show weak correlations with its other lakes. We therefore calculated the mean correlation coefficients for the lakes of a cluster and removed the lakes that fall below 2σ (see SI material for a detailed description). The output of the HCA is a dendrogram that represents the stepwise aggradation of time series pairs and the corresponding distances (see SI material for additional information). It can be cut at different thresholds depending on either the intended average distance (d) or the attempted cluster number.

5. Results

5.1. Linear regression analysis

We derived time series from 1058 lakes within the CAP and performed weighted linear regressions on all time series to obtain decadal trends of lake-size (Figure 3a). Significant negative trends prevail

throughout the CAP. Positive significant trends mostly occur in glaciated areas or in areas close to the Amazonian moisture source. Those are areas near glaciers or regions close to the continental lowlands, where moisture is coming from. A large spread within the linear regression coefficients in the northern Altiplano reflects in the large variance of elevations in that region (Fig. 3b and 3c). The central and southern Altiplano show only slight changes in elevation and reveal a low variance in the regression coefficients. At the transition from the Altiplano to the Puna (northern to southern CAP), there is a strong excess in negative regression coefficients (Fig. 3b) that is equally, but inversely, pronounced in the elevations (Fig. 3c). The same pattern is visible in the southern part of the Puna.

5.2. Hierarchical clustering analysis

We have chosen thee clustering thresholds to generate different cluster number through the CAP (Fig. 5) At a distance of d_1=1.5, there are two clusters that spatially subdivide the CAP into a northern part ($C_1$2) that represents the Northern Altiplano and a southern part ($C_1$1) that represents the southern Altiplano and the Puna Plateau (Fig. 5a). The elevation profiles (Fig 5d) of both clusters only slightly deviate from the CAP elevation histograms, showing a bimodal distribution with its peaks at ~3900m and ~4500m. Reducing the threshold to d_2=1.35 results in 3 clusters. The northern Altiplano ($C_2$3) remains unchanged while the southern CAP is divided into the southern Altiplano/northeastern Puna ($C_2$1) and the southwestern Puna ($C_2$2, Fig. 5b). The elevation histograms of the first two clusters (Fig. 5e) again resemble the elevation histogram of the CAP, but the third cluster $C_2$3 forms a single, broad distribution mostly covering lower elevations. As a last threshold we use a distance of d_3=1.08 to obtain 24 clusters, of which 17 have been removed as they include less than 10 lakes and have a large spatial spread.

With the lowest threshold we generate 7 major clusters that segment the CAP (Fig. 5c). The northern Altiplano is divided into four clusters. Cluster 5 ($C_3$5) represents the northernmost sector of the Altiplano and Cluster 4 ($C_3$4) the southern parts of the northern Altiplano. Cluster 6 ($C_3$6) and Cluster 7 ($C_3$7) represents highly elevated lakes at the eastern margin of the northern CAP. The southern Altiplano is represented by Cluster 1 ($C_3$1). The Puna Plateau is divided into a northeastern part, Cluster 2 ($C_3$2), and a southwestern part, Cluster 3 ($C_3$3). A more detailed description of the clusters and its lakes is given in Table 1. The further segregation of the southern Altiplano and the northwestern Puna ($C_2$1), as well as the northern Altiplano ($C_2$3) with the third threshold d_3 not only reflects in the distribution of the lake locations but also evident in the elevation histograms (Fig. 5f).

5.3. Constructing mean cluster time series

The clustering analysis strongly reduces the dimensionality of the lake area dataset and we derived 7 distinct regions of synchronous lake-size variability. As a representative time series for every cluster we use the mean of all 3-month mean time series within a cluster (Fig. 6), whereas we rly only on time steps with more than 10 measurement. In order to account for the large range of lake sizes (1km² to >1000 km²) we standardized all time series to z-scores and refer to the resulting time series as mean cluster time series.

The mean cluster time series show a pronounced inter-seasonal variability with the maximum lake areas occurring during austral summer and autumn (DJF/MAM), expect for $C_3 3$, where the largest lakes sizes occur in austral winter (JJA). The inter-seasonal variability is most pronounced in the southern Altiplano ($C_3 1$), where rainfall rapidly accumulates in flat lake basins and forms shallow water layers over large areas. These flat lakes are not persistent and most of the water evaporates equally rapid within a few weeks to months as is best illustrated by the Salar de Uyuni (Fig. S9 in the SI material). The sharp and narrow peaks of $C_3 1$ are apparent in the mean time series of the southwestern Puna ($C_3 3$) as well. The lakes around the Lake Titicaca ($C_3 4$) and in the northeastern Puna ($C_3 2$) show broader peaks of lower magnitudes. The data density of the remaining clusters ($C_3 5$, $C_3 6$ and $C_3 7$) does not sufficiently resolve inter-seasonal differences.

On inter-annual timescales large differences can be observed within and among the clusters. There are substantial differences in annual lake sizes with low lake-area increases in e.g. 1990 or 1998 and large lake-area increases in e.g. 2000 or 2006. This becomes especially obvious in the positive differences of the mean cluster difference time series (Fig. S6 in the SI material). The difference time series represent the differences of the mean cluster time series measurements to their relative last measurements (see SI material for more information). Some years show opposing wet and dry years among the clusters, e.g. 1987 or 2008, which have been anomalously wet years in $C_3 1$, but rather dry years in $C_3 4$. The succeeding years, 1988 and 2009, have been wetter in $C_3 4$ and distinctively drier in $C_3 1$. Nevertheless the majority of years are similar, so that strong or weak lake size increases of one cluster are reflected by the other. As mentioned before $C_3 3$ shows little inter-annual variability and its variations do not generally reflect in the other clusters. The lowest lake sizes of $C_3 3$ however occurred in 1997/1998 that is also the time of lake size minima in $C_3 1$ and $C_3 2$. In contrast the lowest lake areas in $C_3 4$, $C_3 6$ and $C_3 7$ occurred in 1991/1992. The highest lake levels observed within the study period are not uniform throughout the clusters, however all clusters, except for $C_3 7$, show

high lake levels in the mid to late 1980s followed by a continuous decline until the mid 1990s with another common high phase at 1999/2001.

We also performed weighted linear regressions on the mean cluster time series (Table S4 in the SI material) in order to find linear trends in the data. The strongest observed trend can be found in $C_3 2$ with a significant negative linear trend (at 95% significance level) throughout the study period. A similar but less clear and not significant decline is observable in $C_3 1$ and $C_3 5$. The lakes of $C_3 4$ do not show a linear trend but a rather decadal-like oscillation with high lake sizes phases in the mid-1980s or the early 2000s and low lake size episodes in the early to mid-1990s and at least partially in the early 2010s. The largest lake area measured during the study period occurred in 2006 and not in the 1980s, when the largest lake sizes of the southern Altiplano and the northwestern Puna have been observed. The clusters $C_3 6$ and $C_3 7$ on the contrary show positive trends, whereby the positive trend of $C_3 6$ is significant at the 95% level of significance.

5.4. Climate indices

In order to find relationships between lake size variability and atmospheric oscillations, especially regarding the influence of changes in the Pacific and Atlantic Oceans, we performed cross-correlations with various climate indices (Table S3 in the SI material). However, the cross correlation did not yield conclusive results (the detailed analysis and the cross-correlation results are given in the SI material). For the visual and qualitative assessment of apparent influences of the El Nino Southern Oscillation (ENSO), we compared the total annual lake growth with ENSO phases (Fig. 7) as determined by the National Oceanic and Atmospheric Administration (NOAA) with the Oceanic Nino Index (ONI). Following this classification, ENSO is subdivided into strong and moderate El Nino events (ONI>1 and 0.5<ONI<1, respectively), strong and moderate La Nina phases (ONI<-1 and -1<ONI<-0.5, respectively), as well as neutral episodes (-0.5<ONI<0.5). The annual lake growth is represented by the annual lake growth time series and corresponds to the annual sums of the positive differences (DJF(0) to DJF(+1)) of mean cluster difference time series.

Most of the maximum annual lake area increases in $C_3 1$, $C_3 2$, $C_3 4$ and $C_3 5$ occurred during negative ENSO modes, e.g. 1991, 2001 or 2006, or in the year after an El Nino, e.g. 1993. However there are also El Nino years with large lake increases (1987) and La Nina years with small lake increases (1989). In $C_3 3$ the largest annual growths occurred equally during El Nino or La Nina episodes. In $C_3 6$ and $C_3 7$ the largest annual lake increases occurred during El Nino years.

To quantitatively assess the ENSO influence we cross-correlated the mean cluster difference time series (see SI material) with the MEI and obtained significant negative correlations for the clusters C_31, C_32, C_34 and C_35, but not for the remaining clusters (Fig. S8 in the SI material). Significance was tested with a random permutation model at the 95% significance level. The maximum correlation coefficients occur at a lag of 2 months for C_31 and C_34 (r = -0.34 and r = -0.42, respectively) and at a lag of 4 months for C_32 and C_35 (r = -0.36 and r= -0.52, respectively). Cross-correlations of the mean cluster difference time series with climate indices of Atlantic sea surface temperature (e.g. TSA (Tropical South Atlantic Index), TNA (Tropical North Atlantic Index) and AMO (Atlantic Multidecadal Oscillation Index) did not show significant correlations.

6. Discussion

6.1. Spatial patterns

The clustering analysis allowed us to segment the CAP into 7 distinct regions of spatial variability. The N-S climatic gradient (e.g. Garreaud, 2009; Placzek et al., 2009) seems to represent the dominant factor on lake size variability, while the elevation plays a major role in the further segregation of the main clusters. There is no clear evidence in the HCA for the previously described E-W precipitation gradient (e.g. Vuille and Keimig, 2004) in the CAP with larger annual precipitation amounts in the eastern sectors of the plateau due to moisture from the east. However, the clustering of the CAP with correlation coefficients is based on similarities in the temporal coevolution and not on differences in magnitudes. The spatial pattern of the 7 clusters we obtained from the lowest threshold at d_3 resembles SE-striking bands from north to south so that the main mode of spatial variability arranges rather along a NE-SW gradient. Those bands show similarities to spatial patterns identified by Vuille and Keimig (2004), who found three coherent modes of variability that we could further refine with the HCA. The NE-SW gradient also corresponds to the northeasterly source of moisture from the Amazon lowlands (e.g. Garreaud et al., 2003). A reason for this gradient is given by a combination of the source of moisture and the geometry of the Andean belt with the broadest width of the eastern Cordillera and the Central Andes in general at the height of the southern Altiplano and roughly at the transition of C_34 and C_31, separating the northern and southern Altiplano, respectively. The subdivision of the CAP is of particular interest for paleoclimate analyses (e.g. Placzek et al., 2006; Placzek et al., 2009; review in Baker and Fritz, 2015) when samples of different regions within the plateau region are compared (Vuille and Keimig, 2004).

6.2. Inter-seasonal variability

The general behavior of inter-seasonal variability is similar for all clusters, despite differences in the slope of periods of lake level decreases that are rather steep in the case of C_31 or C_33 and shallower in C_32 or C_34. C_33 is special in a sense that the largest lake areas do not occur in austral summer (DJF) as is the case for the other clusters but in austral winter (JJA). As the southern-most sector of the CAP, moisture from the eastern and north-eastern regions in the continental foreland rarely reaches this region. It is rather moisture from the Pacific that is responsible for precipitation events in that area (Placzek et al., 2009). This wintertime rainfall might be related to northward extending fronts of a midlatitude precipitation band related to the westerlies of the southern Hemisphere (Vuille, 1999; Strecker et al., 2007; Placzek et al., 2009; Boers et al., 2016).

6.3. Inter-annual variability

The strong inter-annual variability in C_31 and C_33 can be related to the characteristics of lakes in those regions. Rainfall rapidly accumulates in flat lake basins and forms shallow water layers over large areas. These flat lakes are not persistent and most of the water evaporates equally rapid within a few weeks to months as is best illustrated by the Salar de Uyuni in C_31. Lakes in C_32 and C_34 seem to be more stable throughout the study period that might be in the case of C_32 due to higher elevations and lower evaporation rates and in the case of C_34 due to higher levels of moisture in that region compared to other areas of the CAP. The latter is not only caused by the evaporation of Lake Titicaca and the corresponding atmospheric moistening, but also to the configuration of the eastern Cordillera in the northern Altiplano that allows for upslope transport of foreland moisture despite prevailing westerly flow above the CA (Falvey and Garreaud, 2005). The differences of the lakes in C_34 and C_31 are especially visible comparing Lake Titicaca (C_34), a large freshwater lake with a depth of 281m and Lake Poopó (C_31), a large saline and shallow lake with a depth of only few meters. This further amplifies the northeastern source of moisture and the NE-SW gradient (Garreaud, 2009). Some studies however presented evidence that moisture from the subtropical plains to the SE of the CAP also contribute to rainfall in the southern parts of the Altiplano and the Puna via the Chaco low-level jet and cold front surges from the south (e.g. Placzek, 2009; Rohrmann et al., 2014; Boers et al., 2016). Opposing wet and dry phases (e.g. 1987 or 2008) in the northern and southern Altiplano might indicate an additional southeasterly source of moisture in our data. Additionally the Puna (C_32 and C_33) shows no anomalies in the annual lake growth time series during the strong El Nino of

1997/1998, despite the preceding, anomalous lake-area decrease due to El Nino related temperature increases. While humidity in the Amazonian lowlands to the NE of the CAP is not important for rainfall variability over the CAP (Garreaud et al., 2003), moisture levels in southeastern South America strongly control the intensity of precipitation that reaches the southern CAP from the SE (Quade et al., 2008). The SALLJ that brings moisture to the subtropical plains of southeastern South America was particularly strong during the El Nino of 1997/1998 (Ferreira et al., 2003) and thus might be an explanation for non anomalous lake-area increases during that year compared to the anomalously low increases in the Altiplano.

6.4. Decadal trends

The linear patterns of temporal variability in lake areas provide further insights on the characteristics, the behavior and the trends of the clusters and the CAP. The clusters $C_3 6$ and $C_3 7$ show positive trends that are in consideration of the warming of the Central Andes (e.g. Vuille et al., 2015) rather unexpected. However, the location of the corresponding lakes close to the moist continental foreland and their proximity to glaciers of the Central Andes justify lake area increases throughout the study period. This further suggests that most of these lakes are proglacial lakes and that the decline of most Andean glaciers (e.g. Vuille et al., 2008; Rabatel et al., 2013) reflects in the growth of those lakes. Despite the distinctive decline of lakes in $C_3 4$ from the 1980s towards the 1990s and during the end of the study period, the lake sizes of $C_3 4$ remained rather stable. That can be, as described beforehand, attributed to the structure of the eastern central Andes and the evaporation of Lake Titicaca (Falvey and Garreaud, 2005). Due to the large volume of water within the Titicaca basin, the Lake Titicaca is more stable than flat salt lakes in the southern Altiplano and the southward transport of water by the Rio Desaguadero provides water to the lakes in the southern and western vicinity of Lake Titicaca. However severe declines in the water budget of Lake Titicaca will strongly affect the stability of the surrounding lakes. In the northernmost parts of the Altiplano the lakes of $C_3 5$ show a slight areal decrease since 1985 that could be related to a smaller influence of moisture from the Lake Titicaca and to less moisture from the east reaching that area. This image is similar in the southern Altiplano, where flat and unstable salt lakes prevail that react very sensitive to climatic changes. Indeed the lakes of $C_3 1$ did not return to the large areas that emerged during the 1980s, except for some years of short-term lake size highs (e.g. 2001 or 2006). This is indicative of less moisture entering the southern parts of the Altiplano compared to $C_3 4$. The northeastern parts of the Puna plateau show a significant decline throughout the timeframe of the study that is also

strongest among the clusters. This decline and the high elevations of lakes in $C_3 2$ (Fig. 3f) suggest a more pronounced susceptibility of lakes at greater altitude. Despite the strong negative decadal trends observed in $C_3 2$, the annual lake growth (Fig. 7) does not noticeably drop within the study period and we hypothesize that the continuous decrease of lake areas in the northeastern Puna ($C_3 2$) is not primarily induced by lower precipitation, but rather by increasing temperatures. It is especially evident in 2001 or 2010 that strong or frequent rainfalls are capable of bringing lake sizes to the level of the 1980s, when lake areas in this region were at its maximum.

6.5. Influence of regional-scale circulations

The inter-annual variability is mostly controlled by large-scale circulations and as shown in previous studies it is linked to the ENSO cycles (e.g. e.g. Vuille, 1999; Vuille et al., 2000; Garreaud et al., 2003). Nonetheless we could not find a clear influence of El Nino and La Nina on lake size variability in the cross-correlation analysis of the mean cluster time series with climate indices (see SI material). However the role of ENSO is apparent with regard to the mean cluster difference time series (see SI material) and the derived annual lake growth time series and a cross-correlation of the difference time series and the MEI yields significant negative correlations in the inner parts of the Altiplano and the southwestern Puna. The other clusters do not show a clear relationship with ENSO. In the case of $C_3 3$ this could be related to different afore mentioned mechanisms of rainfall that are rather related to the Pacific. The clusters $C_3 6$ and $C_3 7$ are located at the eastern border of the CAP where moisture contents are higher compared to other regions of the CAP due to enhanced upslope transport of moisture through the dissected eastern Cordillera (Falvey and Garreaud, 2005). This upslope transport also prevails during an El Nino related westerly flow regime over the CAP. The significant negative correlations of $C_3 1$, $C_3 2$, $C_3 4$ and $C_3 5$ further support the La Nina/wet, El Nino/dry pattern (e.g. Garreaud and Aceituno, 2001; Garreaud et al., 2003). Despite the conditions after an El Nino year, the following year tends to be wetter due to a monsoon return flow and a strengthening of the Atlantic trade winds in the next summer (e.g. Vuille et al., 2000). This can be seen in all clusters except for $C_3 3$ in 1992/1993 after a strong El Nino in 1991/1992 and also in 1998/1999, 2005/2006 and 2010/2011, which all exhibited a La Nina signal. It is therefore not clear if the large lake size increases during those years are related due to the monsoon return flow or the La Nina conditions. In contrast the year after a moderate to strong La Nina event (e.g. 1999/2000 or 2008/2009) seems to be drier (Vuille et al., 2000). If there are three consecutive years of prevailing La Nina conditions, the third year shows large lake area increases (e.g. 2000/2001). The heating of

the eastern Pacific during El Nino heats the overlying atmosphere and thus temperature increases are observed over the CAP during El Nino (e.g. Vuille et al., 2000). That is especially evident in 1997/1998, where lake areas in C_31, C_32 and C_33 dropped to their lowest point in the study period coinciding with the strongest El Nino in the last 30 years. Even though there is a tight coupling of the Pacific Decadal Oscillation (PDO) to the ENSO cycles we found no distinct relationship of lake size variability and the PDO from the cross-correlation analysis of the PDO-index and the annual lake growth time series. As other studies before (e.g. Garreaud et al., 2003), we also found no conclusive influence of the northern and southern Atlantic on lake size variability despite different mechanisms that are capable of putting changes of central Andean lakes and the Atlantic into relation. The continuous warming of the northern Atlantic and the steady increase of the AMO within the last decades might shift the ITCZ and convection over the Amazon basin northward (e.g. Broccoli et al., 2006; Knight et al., 2006), thus rendering especially the southern parts of the CAP drier. The decadal decline of lake sizes within C_31 and C_32 supports this theory. Furthermore changes in the tropical Atlantic temperatures and also the temperature gradient between the northern and southern Atlantic strongly control the strength of the trade winds and therefore how much moisture is transported into the subtropics of South America. The Chaco region to the SE of the CAP that substantially contributes to the rainfall of the Puna and the southern Altiplano shows a large variability in moisture levels that are related to the trade wind strengths and the intensity and mode of the SALLJ (Vera et al., 2006). Additionally there have been described several teleconnections between the Atlantic and the Pacific Oceans (e.g. Giannini et al., 2001; Kayano et al., 2011; Kayano et al., 2013; Kayano and Capistrano, 2014). As we did not find a clear influence of the Atlantic, any apparent relationships are either not recognizable with out dataset or act on decadal timescales. This may be also the reason why we found no association of lake area changes and the PDO. Garreaud et al. (2003) found that moisture availability in the continental foreland of the CAP shows little variability and that changes in the wind regimes control moisture transport into the plateau region. Hence on inter-annual timescales, changes in the strength and the position of the Bolivian High that are closely tied to variations in the ENSO cycles seem to be the main driver of lake size variability. Nonetheless any decadal-scale oscillation like the AMO or the PDO might still contribute to the long-term signal.

7. Conclusion

In this study we semi-automatically extracted lakes and their corresponding areas from Landsat satellite images from 1985 to 2011 for the Central Andean Plateau (CAP) in South America. We generated time series of lake-area changes for 1058 lakes and we clustered the data in order to segment the CAP into regions of similar climatic variability. We make the following three key observations:

First, hierarchical clustering analysis subdivided the CAP into 7 coherent climatic regions: The northern and northwestern Altiplano, the upper northeastern Altiplano and the lower northeastern Altiplano, the southern Altiplano and the northeastern and southwestern Puna.

Second, linear regression analyses of single lake time series and the mean cluster time series indicate mainly negative trends throughout the inner parts of the plateau region, especially in the southern Altiplano (northern CAP) and the northern Puna (southern CAP). Negative trends seem to be stronger at higher elevations as is evident in the case of the northeastern Puna. Positive trends (i.e. increasing lake areas) are observed at the plateau boundary of the northern Altiplano and are associated to glacial melting and the proximity to the Amazonian moisture.

Third, analyses of lake size variability and climate indices indicates no clear relationship of lake area changes with variations in the Atlantic (TSA, AMO) or the PDO. However, decadal-scale influences of the AMO or the PDO are possible, but cannot be verified explicitly with our data. Instead we found negative significant correlations of annual lake growth time series with the MEI that supports the general view of wetter (drier) conditions during La Nina (El Nino) due to prevailing easterly (westerly) winds associated with both ENSO cycles. We therefore conclude that ENSO is the dominant driver of inter-annual variability in most parts of the CAP.

This study is the first to analyze all lakes within the CAP on decadal timescales and our results shed new light on the spatial and temporal climatic variability of the CAP. Prevalent decadal trends of the central Andean lakes display pronounced impacts of climatic changes on water resources of the CAP. The data from our study not only improves the understanding of past climatic variations in that region but also presents prospects on how climatic changes may affect the CAP in the future.

List of References

Abarca-Del-Rio, R., Crétaux, J.-F., Berge-Nguyen, M., Maisongrande, P., 2012. Does Lake Titicaca still control the Lake Poopó system water levels? An investigation using satellite altimetry and MODIS data (2000–2009). Remote Sens. Lett. 3, 707–714. doi:10.1080/01431161.2012.667884

Allmendinger, R.W., Jordan, T.E., Kay, S.M., Isacks, B.L., 1997. the Evolution of the Altiplano-Puna Plateau of the Central Andes. Annu. Rev. Earth Planet. Sci. 25, 139–174. doi:DOI: 10.1146/annurev.earth.25.1.139

Baker, P.A., Fritz, S.C., 2015. Nature and causes of Quaternary climate variation of tropical South America. Quat. Sci. Rev. 124, 31–47. doi:10.1016/j.quascirev.2015.06.011

Bennett, M., New, M., Marino, J., Sillero-Zubiri, C., 2016. Climate complexity in the Central Andes: A study case on empirically-based local variations in the Dry Puna. J. Arid Environ. 128, 40–49. doi:10.1016/j.jaridenv.2016.01.004

Blacutt, L.A., Herdies, D.L., de Gonçalves, L.G.G., Vila, D.A., Andrade, M., 2015. Precipitation comparison for the CFSR, MERRA, TRMM3B42 and Combined Scheme datasets in Bolivia. Atmos. Res. 163, 117–131. doi:10.1016/j.atmosres.2015.02.002

Boers, N., Barbosa, H.M.J., Bookhagen, B., Marengo, J.A., Marwan, N., Kurths, J., 2015. Propagation of Strong Rainfall Events from Southeastern South America to the Central Andes. J. Clim. 28, 7641–7658. doi:10.1175/JCLI-D-15-0137.1

Boers, N., Bookhagen, B., Marwan, N., Kurths, J., 2016. Spatiotemporal characteristics and synchronization of extreme rainfall in South America with focus on the Andes Mountain range. Clim. Dyn. 46, 601–617. doi:10.1007/s00382-015-2601-6

Bookhagen, B., Strecker, M.R., 2008. Orographic barriers, high-resolution TRMM rainfall, and relief variations along the eastern Andes. Geophys. Res. Lett. 35, 1–6. doi:10.1029/2007GL032011

Bookhagen, B., Strecker, M.R., 2012. Spatiotemporal trends in erosion rates across a pronounced rainfall gradient: Examples from the southern Central Andes. Earth Planet. Sci. Lett. 327–328, 97–110. doi:10.1016/j.epsl.2012.02.005

Bookhagen, B., 2016. Glaciers and Monsoon Systems. Monsoons Clim. Chang. Obs. Model. 225–249. doi:10.1007/978-3-319-21650-8_11

Bradley, R.S., Vuille, M., Diaz, H.F., Vergara, W., 2006. Threats to Water Supplies in the Tropical Andes. Science (80-.). 312, 1755–1756.

Broccoli, A.J., Dahl, K.A., Stouffer, R.J., 2006. Response of the ITCZ to Northern Hemisphere cooling. Geophys. Res. Lett. 33, 1–4. doi:10.1029/2005GL024546

Bush, M.B., Hanselman, J.A., Gosling, W.D., 2010. Nonlinear climate change and Andean feedbacks: An imminent turning point? Glob. Chang. Biol. 16, 3223–3232. doi:10.1111/j.1365-2486.2010.02203.x

Carvalho, L.M. V, Jones, C., Liebmann, B., 2004. The South Atlantic convergence zone: Intensity, form, persistence, and relationships with intraseasonal to interannual activity and extreme rainfall. J. Clim. 17, 88–108. doi:10.1175/1520-0442(2004)017<0088:TSACZI>2.0.CO;2

Carvalho, L.M. V, Jones, C., Posadas, A.N.D., Quiroz, R., Bookhagen, B., Liebmann, B., 2012. Precipitation characteristics of the South American monsoon system derived from multiple datasets. J. Clim. 25, 4600–4620. doi:10.1175/JCLI-D-11-00335.1

Castino, F., Bookhagen, B., Strecker, M.R., 2016. Rainfall variability and trends of the past six decades (1950-2014) in the subtropical NW Argentine Andes. Clim. Dyn. 1–19. doi:10.1007/s00382-016-3127-2

Chander, G., Markham, B., 2003. Revised Landsat-5 TM Radiometrie Calibration Procedures and Postcalibration Dynamic Ranges. IEEE Trans. Geosci. Remote Sens. 41, 2674–2677. doi:10.1109/TGRS.2003.818464

Christensen, J.H., Hewitson, B., Busuioc, A., Chen, A., Gao, X., Held, I., Jones, R., Kolli, R.K., Kwon, W.-T., Laprise, R., Rueda, V.M., Mearns, L., Menéndez, C.G., Räisänen, J., Rinke, A., Sarr, A., Whetton, P., 2007. Regional Climate Projections. Clim. Chang. 2007 Phys. Sci. Basis. Contrib. Work. Gr. I to Fourth Assess. Rep. Intergov. Panel Clim. Chang. 847–940. doi:10.1080/07341510601092191

Diaz, H.F., Bradley, R.S., Ning, L., 2014. Climatic Changes in Mountain Regions of the American Cordillera and the Tropics: Historical Changes and Future Outlook. Arctic, Antarct. Alp. Res. 46, 735–743. doi:10.1657/1938-4246-46.4.735

Eichler, T.P., Londoño, A.C., 2013. South American climatology and impacts of El Niño in NCEP's CFSR data. Adv. Meteorol. 2013. doi:10.1155/2013/492630

Espinoza, J. C., S. Chavez, J. Ronchail, C. Junquas, K. Takahashi, and W.L., 2015. Rainfall hotspots over the southern tropical Andes: Spatial distribution, rainfall intensity, and relations with large-scale atmospheric circulation. Water Resour. Res. 51, 506–523. doi:10.1016/0022-1694(68)90080-2

Falvey, M., Garreaud, R.D., 2005. Moisture variability over the South American Altiplano during the South American low level jet experiment (SALLJEX) observing season. J. Geophys. Res. Atmos. 110, 1–12. doi:10.1029/2005JD006152

Ferreira, R.N., Rickenbach, T.M., Herdies, D.L., Carvalho, L.M. V., 2003. Variability of South American Convective Cloud Systems and Tropospheric Circulation during January–March 1998 and 1999. Mon. Weather Rev. 131, 961–973. doi:10.1175/1520-0493(2003)131<0961:VOSACC>2.0.CO;2

Garreaud, R., 1999. Multiscale Analysis of the Summertime Precipitation over the Central Andes. Mon. Weather Rev. 127, 901–921. doi:10.1175/1520-0493(1999)127<0901:MAOTSP>2.0.CO;2

Garreaud, R., Aceituno, P., 2001. Interannual Rainfall Variability over the South American Altiplano. J. Clim. 14, 2779–2789. doi:10.1175/1520-0442(2001)014<2779:IRVOTS>2.0.CO;2

Garreaud, R., Vuille, M., Clement, A.C., 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

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

Garreaud, R.D., Vuille, M., Compagnucci, R., Marengo, J., 2009. Present-day South American climate. Palaeogeogr. Palaeoclimatol. Palaeoecol. 281, 180–195. doi:10.1016/j.palaeo.2007.10.032

Giannini, A., Chiang, J.C.H., Cane, M.A., Kushnir, Y., Seager, R., 2001. The ENSO teleconnection to the Tropical Atlantic Ocean: Contributions of the remote and local SSTs to rainfall variability in the Tropical Americas. J. Clim. 14, 4530–4544. doi:10.1175/1520-0442(2001)014<4530:tetttt>2.0.co;2

Hanshaw, M.N., Bookhagen, B., 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. Cryosph. 8, 1–18. doi:10.5194/tc-8-1-2014

Jain, a. K., Murty, M.N., Flynn, P.J., 1999. Data clustering: a review. ACM Comput. Surv. 31, 264–323. doi:10.1145/331499.331504

Kayano, M.T., Andreoli, R. V., Ferreira de Souza, R.A., 2013. Relations between ENSO and the South Atlantic SST modes and their effects on the South American rainfall. Int. J. Climatol. 33, 2008–2023. doi:10.1002/joc.3569

Kayano, M.T., Capistrano, V.B., 2014. How the Atlantic multidecadal oscillation (AMO) modifies the ENSO influence on the South American rainfall. Int. J. Climatol. 34, 162–178. doi:10.1002/joc.3674

Knight, J.R., Folland, C.K., Scaife, A.A., 2006. Climate impacts of the Atlantic multidecadal oscillation. Geophys. Res. Lett. 33, 1–4. doi:10.1029/2006GL026242

Lavado Casimiro, W.S., Labat, D., Ronchail, J., Espinoza, J.C., Guyot, J.L., 2013. Trends in rainfall and temperature in the Peruvian Amazon-Andes basin over the last 40years (1965-2007). Hydrol. Process. 27, 2944–2957. doi:10.1002/hyp.9418

Lenters, J.D., Cook, K.H., 1997. On the Origin of the Bolivian High and Related Circulation Features of the South American Climate. J. Atmos. Sci. 54, 656–678. doi:10.1175/1520-0469(1997)054<0656:OTOOTB>2.0.CO;2

Lenters, J.D., Cook, K.H., 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

Liebmann, B., Mechoso, C., 2011. The South American Monsoon System. Glob. Monsoon Syst. Res. Forecast (2nd Ed. 137–157.

López-Moreno, J.I., Morán-Tejeda, E., Vicente-Serrano, S.M., Bazo, J., Azorin-Molina, C., Revuelto, J., Sánchez-Lorenzo, A., Navarro-Serrano, F., Aguilar, E., Chura, O., 2016. Recent temperature variability and change in the Altiplano of Bolivia and Peru. Int. J. Climatol. 36, 1773–1796. doi:10.1002/joc.4459

Manning, C.D., Raghavan, P., Schütze, H., 2009. Hierarchical clustering. Introd. to Inf. Retr. 377–401. doi:10.1017/CBO9780511809071.017

Marengo, J.A., Liebmann, B., Grimm, A.M., Misra, V., Silva Dias, P.L., Cavalcanti, I.F.A., Carvalho, L.M. V, Berbery, E.H., Ambrizzi, T., Vera, C.S., Saulo, A.C., Nogues-Paegle, J., Zipser, E., Seth, A., Alves, L.M., 2012. Recent developments on the South American monsoon system. Int. J. Climatol. 32, 1–21. doi:10.1002/joc.2254

Marwan, N., Trauth, M.H., Vuille, M., Kurths, J., 2003. Comparing modern and Pleistocene ENSO-like influences in NW Argentina using nonlinear time series analysis methods. Clim. Dyn. 21, 317–326. doi:10.1007/s00382-003-0335-3

McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. Int. J. Remote Sens. 17, 1425–1432. doi:10.1080/01431169608948714

Michelutti, N., Wolfe, A.P., Cooke, C.A., Hobbs, W.O., Vuille, M., Smol, J.P., 2015. Climate change forces new ecological states in tropical Andean lakes. PLoS One 10, 1–10. doi:10.1371/journal.pone.0115338

Mohr, K.I., Slayback, D., Yager, K., 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

Morales, M.S., Carilla, J., Grau, H.R., Villalba, R., 2015. Multi-century lake area changes in the Southern Altiplano: A tree-ring-based reconstruction. Clim. Past 11, 1139–1152. doi:10.5194/cp-11-1139-2015

Mountain Research Initiative EDW Working Group, 2015. Elevation-dependent warming in mountain regions of the world. Nat. Clim. Chang. 5, 424–430. doi:10.1038/nclimate2563\rhttp://www.nature.com/nclimate/journal/v5/n5/abs/nclimate2563.html#supplementary-information

Murtagh, F., Contreras, P., 2012. Algorithms for hierarchical clustering: An overview. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 2, 86–97. doi:10.1002/widm.53

Nelson, R.F., 1985. Sensor-induced temporal variability of Landsat MSS data. Remote Sens. Environ. 18, 35–48. doi:10.1016/0034-4257(85)90036-7

Neukom, R., Rohrer, M., Calanca, P., Salzmann, N., Huggel, C., Acuña, D., Christie, D.A., Morales, M.S., 2015. Facing unprecedented drying of the Central Andes? Precipitation variability over the period AD 1000–2100. Environ. Res. Lett. Environ. Res. Lett 10, 84017–84017. doi:10.1088/1748-9326/10/8/084017

Placzek, C., Quade, J., Betancourt, J.L., Patchett, P.J., Rech, J.A., Latorre, C., Matmon, A., Holmgren, C., English, N.B., 2009. CLIMATE IN THE DRY CENTRAL ANDES OVER GEOLOGIC, MILLENNIAL, AND INTERANNUAL TIMESCALES [1]. Ann. Missouri Bot. Gard. 96, 386–397. doi:10.3417/2008019

Placzek, C., Quade, J., Patchett, P.J., 2006. Geochronology and stratigraphy of late Pleistocene lake cycles on the southern Bolivian Altiplano: Implications for causes of tropical climate change. Bull. Geol. Soc. Am. 118, 515–532. doi:10.1130/B25770.1

Quade, J., Rech, J.A., Betancourt, J.L., Latorre, C., Quade, B., Rylander, K.A., Fisher, T., 2008. Paleowetlands and regional climate change in the central Atacama Desert, northern Chile. Quat. Res. 69, 343–360. doi:10.1016/j.yqres.2008.01.003

Rabatel, A., Francou, B., Soruco, A., Gomez, J., Cáceres, B., Ceballos, J.L., Basantes, R., Vuille, M., Sicart, J.E., Huggel, C., Scheel, M., Lejeune, Y., Arnaud, Y., Collet, M., Condom, T., Consoli, G., Favier, V., Jomelli, V., Galarraga, R., Ginot, P., Maisincho, L., Mendoza, J., Ménégoz, M., Ramirez, E., Ribstein, P., Suarez, W., Villacis, M., Wagnon, P., 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

Ragettli, S., Immerzeel, W.W., Pellicciotti, F., 2016. Contrasting climate change impact on river flows from high-altitude catchments in the Himalayan and Andes Mountains. Proc. Natl. Acad. Sci. U. S. A. 113, 9222–7. doi:10.1073/pnas.1606526113

Rangecroft, S., Harrison, S., Anderson, K., Magrath, J., Castel, A.P., Pacheco, P., 2013. Climate change and water resources in arid mountains: An example from the bolivian andes. Ambio 42, 852–863. doi:10.1007/s13280-013-0430-6

Rangwala, I., Miller, J.R., 2012. Climate change in mountains: A review of elevation-dependent warming and its possible causes. Clim. Change 114, 527–547. doi:10.1007/s10584-012-0419-3

Rheinwalt, A., Goswami, B., Boers, N., Heitzig, J., Marwan, N., Krishnan, R., Kurths, J., 2015. Teleconnections in Climate Networks: A Network-of-Networks Approach to Investigate the Influence of Sea Surface Temperature Variability on Monsoon Systems, in: Lakshmanan, V., Gilleland, E., McGovern, A., Tingley, M. (Eds.), Machine Learning and Data Mining Approaches to Climate Science: Proceedings of the 4th International Workshop on Climate Informatics. Springer International Publishing, Cham, pp. 23–33. doi:10.1007/978-3-319-17220-0_3

Rohrmann, A., Strecker, M.R., Bookhagen, B., Mulch, A., Sachse, D., Pingel, H., Alonso, R.N., Schildgen, T.F., Montero, C., 2014. Can stable isotopes ride out the storms? The role of convection for water isotopes in models, records, and paleoaltimetry studies in the central Andes. Earth Planet. Sci. Lett. 407, 187–195. doi:10.1016/j.epsl.2014.09.021

Rokach, L., Maimon, O., 2010. Chapter 15— Clustering methods. Data Min. Knowl. Discov. Handb. 32. doi:10.1007/0-387-25465-X_15

Romatschke, U., Houze, R. a., 2013. Characteristics of Precipitating Convective Systems Accounting for the Summer Rainfall of Tropical and Subtropical South America. J. Hydrometeorol. 120927133830006. doi:10.1175/JHM-D-12-060.1

Schauwecker, S., Rohrer, M., Acuña, D., Cochachin, A., Dávila, L., Frey, H., Giráldez, C., Gómez, J., Huggel, C., Jacques-Coper, M., Loarte, E., Salzmann, N., Vuille, M., 2014. Climate trends and glacier retreat in the Cordillera Blanca, Peru, revisited. Glob. Planet. Change 119, 85–97. doi:10.1016/j.gloplacha.2014.05.005

Scheel, M.L.M., Rohrer, M., Huggel, C., Santos Villar, D., Silvestre, E., Huffman, G.J., 2010. 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. Discuss. 7, 8545–8586. doi:10.5194/hessd-7-8545-2010

Seiler, C., Hutjes, R.W.A., Kabat, P., 2013. Likely ranges of climate change in Bolivia. J. Appl. Meteorol. Climatol. 52, 1303–1317. doi:10.1175/JAMC-D-12-0224.1

Silva, V.B.S., Kousky, V.E., 2012. The South American Monsoon System : Climatology and Variability. Mod. Climatol. 398.

Strecker, M.R., Alonso, R.N., Bookhagen, B., Carrapa, B., Hilley, G.E., Sobel, E.R., Trauth, M.H., 2007. Tectonics and Climate of the Southern Central Andes. Annu. Rev. Earth Planet. Sci. 35, 747–787. doi:10.1146/annurev.earth.35.031306.140158

Thibeault, J.M., Seth, A., Garcia, M., 2010. Changing climate in the Bolivian Altiplano: CMIP3 projections for temperature and precipitation extremes. J. Geophys. Res. Atmos. 115, 1–18. doi:10.1029/2009JD012718

Thibeault, J., Seth, A., Wang, G., 2012. Mechanisms of summertime precipitation variability in the bolivian altiplano: Present and future. Int. J. Climatol. 32, 2033–2041. doi:10.1002/joc.2424

Toshie Kayano, M., Valéria Andreoli, R., Ferreira de Souza, R.A., 2011. Evolving anomalous SST patterns leading to ENSO extremes: Relations between the tropical Pacific and Atlantic Oceans and the influence on the South American rainfall. Int. J. Climatol. 31, 1119–1134. doi:10.1002/joc.2135

Urrutia, R., Vuille, M., 2009. Climate change projections for the tropical Andes using a regional climate model: Temperature and precipitation simulations for the end of the 21st century. J. Geophys. Res. Atmos. 114, 1–15. doi:10.1029/2008JD011021

Veettil, B.K., Bremer, U.F., de Souza, S.F., Maier, É.L.B., Simões, J.C., 2016. Influence of ENSO and PDO on mountain glaciers in the outer tropics: case studies in Bolivia. Theor. Appl. Climatol. 125, 757–768. doi:10.1007/s00704-015-1545-4

Vera, C., Baez, J., Douglas, M., Emmanuel, C.B., Marengo, J., Meitin, J., Nicolini, M., Nogues-Paegle, J., Paegle, J., Penalba, O., Salio, P., Saulo, C., Silva Dias, M.A., Silva Dias, P., Zipser, E., 2006. The South American low-level jet experiment. Bull. Am. Meteorol. Soc. 87, 63–77. doi:10.1175/BAMS-87-1-63

Vergara, W., Deeb, A.M., Valencia, A.M., Bradley, R.S., Francou, B., Zarzar, A., Grünwaldt, A., Haeussling, S.M., 2007. Economic Impacts of Rapid Glacier Retreat in the Andes. Eos, Trans. Am. Geophys. Union 88, 261. doi:10.1029/2007EO250001

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., Bradley, R., Keimig, F., 2000. Climate Variability in the Andes of Ecuador and Its Relation to Tropical Pacific and Atlantic Sea Surface Temperature Anomalies. J. Clim. 13, 2520–2535. doi:10.1175/1520-0442(2000)013<2520:CVITAO>2.0.CO;2

Vuille, M., Bradley, R., Werner, M., Keimig, F., 2003. 20th century climate change in the tropical Andes: observations and model results. Clim. Change 59, 75–99.

Vuille, M., Keimig, F., 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., Francou, B., Wagnon, P., Juen, I., Kaser, G., Mark, B.G., Bradley, R.S., 2008. Climate change and tropical Andean glaciers: Past, present and future. Earth-Science Rev. 89, 79–96. doi:10.1016/j.earscirev.2008.04.002

Vuille, M., Franquist, E., Garreaud, R., Lavado Casimiro, W.S., Cáceres, B., 2015. Impact of the global warming hiatus on Andean temperature. J. Geophys. Res. Atmos. 120, 3745–3757.

Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int. J. Remote Sens. 27, 3025–3033. doi:10.1080/01431160600589179

Zhou, J., Lau, K.M., 1998. Does a monsoon climate exist over South America? J. Clim. 11, 1020–1040. doi:10.1175/1520-0442(1998)011<1020:DAMCEO>2.0.CO;2

List of Figures
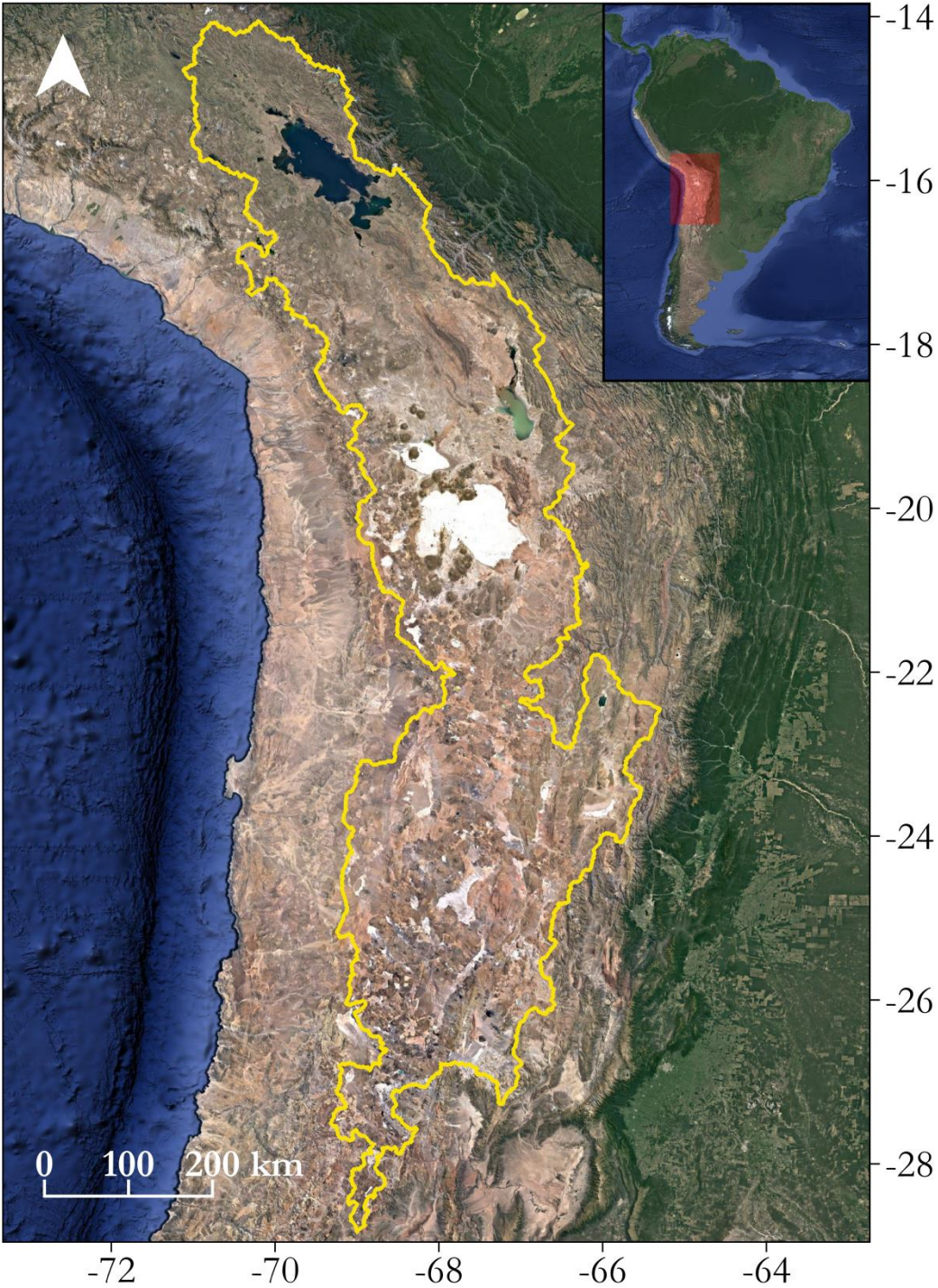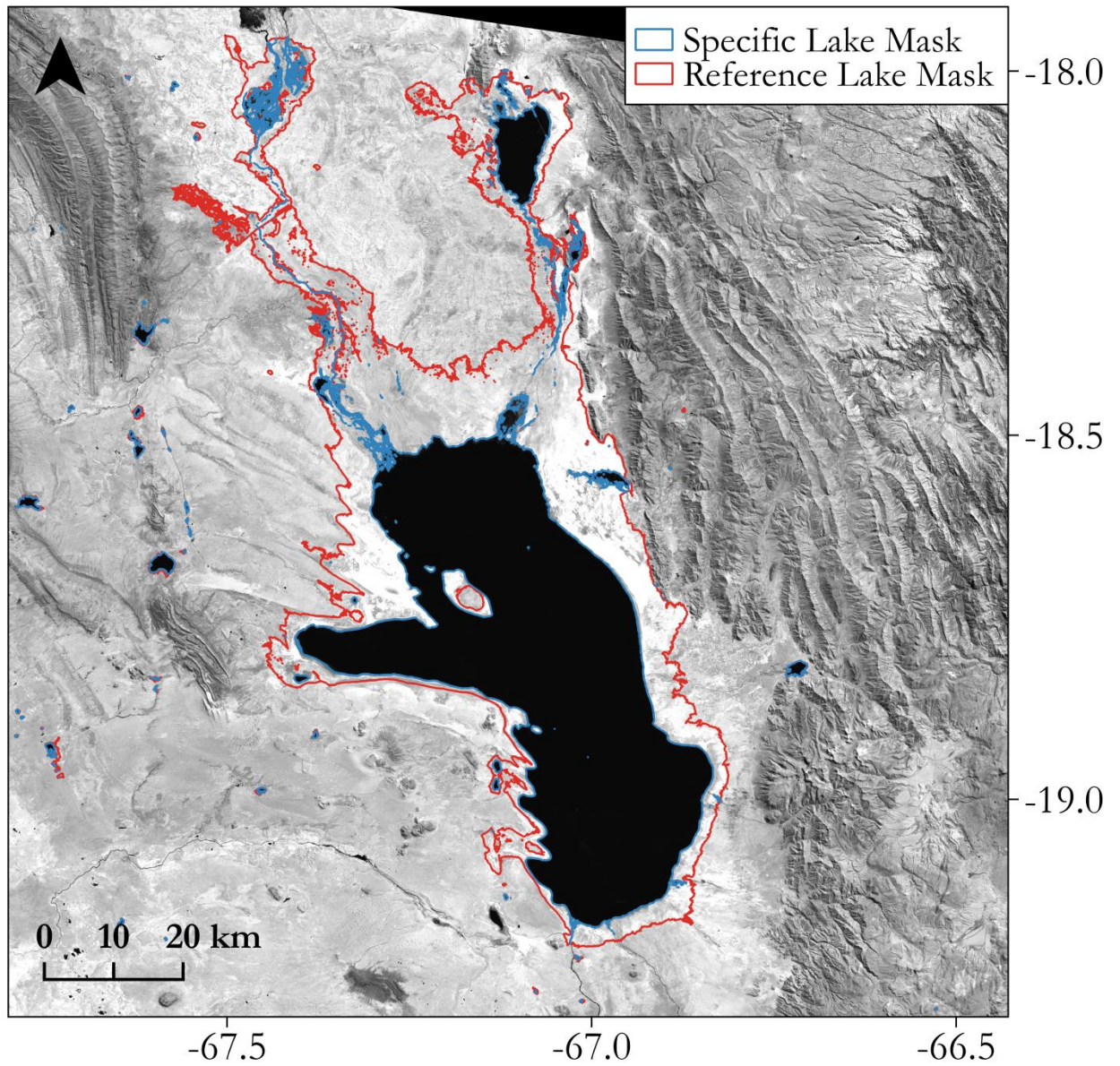
Fig. 2:   Landsat image from 14.March of 1999 showing the Lake Poopó (in this study combined with lake Uru-Uru in the north) in Bolivia and the products from the semi-automatic classification process: A reference lake mask (red) that represents the maximum lake extents from 1985 to 2011 and a specific lake mask (blue) that shows the lake extents at the date of image acquisition (1999).
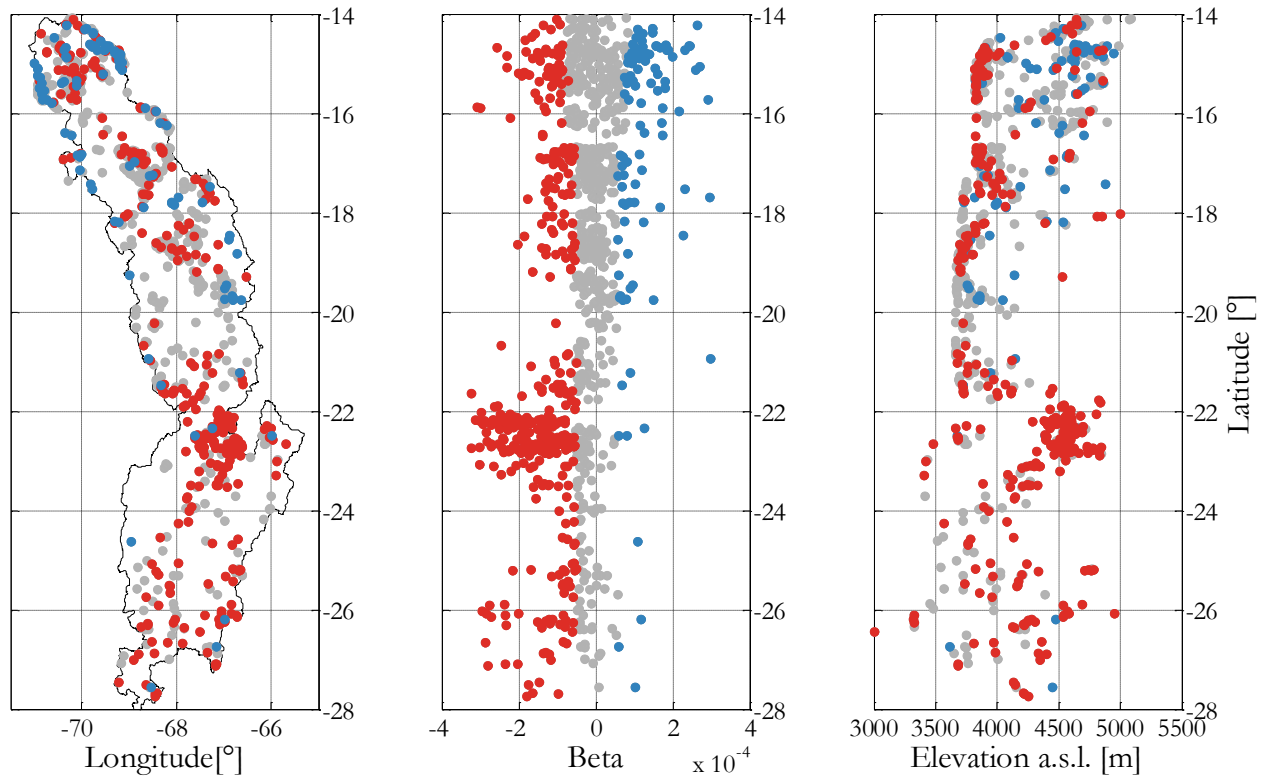
Fig. 3: Overview of weighted linear regression coefficients (Beta) and 95% level of significance (red – negative significant, blue – positive significant, grey – not significant). (a) Spatial distribution of lake centroids, (b) linear regression coefficients versus Latitude, (c) linear regression coefficients versus elevation
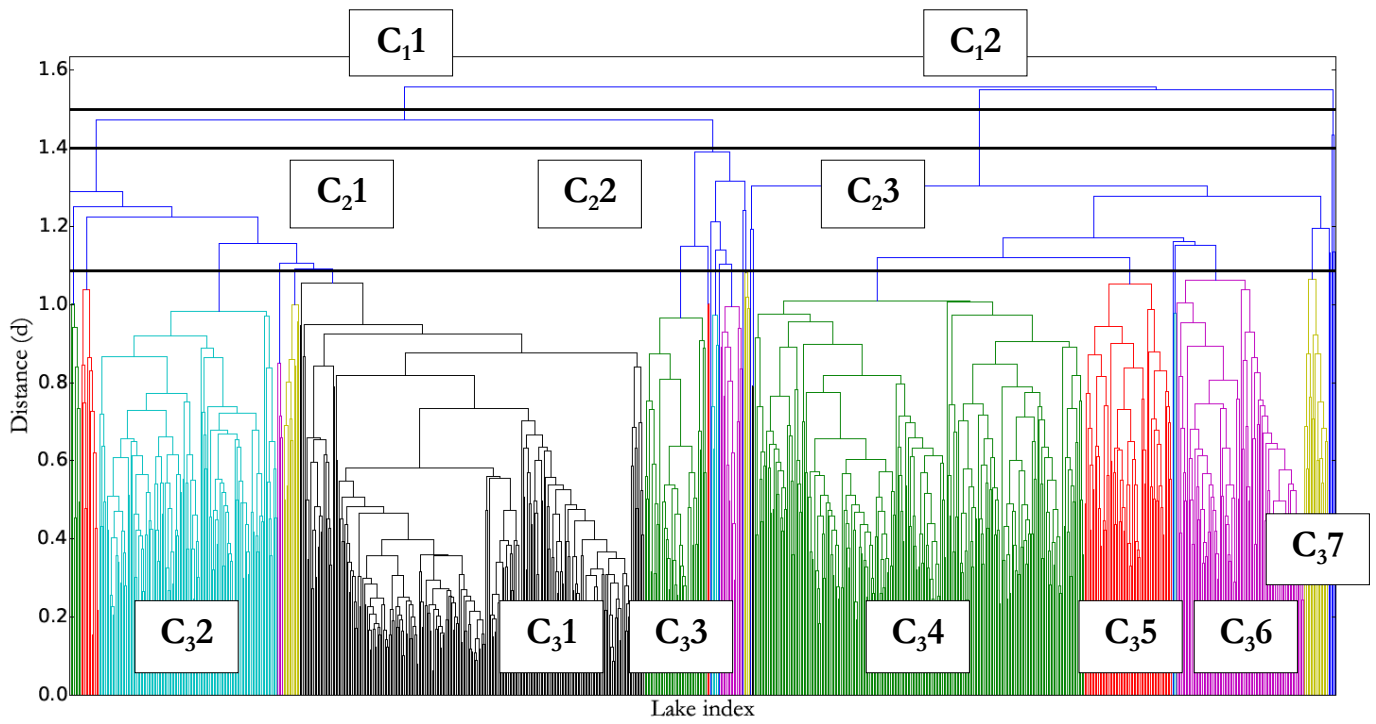
Fig. 4: HCA dendrogram with black lines indicating the chosen thresholds (d_1=1.5, d_2=1.3 and d_3=1.08). The obtained clusters are labeled within the figure. The clusters that result from cutting the dendrogram at the lowest threshold d_3 are colorized.
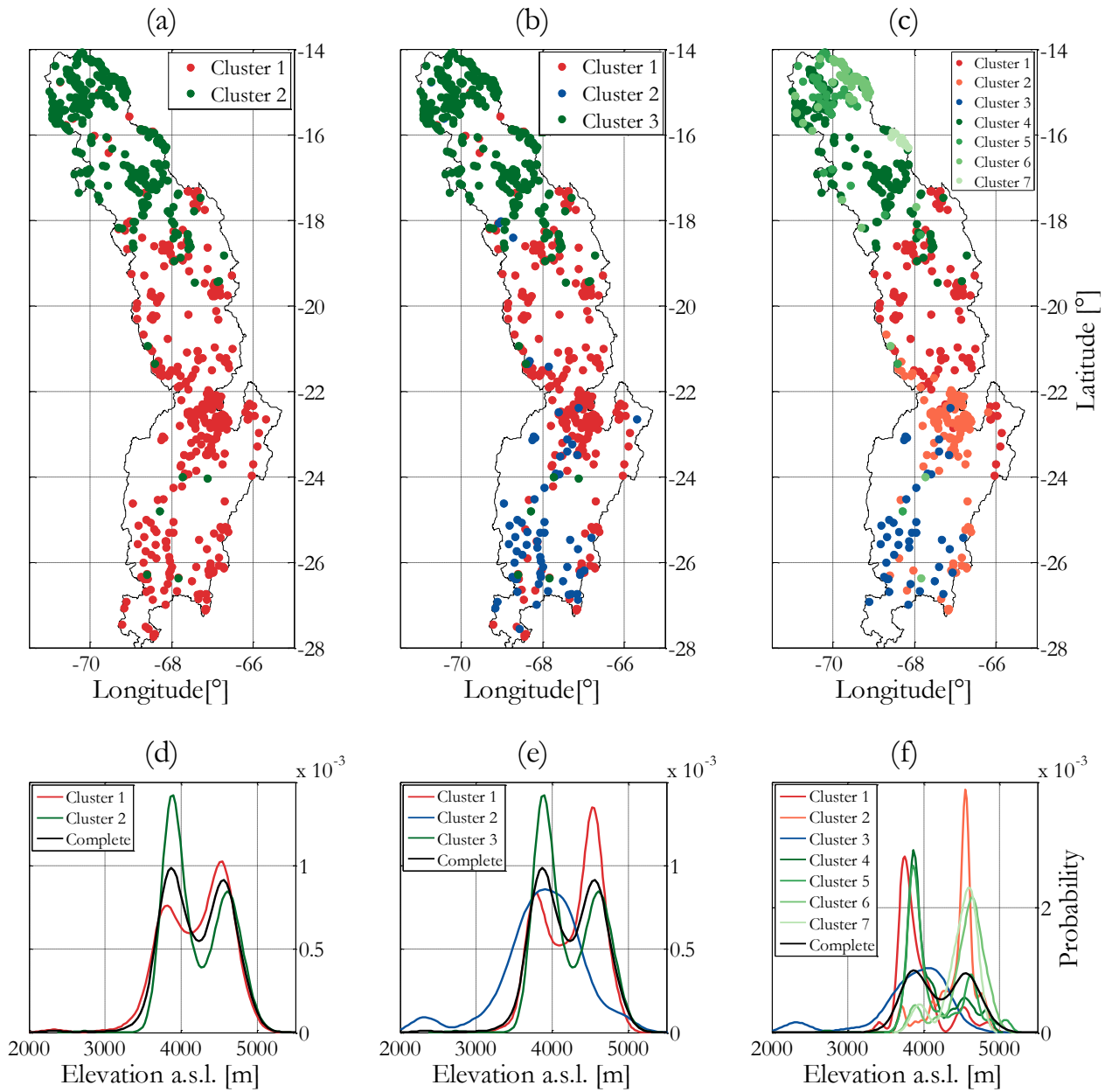
Fig. 5: Spatial distribution (a-c) and elevation histograms (d-f) of clusters for different HCA dendrogram thresholds (a, d) $d_1 = 1.5$, (b, e) $d_2 = 1.35$, (c, f) $d_3 = 1.08$.
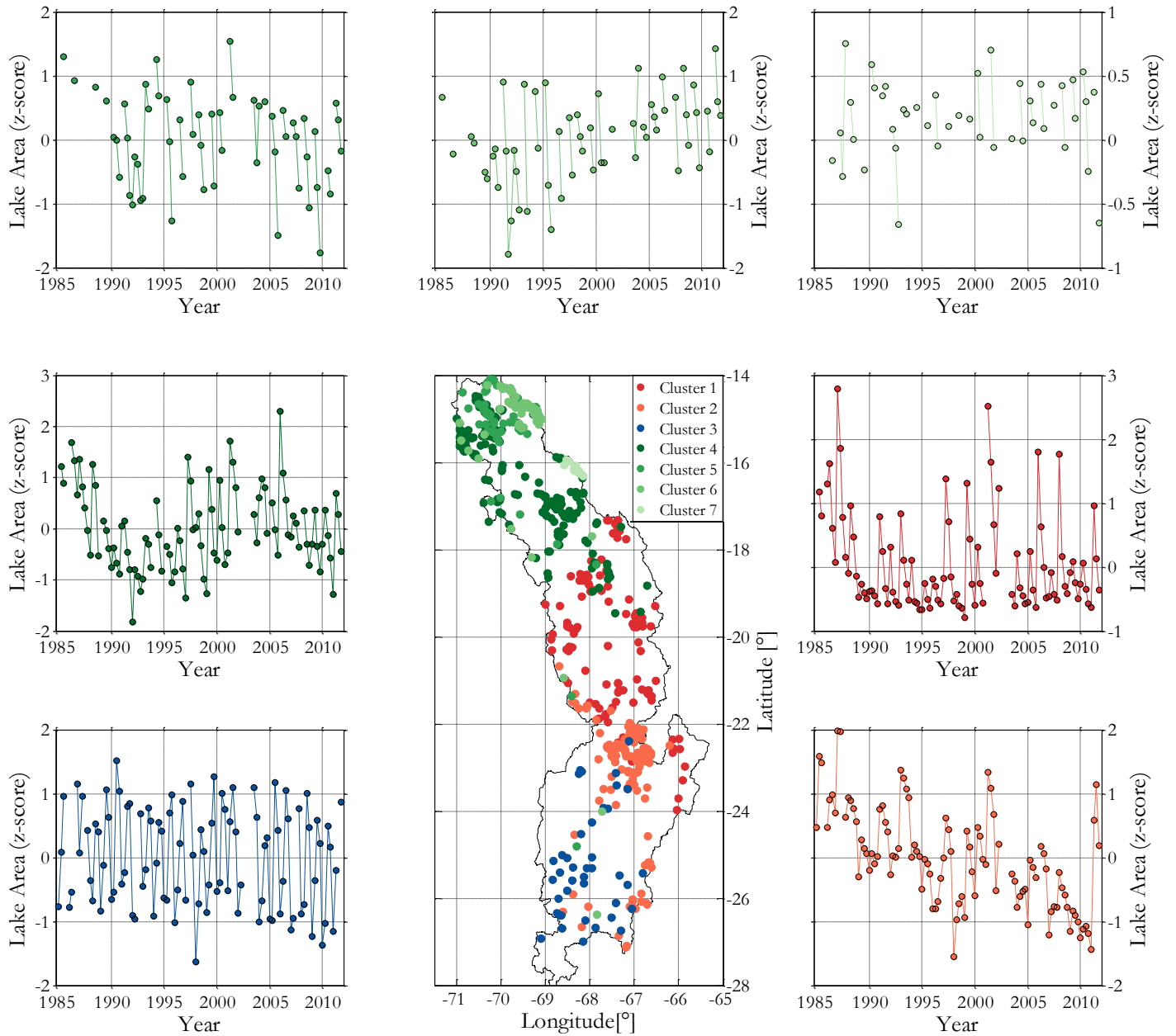
Fig. 6: Overview of clusters from HCA at the lowest threshold (d₃) with equally colored mean cluster time series. Lake areas have been standardized to z-scores.
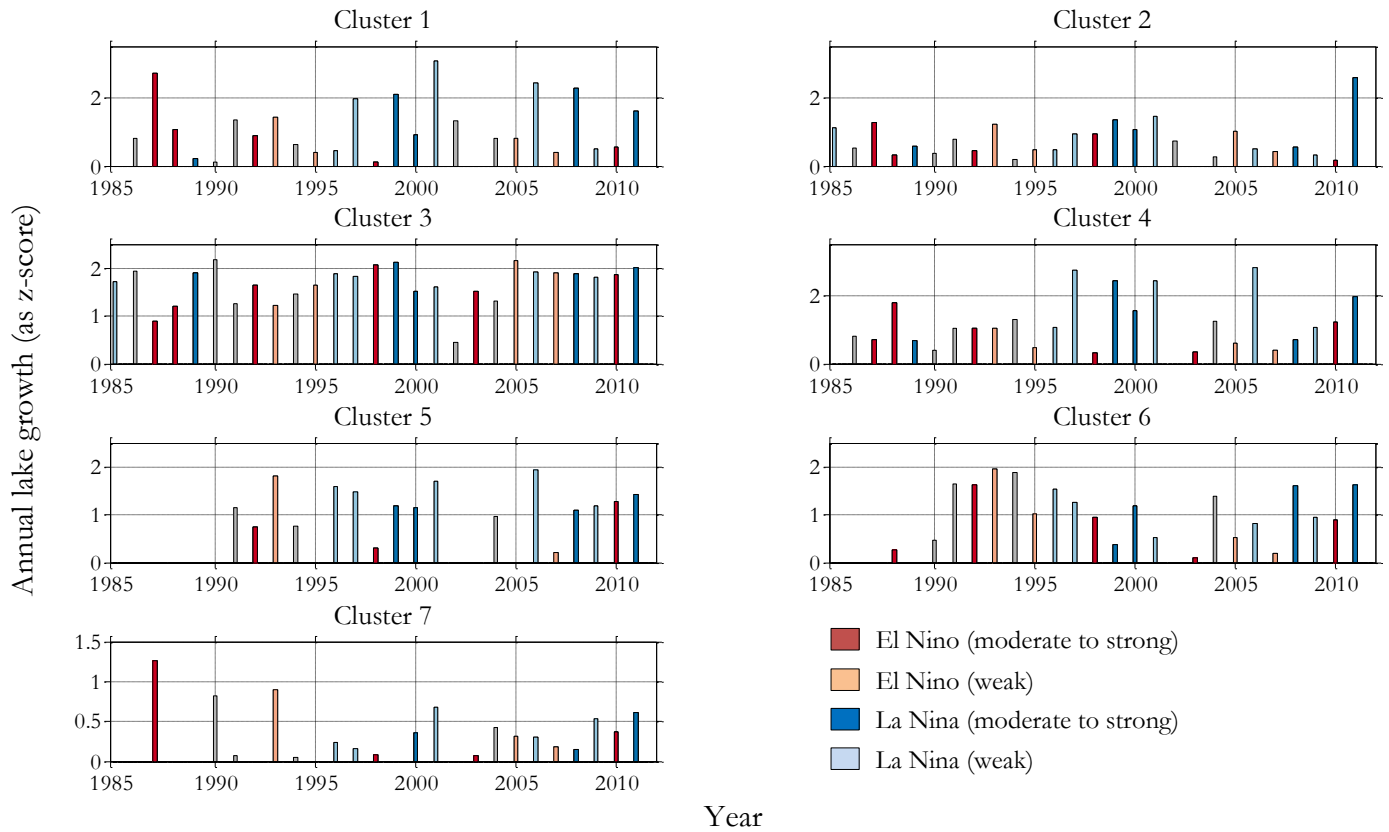
Fig. 7:   Annual lake growth time series as bar plots showing the annual lake growth (as z-scores) per year. Annual lake growths are the summed differences from DJF(0) until DJF(+1) of the mean cluster difference time series (see SI material) representing the difference of a measurement to the last measurement. Bars are colorized by the prevailing ENSO mode during the corresponding year, with dark blue and light blue indicating moderate to strong and weak La Nina events (ONI<-1 and -1<ONI<-0.5, respectively), neutral conditions (gray) and dark red and light red showing moderate to strong and weak El Nino episodes (ONI>1 and 0.5<ONI<1, respectively).

List of Tables

Cluster	Location	Description
C_31	Southern Altiplano	Mostly flat saline lakes, e.g. Lago Poopó, Salar de Uyuni, and Laguna de Pozuelo. Some of the lakes have areas greater than 100km² but are only up to a few meters deep. React very sensitive to changes in precipitation and evaporation.
C_32	Northeastern Puna	Similar to Cluster 1, but at higher elevations. Also flat and saline lakes (not as extensive as in Cluster 1) and in part deeper lakes.
C_33	Southwestern Puna	Small, flat and saline lakes south and east of the Atacama desert.
C_34	Southeastern northern Altiplano	Many lakes of different depths around the Lake Titicaca and along the Rio Desaguadero.
C_35	Northeastern northern Altiplano	Small and flat lakes in the mountainous areas of the northernmost parts of the Altiplano.
C_36	Northern East-Altiplano	Mostly proglacial lakes in deep valleys of the eastern Cordillera.
C_37	Southern East-Altiplano	Mostly proglacial lakes of the southernmost tropical glaciers.

Table 1: Overview of the clusters that we obtained from the HCA with the lowest threshold d_3 and description of the corresponding lakes and the location.

Supplementary Material

A. Landsat dataset

This study uses Landsat imagery to create time-series of lake area changes within the timeframe of 1985-2011. We obtained a total of 3365 images within this period from the Landsat 4 and Landsat 5 missions. Both missions work with the same sensor, the Thematic Mapper (TM) that covers 7 bands within a spectral range of 0.45 μm - 12.5 μm and at a resolution of 30 m (band 1-5 and band 7) and 120 m (band 6). More details on the bands are given in table S1.

We used Level-1G Landsat products, which have been radiometrically (e.g. Chander and Markham, 2007) and geometrically (http://landsat.usgs.gov/geometry.php) corrected. Landsat images are recorded as 185 km * 172 km scenes that correspond to a specific row in a path (orbit). There are 22 path/row combinations with 5 different paths that cover the CAP (Fig. S1)

B. Lake classification

One of the most common approaches in water classification is to utilize the spectral properties of water. In contrast to most land classes, water strongly absorbs the infrared portion of the radiation that is emitted from the sun and therefore has low intensities in the infrared-bands of optical satellite sensors. Thresholding those bands yields a binary image that separates water and other classes. However different sensors and even different scenes of a single sensor require different thresholds. In order to establish a unique threshold for water classification, McFeeters (1996) introduced the Normalized Difference Water Index (NDWI), a band ratio of the green and the near-infrared (NIR) sensor bands, normalized to values between -1 and 1:

$$NDWI = \frac{Green - NIR}{Green + NIR}$$

Pixels that contain water have positive, whereas pixels of other classes, e.g. vegetation, soil or rocks have negative values, as their reflectance is higher in the NIR spectral range than in the green range. The NDWI works very well for open and clear water, but cannot be used in the Central Andean Plateau. Most of the lakes in that area are very flat and saline and large algae mats form within those

lakes. The NDWI of those lakes can be negative. There are also miss-classifications of other land classes, such as built-up land (Xu, 2006). Water features are more pronounced in the short-wave infrared bands of satellite sensors. Xu (2006) therefore proposed a modified version of the NDWI, the Modified Normalized Difference Water Index (MNDWI), using a SWIR-band instead of the NIR-band:

$$MNDWI = \frac{Green - SWIR}{Green + SWIR}$$

In the study region the MNDWI and the SWIR-bands in general are less affected by algae or solubles within the water, compared to the NDWI or the NIR-band. However, the MNDWI indicated shore areas as water that cannot be visually perceived as water. Supposedly those are either areas with a thin layer of water on top or wet sediments. As both indices, the NDWI and the MNDWI, resulted in false or no classification in the direct vicinity of the lakes, the SWIR spectral range was used solely in this study instead of a band ratio. Classification was done on Landsat images from the Landsat 4 and 5 sensors. The band that is used is the first SWIR-band (band 5 for Landsat 4-5). The intensities of that band are thresholded to create a binary mask with the value 1 for water and zero for other land classes. As mentioned above, the threshold is quite variable. In order to automatically process a large amount of images, the threshold would have to adapt to changing scene characteristics. One option would be Otsu's methods (Otsu, 1979). The SWIR-histogram of most satellite images with water resembles a bimodal distribution, whereby water is one distribution and other surface types form a second distribution. Otsu developed a thresholding method to separate those distributions. In reality, however, the bimodal distribution is barely pronounced and the method fails to provide a reliable threshold. We found no other method to automatically adapt the threshold for every scene in a way that water is doubtlessly classified. Instead of using an adaptive threshold, one can also adjust all images to varying sensor parameters, e.g. the sensor-specific gain or the angle of the solar irradiation (compare Eq. 1 and Eq. 2), as variations of those parameters result in image intensity deviations of the same target. Nelson (1985) proposed a transformation of Landsat data to reflectance values in order to reduce inter-scene variability, as it is described by Chander and Markham (2003). During the processing of level-1G Landsat products pixel values are first converted to absolute radiance and then to 8-bit integer digital number (DN) values. DN-values can be converted back to radiance with the Landsat metadata that is provided with every Landsat product, as follows:

$$L_\lambda = G_{rs} * Q_{cal} + B_{rs}$$

where L_λ is the radiance, G_{rs} is the rescaled sensor gain, B_{rs} is the rescaled sensor bias and Q_{cal} is the pixel value as DN. With the additional information in the metadata, the planetary reflectance ρ_p can be calculated from the radiance:

$$\rho_p = \frac{\pi * L_\lambda * d^2}{ESUN_\lambda * \cos \theta_s}$$

where d is the earth-sun distance, $ESUN_\lambda$ is the mean solar exoatmospheric irradiance (values from Chander and Markham, 2003) and θ_s is the solar zenith angle in degrees, that is the difference of 90° minus the solar elevation (given in the Landsat metadata).

All Landsat images of the SWIR-band are converted to the top-of-atmosphere (TOA) reflectance and then segmented into water and non-water areas based on a threshold that captures the variability of water reflectivity. While most water pixels have very low intensities in the SWIR-reflectance, there are many pixels that show significantly higher values, either due to algae, solubles or due to mixing with other land types within the same pixel. The used threshold was empirically chosen to capture the complete range. This implies that there are also misclassifications of mostly cloud and mountain shadows. To avoid some of the false classifications, especially hillslope shadows, pixels with a slope larger than 20° are masked out. Slopes are calculated from the SRTM-30m digital elevation model (DEM). Originally a critical slope threshold of 5° was intended. However this led to masking out parts of the lakes in mountainous regions as (1) pixels of the DEM can be offset to the Landsat images and (2) the maximum lake extent can exceed the lake size at the time of the STRM recording (February 2000) and previously exposed hillslopes are covered by water during other periods within the timeframe.

The extraction of lakes from Landsat imagery requires additional steps to finally derive the time-series of lake areas. There are two parts of a program that was developed with Matlab and the Geospatial Data Abstraction Library (GDAL) python bindings. The first part (Fig. S2) creates, based on a cumulative classification of all images, an initial lake mask, which is afterwards used in the second part (Fig. S3) to compute the distinct areas of the lakes in every image.

B.1. Multi-temporal cumulative classification

At first, due to differences in the inter-scene sizes of Landsat products, all images of a scene are clipped to their minimum extent. This not only excludes areas that are not present throughout the complete stack, but also registers the images in the intrinsic image space, which is necessary for later processing. As the images are registered a visual inspection indicates images that are not precisely georeferenced or geometrically distorted. These images are deleted and the stack is reevaluated and clipped again. To get an idea of the occurrence and the distribution of lakes within the timeframe of 30 years, a first script loops through every image and cumulatively adds classified water to an initial water mask. As mentioned earlier, the classification of water is done by thresholding SWIR-reflectance intensities. The reflectance images are binarized using a threshold of $t_1 = 0.05$. A scene specific water mask is created by replacing all pixel values below that threshold with 1 and others with 0. In an additional step, pixels with a slope larger than 20° are also set to 0 to avoid misclassifying hill shadows. The extent of the clipped Landsat scene is used to crop the SRTM-DEM of the Central Andean Plateau and compute a slope map for the image region. In the process, the water mask of every image is added to an initial water mask (cumulative version) in order to not only find the maximum extent of every lake, but also to get an idea about the occurrence throughout the studied period. A binarized version of this initial water mask is converted to a shapefile with gdal, where all water bodies are separate polygons. This shapefile, as well as the cumulative initial water mask and additional satellite images are used to select the lakes of interest in a Geographical Information System (GIS) and to create a reference lake mask. The manual selection was necessary to delete small ponds, transient water bodies, river segments, as well as a large number of falsely classified mountain and cloud shadows. Some lakes have been manually edited, where nearby hillslope shadows or glaciers and snow have been falsely classified as part of the respective lakes.

B.2. Single-temporal lake classification

The first part is intended to give an estimation of where lakes have been present within the last 30 years, but does not give insights on the exact lake extents at every date covered by the Landsat images. The second part uses the reference lake mask and calculates the respective lake areas throughout the images. The updated reference lake mask shapefiles are first rasterized to create a binary lake mask that indicates the location of the selected lakes. Afterwards the lakes of all scenes are labeled and assigned an unique index in a reference water mask. The algorithm then loops

through the reflectance images of every stack and classifies water as described in the previous section, but only in the areas where the water mask assumes water. Every classified object is labeled according to the corresponding index in the reference water mask. Finally the amount of pixels for every object is calculated and together with the image resolution the area of every lake can be calculated.

C. Hierarchical clustering analysis

We derived 1058 lake-area time series from the lake classification process, but assume that several time series show either systematic errors or specific errors at certain acquisition dates, that are related to the semi-automatic classification, but also to the averaging to 3-month intervals. To increase the robustness of the latter analyses, we decided to combine lake-area time series with a similar coevolution. For this matter we used the hierarchical clustering analysis (HCA; e.g. Jain et al., 1999; Manning et al., 2008; Rokach and Maimon, 2010; Murtagh and Contreras, 2012) that agglomeratively connects similar time series until every time series is assigned to the same cluster. The iterative process of the HCA is as follows: First, each time series is assigned a unique cluster. Second, the unique time series clusters are combined based on a distance metric (e.g. Murtagh and Contreras, 2012), which is a measure for similarity. In this study similarity is expressed as the correlation of the lake-area time series. The distance (d) is the arcos of the Pearson correlation coefficient r. The more different (similar) time series are from each other, the higher (lower) the angle in vector space and the distance between time series. In the further iterations, as clusters consist of more than one time series, a linkage criterion is used to in addition to the distance metric to merge clusters (Table S2).

We chose the average linkage criterion for our study. The complete linkage criterion tends to generate many small clusters and is very sensitive to outliers as the maximum distance is considered when combining clusters. As we assume errors and uncertainties within our dataset, the complete linkage criterion is not suitable for the clustering of the lake-area time series. The single linkage criterion tends to chaining effects and as it produces very large clusters. As this might override apparent differences in the lake-area time series, we further decided against the single linkage criterion. The distance between two time series cluster (u,v) and its cardinalities with the average linkage criterion and the distance metric d for all time series (i,j) within both clusters is then given by:

$$d(u,v) = \sum_{ij} \frac{d(u[i], v[j])}{(|u| * |v|)}$$

Correlation coefficients of the time series have been calculated pairwise, whereby rows of a pair with no value have been ignored. Additionally lakes with more than 80 percent of either no occurrence or no value within the 3-month intervals have been deleted. The stepwise aggradation of time series to common clusters is represented by a dendrogram that plots the distance versus the time series clusters. At a distance of d=0, every time-series is assigned to its own cluster and every node in the consequent cluster merging process represents the unification of two clusters and the height indicates the average distance of both clusters. The dendrogram can be cut at every distance in order to either get a specific number of clusters or to ensure an average distance of the obtained clusters.

We thresholded the dendrogram at three different distances (d_1 = 1.5, d_2 = 1.35 and d_3 = 1.08) in order to get different amounts of clusters (2, 3 and 7, respectively, after the removal of outlier clusters). Clusters that contain less than 10 lake-area time series and have a large spatial spread have been removed as they are regarded as outliers.

The distance (d) at which we cut the dendrogram indicates that the cluster specific lake-area time series correlate on average with cos(d). This however implies that there can be time series pairs of the same cluster that have much lower correlation coefficients, as it is indicated by Fig. C1 and Fig. C2. The correlation matrices (Fig. C1) show that several time series pairs do not correlate at all, even though they belong to the same cluster. In order to find lake-area time series that generally correlate poorly with the other time series of a cluster, we averaged the sums of the correlation matrix rows. This gives, for every lake-area time series of a cluster, the mean correlation coefficient (r_{CML}) within the same cluster. Lakes of a cluster with $r_{CML} < 2\sigma$ are regarded as outlier and are removed (Fig. C2). This process removed several lake time series of $C_3 2$ but only few of the other clusters. The removal was applied to increase the quality of the average time series of every cluster (mean cluster time series).

D. Cluster mean difference time series

To facilitate inter-annual comparisons within a mean cluster time series, we calculated the corresponding difference time series that represent the difference of a measurement to the preceding measurement. Any apparent trends within the mean cluster time series are removed in the difference

time series. If X is a mean cluster time series of length m and Y is the difference time series of X, then the mean cluster difference time series (Y) can be calculated as:

$$Y(i\text{-}1) = X(i) - X(i\text{-}j),$$

whereby j (0<j<i) represents the distance to the last measurement and i (i = [2,3,…,m]) represents the measurement index of the time series and therefore a 3-month interval. The resulting mean cluster difference time series are shown in Fig. S6. In order to keep the length of the mean cluster time series, the difference time series are appended by NaN at Y(1).

We further summed the positive difference (Y(i)>0) of every year, i.e. from DJF to the subsequent DJF season. The yearly sums of the positive difference represent the annual lake growth time series that cumulate annual increases in lake areas within the DJF season of a year under the assumptions that most rainfall over the CAP occurs in austral summer. The annual lake growth time series indicate how much total area a lake gained throughout a year.

E. Cross-correlation with climate indices

We performed cross-correlation analysis with several climate indices (Table S3) that represent change characteristics of the Atlantic and the Pacific Ocean in order to decipher any apparent relationships between large-scale oscillations of the western Hemisphere and variability of lake sizes in the Central Andes. As a correlation metric we used the Pearson correlation coefficient with a minimum (maximum) lag of 0 (6) months and with a step length of 1 month. Since the mean cluster time series consist of 3-month intervals, we had to adapt the climate index time series, which are sampled in monthly steps, for every lag. When the lag increases by 1 month, the monthly climate index time series is shifted by 1 month and the months, which overlap with a 3-month bin of the cluster mean time series, are averaged. The resulting cross-correlation plots are shown in Fig. S7. Significance was tested at the 95% level by comparing correlation coefficients with values obtained from correlations of the climate index time series with 10000 random permutations of the cluster mean time series.

The cross-correlations reveal several significant correlations, especially with the Nino-indices (Nino12, Nino3, Nino34, Nino4), but also with temperature anomalies of the Atlantic (TSA, TNA, AMO) or the Pacific (PDO). The previously mentioned climate indices show significant correlations with almost all clusters. In the case of Nino12, the cluster of the Altiplano ($C_3 1$, $C_3 4$, $C_3 5$, $C_3 6$ and $C_3 7$) show moderate to strong positive correlation coefficients of around 0.5 within the first 3

months and moderate negative correlations of around -0.3 after that. $C_3 3$ is shifted with a strong negative correlation ($r \approx -0.6$) at lag=0 and strong positive correlations ($r \approx 0.6$) at lag=5. The results for the other Nino-indices are similar but shifted. With Nino12 being farthest to the East and Nino4 farthest to the West, the cross-correlations not only decrease in amplitude but also shift towards smaller lags. Hence the Nino4 region shows no more positive correlations with the Altiplano in the first 3 months but with the southwestern Puna ($C_3 3$). The PDO shows mainly positive correlations with $C_3 1$, $C_3 2$, $C_3 3$, $C_3 4$ and $C_3 5$, but also negative correlations with $C_3 6$. Cross-correlations with the TSA are positive in the northern CAP ($C_3 4$, $C_3 5$, $C_3 6$ and $C_3 7$) and negative in the southern CAP ($C_3 1$, $C_3 2$ and $C_3 3$). $C_3 6$ shows the strongest positive correlations with the TSA ($r \approx 0.5$). For the southern CAP the AMO shows similar correlations as the TSA, with maximum negative correlations ($r \approx -0.5$) in the northeastern Puna ($C_3 2$). The only positive correlation of the AMO-index is observed in $C_3 6$. ENSO-indices (MEI, BEST) show only few significant and weak correlations with $C_3 3$, $C_3 6$ and $C_3 4$. Only few significant correlations are also observed for the AAO-index ($C_3 6$) and the indices covering the pressure differences over the Pacific, the SOI and the EQSOI ($C_3 2$ and $C_3 6$) The strong correlations with the Nino indices however do not necessarily support an influence of ENSO periods on lake size variability. We argue that the correlations are related to similar oscillations of the Nino time series and the lake area time series with increasing lake sizes (temperatures) in the Central Andes (eastern Pacific Ocean). This is especially obvious comparing $C_3 3$ and Nino12 (Fig. D2), which are strongly anti-correlated, as the largest lake areas occur in austral winter when temperatures in the eastern Pacific reach their lowest point throughout the year. Despite correlations with TSA and AMO it is not evident that positive or negative anomalies in either of both climate indices generally coincide with positive or negative anomalies in lake sizes. The strong negative correlations of AMO and $C_3 2$ that do not change strongly at different lags may rather show an inverse development with a positive (negative) linear trend of the AMO ($C_3 2$) within the last 30 years that may be contributed to similar forcing, e.g. rising global temperatures that warmed the northern Atlantic and lead to the continuous reduction of lake areas in the northeastern Puna.

List of References

Chander, G., Markham, B., 2003. Revised Landsat-5 TM Radiometrie Calibration Procedures and Postcalibration Dynamic Ranges. IEEE Trans. Geosci. Remote Sens. 41, 2674–2677. doi:10.1109/TGRS.2003.818464

Jain, a. K., Murty, M.N., Flynn, P.J., 1999. Data clustering: a review. ACM Comput. Surv. 31, 264–323. doi:10.1145/331499.331504

Manning, C.D., Raghavan, P., Schütze, H., 2009. Hierarchical clustering. Introd. to Inf. Retr. 377–401. doi:10.1017/CBO9780511809071.017

McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. Int. J. Remote Sens. 17, 1425–1432. doi:10.1080/01431169608948714

Murtagh, F., Contreras, P., 2012. Algorithms for hierarchical clustering: An overview. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 2, 86–97. doi:10.1002/widm.53

Nelson, R.F., 1985. Sensor-induced temporal variability of Landsat MSS data. Remote Sens. Environ. 18, 35–48. doi:10.1016/0034-4257(85)90036-7

Rokach, L., Maimon, O., 2010. Chapter 15— Clustering methods. Data Min. Knowl. Discov. Handb. 32. doi:10.1007/0-387-25465-X_15

Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int. J. Remote Sens. 27, 3025–3033. doi:10.1080/01431160600589179
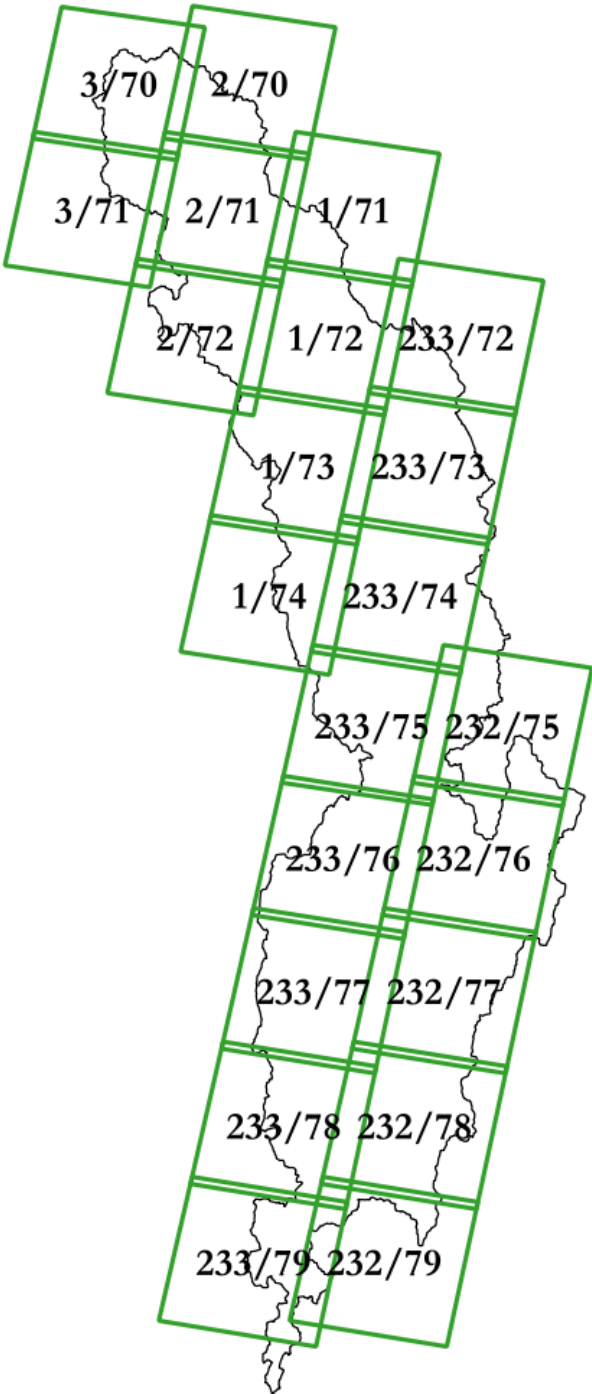
List of Figures

Figure S1:   Overview of the Central Andean Plateau and the corresponding Landsat scenes. Path/row combinations are labeled respectively.
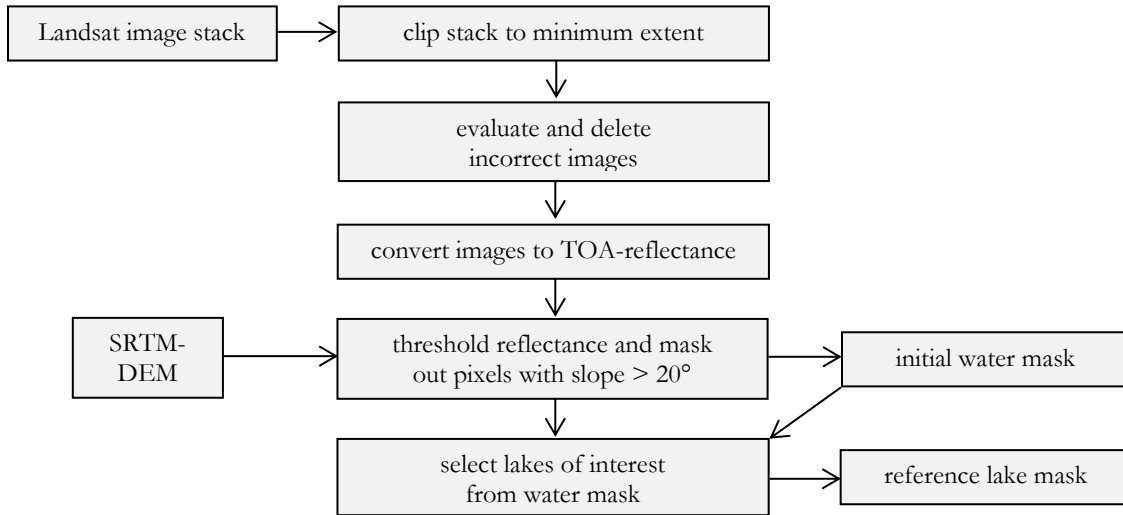
```
┌─────────────────────┐      ┌──────────────────────────────┐
│ Landsat image stack │ ───▶ │  clip stack to minimum extent │
└─────────────────────┘      └──────────────────────────────┘
                                           │
                                           ▼
                             ┌──────────────────────────────┐
                             │       evaluate and delete      │
                             │        incorrect images        │
                             └──────────────────────────────┘
                                           │
                                           ▼
                             ┌──────────────────────────────┐
                             │ convert images to TOA-reflectance │
                             └──────────────────────────────┘
                                           │
   ┌──────────┐                            ▼
   │  SRTM-   │              ┌──────────────────────────────┐      ┌────────────────────┐
   │  DEM     │ ──────────▶  │ threshold reflectance and mask │ ──▶  │ initial water mask │
   └──────────┘              │   out pixels with slope > 20°  │      └────────────────────┘
                             └──────────────────────────────┘
                                           │
                                           ▼
                             ┌──────────────────────────────┐      ┌────────────────────┐
                             │       select lakes of interest │ ──▶  │ reference lake mask │
                             │         from water mask        │      └────────────────────┘
                             └──────────────────────────────┘
```

Figure S2:   Scheme of the cumulative water classification that shows the processing steps to get an initial lake
                  mask from a stack of Landsat images.

```
                             ┌──────────────────────────────┐
                             │       reference lake mask      │
                             └──────────────────────────────┘
                                           │
                                           ▼
                             ┌──────────────────────────────┐
                             │   lake object reference labeling │
                             └──────────────────────────────┘
                                           │
   ┌─────────────────────┐                 ▼
   │ Landsat image stack │    ┌──────────────────────────────┐
   │ (TOA – reflectance) │ ─▶ │  thresholding based on lake mask │
   └─────────────────────┘    └──────────────────────────────┘
                                           │
                                           ▼
                             ┌──────────────────────────────┐
                             │       specific lake mask       │
                             └──────────────────────────────┘
                                           │
                                           ▼
                             ┌──────────────────────────────┐      ┌────────────────┐
                             │    lake object labeling based  │ ──▶  │   lake areas   │
                             │        on reference labels     │      └────────────────┘
                             └──────────────────────────────┘
```
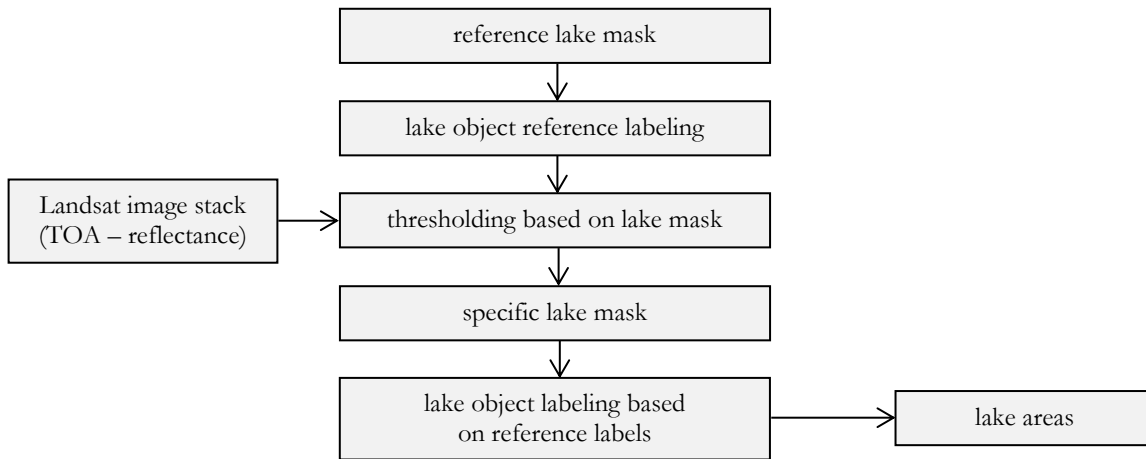
Figure S3:   Scheme of the specific water classification that shows the processing steps to compute date specific
                  lake areas from the previously calculated initial lake mask and the Landsat reflectance images.
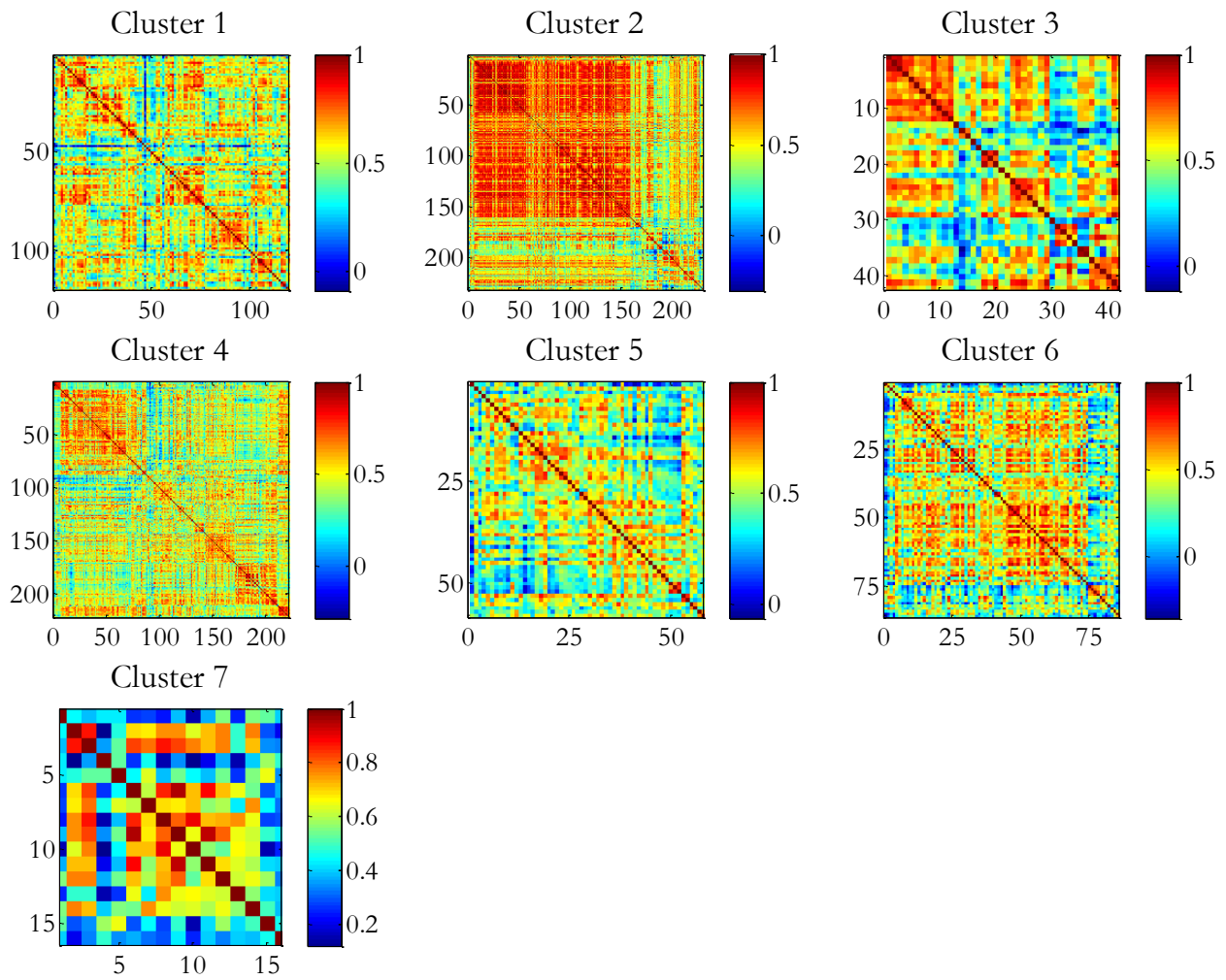
Fig. S4: Matrices of Pearson correlation coefficients for every cluster that indicate the correlation of a lake with the other lakes of the same cluster.
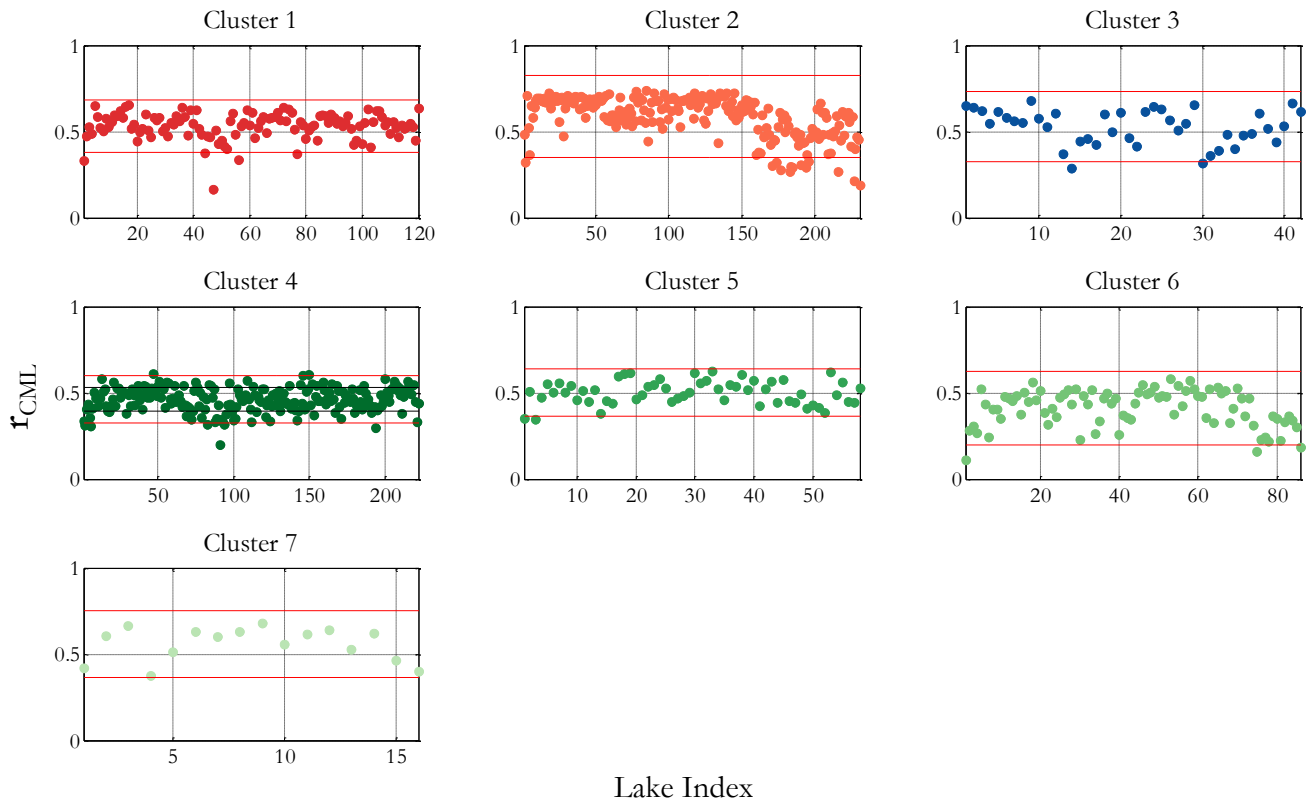
Fig. S5: Overview of the cluster mean lake correlation coefficients (r_{CML}) and 2σ boundaries. The lower 2σ boundary acts as a threshold for the outlier detection. Lakes below 2σ were removed from the clusters.
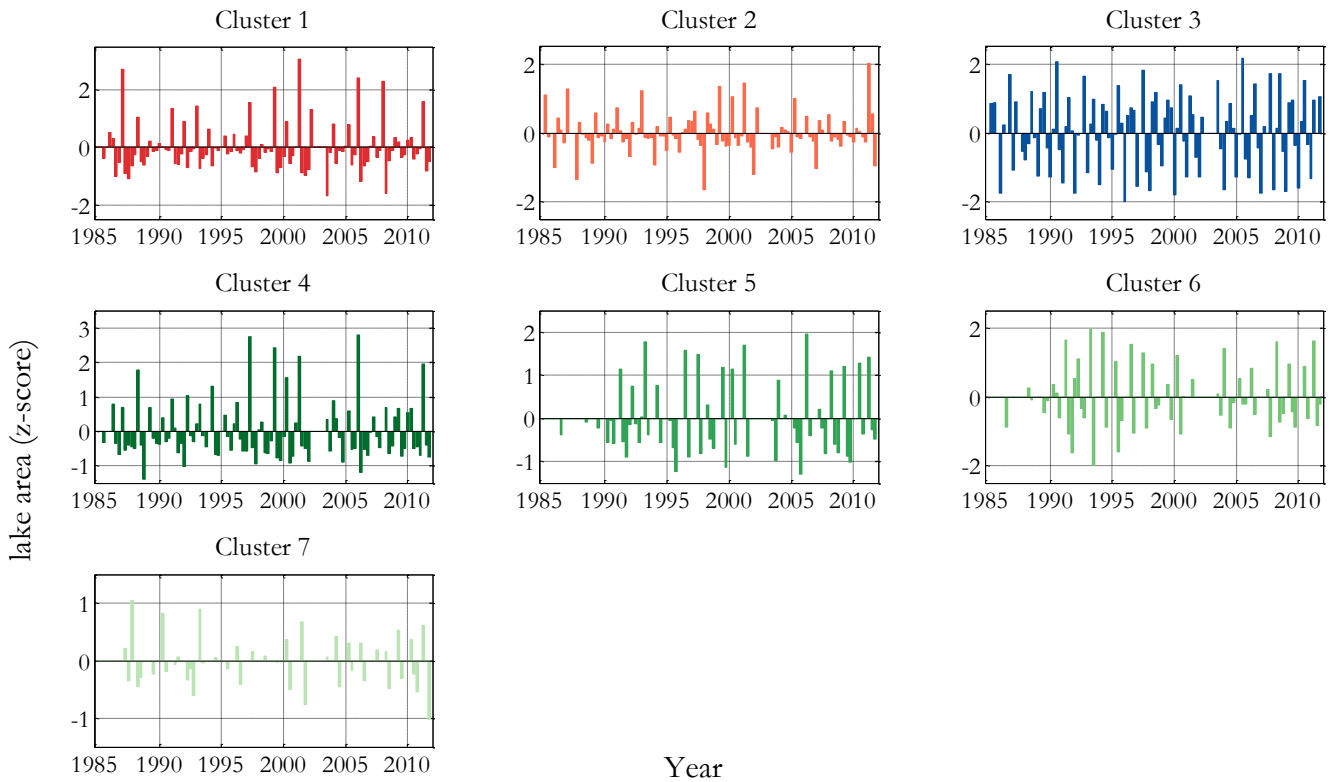

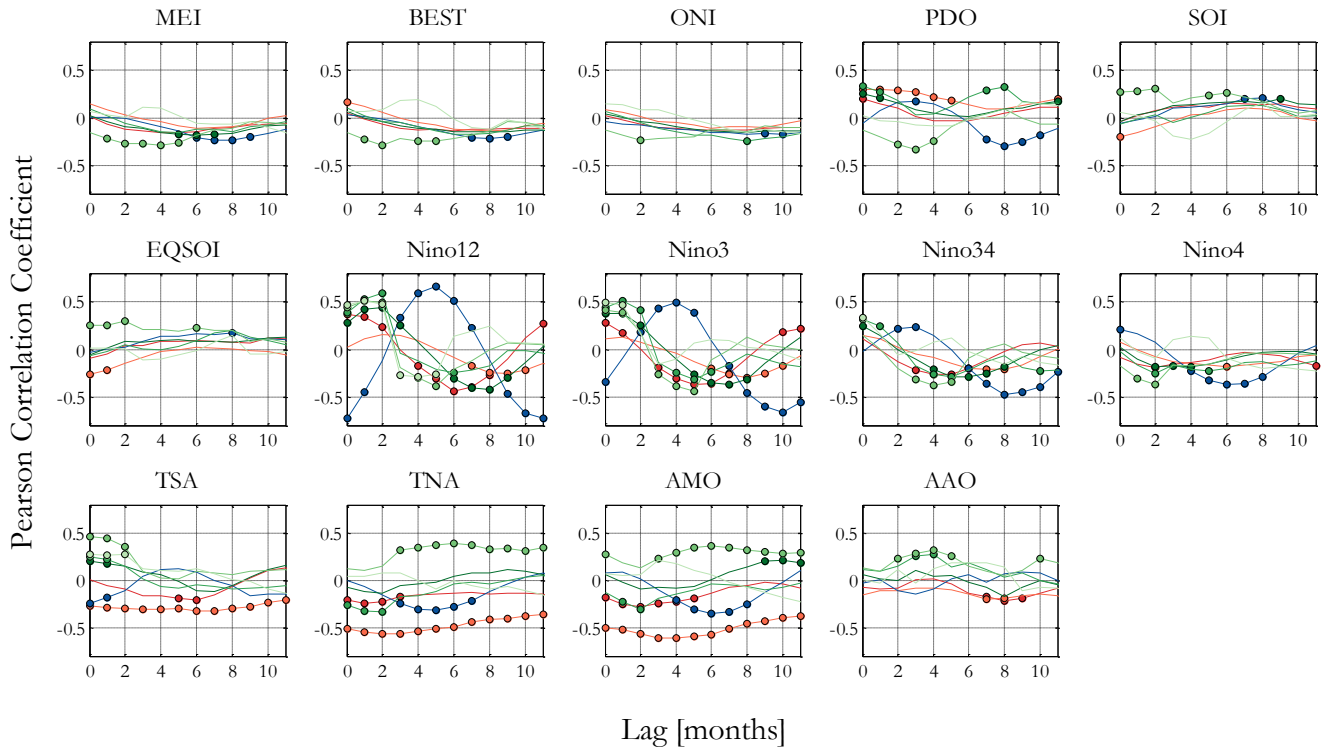Fig. S6: Mean cluster difference time series for every cluster. Plots show the lake areas as z-scores vs. the year.

Fig. S7: Mean cluster difference time series for every cluster. Plots show the lake areas as z- scores vs. the year.
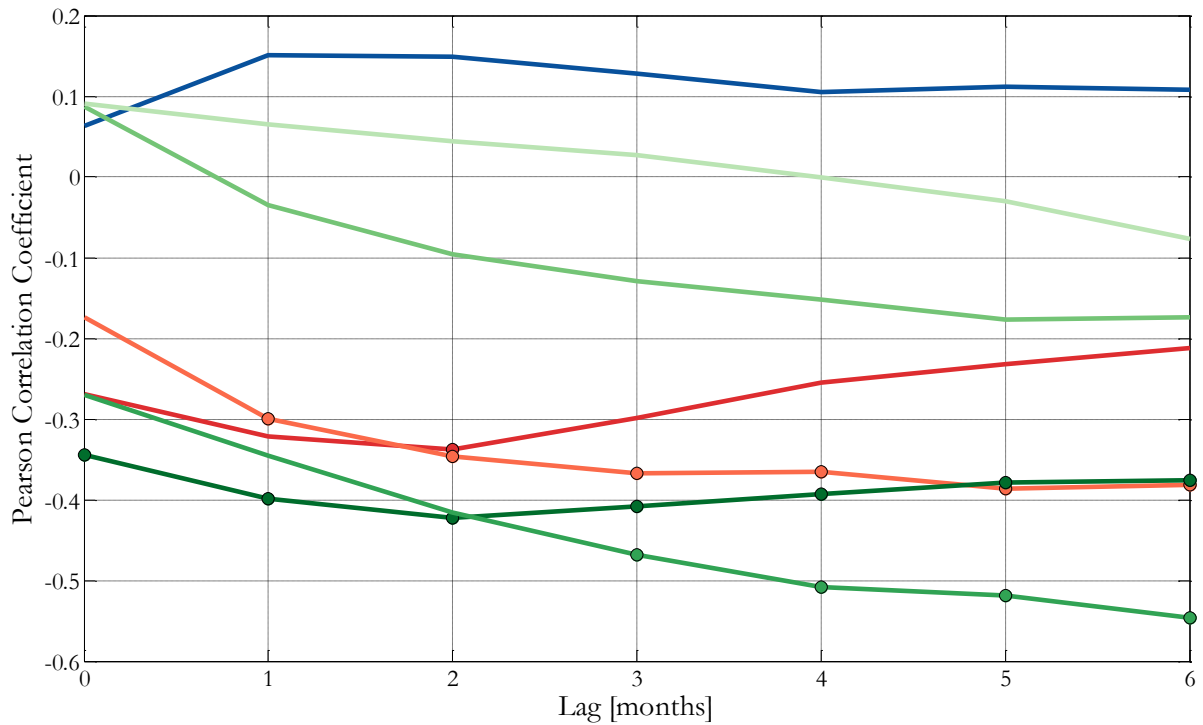

Fig. S8: Cross-correlation of MEI and mean cluster difference time series of every cluster for lags of 0-6 months
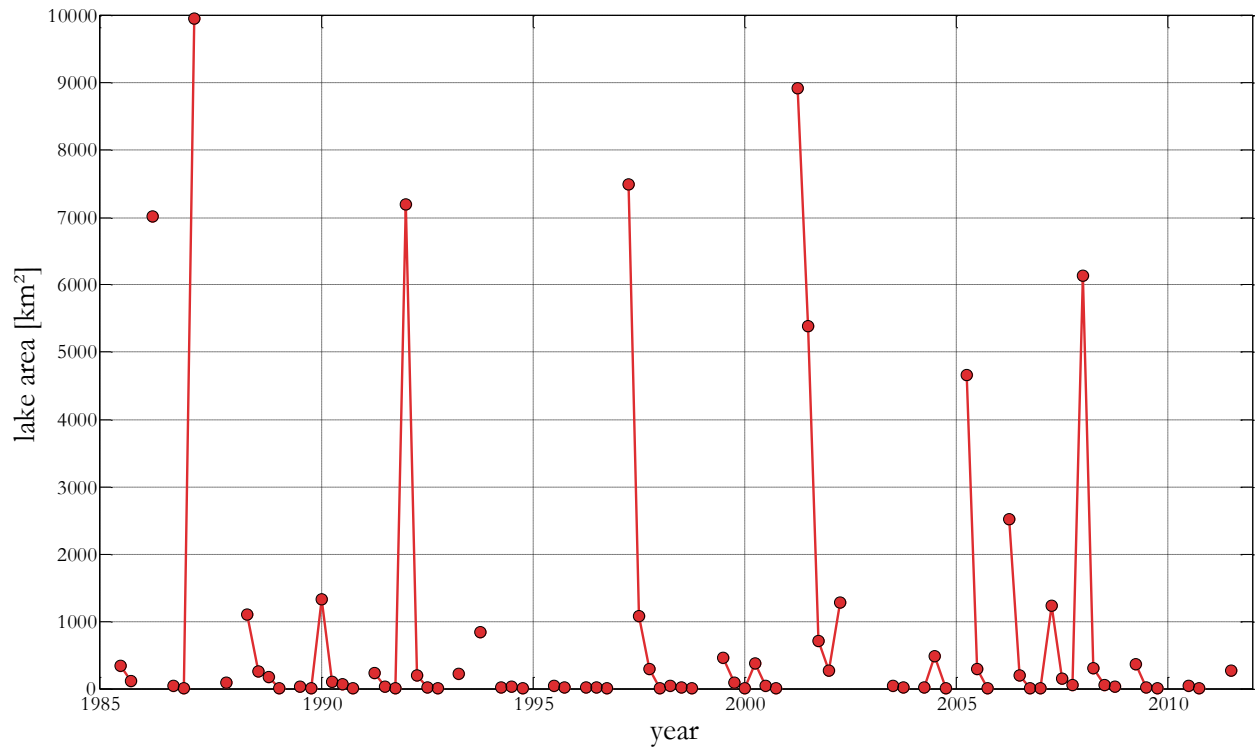
Fig. S9: Time series of the Salar de Uyuni in Bolivia from 1985-2011.

List of Tables

Band	Wavelength [μm]	Resolution [m]
1 – Blue	0.45-0.52	30
2 – Green	0.52-0.60	30
3 – Red	0.63-0.69	30
4 – Near Infrared	0.76-0.90	30
5 – Short-wave infrared I	1.55-1.75	30
6 – Thermal Infrared	10.41-12.5	120
7 – Short-wave infrared II	2.08-2.35	30

(http://landsat.gsfc.nasa.gov/the-thematic-mapper/)

Table S1: Landsat band designations with the corresponding wavelengths and the spatial resolution

Linkage criterion	Formula
Complete (maximum) linkage	$d(u,v) = \max(d(u[i], v[j]))$
Single (minimum) linkage	$d(u,v) = \min(d(u[i], v[j]))$
Average linkage	$d(u,v) = \sum_{ij} \dfrac{d(u[i], v[j])}{(|u| * |v|)}$

https://docs.scipy.org/doc/scipy-0.18.1/reference/generated/scipy.cluster.hierarchy.linkage.html

Table S2: Different linkage criteria and corresponding formulas for the distance metric (d), the clusters (u,v), the cardinalities of the clusters (|u|,|v|) and the points within the clusters (i,j).

Climate Index		Source
AAO	- Antarctic Oscillation Index	Temperature Anomalies (Antarctic Ocean)
AMO	- Atlantic Multidecadal Oscillation Index	Temperature Anomalies (Northern Atlantic)
BEST	- Bivariate ENSO Timeseries Index	Pressure and Temperature Anomalies (Pacific)
EQSOI	- Equatorial Southern Oscillation Index	Sea Level Pressure Differences (East/West Pacific)
MEI	- Multivariate ENSO Index	Sea Level Pressure, Wind Fields, Air and Sea Temperatures, Cloudiness (Pacific)
Nino12	- Nino12	Temperature in the Eastern Pacific
Nino3	- Nino3	Temperature in the Eastern/Central Pacific
Nino34	- Nino34	Temperature in the Eastern/Central Pacific
Nino4	- Nino4	Temperature in the Eastern/Central Pacific
ONI	- Oceanic Nino Index	Temperature Anomalies (Eastern/Central Pacific)
PDO	- Pacific Decadal Oscillation Index	Temperature Anomalies (Northern Pacific)
SOI	- Southern Oscillation Index	Sea Level Pressure Differences (East/West Pacific)
TNA	- Tropical North Atlantic Index	Temperature Anomalies (Northern Atlantic)
TSA	- Tropical Southern Atlantic Index	Temperature Anomalies (Southern Atlantic)

(http://www.esrl.noaa.gov/psd/data/climateindices/list/)

Table S3: Overview and description of climate indices used in our study

Cluster Index	Linear Regression Coefficient (β)	p-value
1	-0.005	0.063
2	-0.015	2.263e-11
3	-0.003	0.261
4	-7.495e-4	0.776
5	-0.006	0.084
6	0.009	9.148e-4
7	4.936e-4	0.735

Table S4: Regression coefficients from weighted linear regression of the mean cluster time series of every cluster. Significance was tested with a t-test distribution.

Acknowledgement

Selbstständigkeitserklärung

Ich versichere, dass ich, Nicolas Werner, die vorliegende Arbeit selbstständig und nur mit den angegebenen Quellen und Hilfsmitteln (z. B. Nachschlagewerke oder Internet) angefertigt habe. Alle Stellen der Arbeit, die ich aus diesen Quellen und Hilfsmitteln dem Wortlaut oder dem Sinne nach entnommen habe, sind kenntlich gemacht und im Literaturverzeichnis aufgeführt. Weiterhin versichere, ich, dass weder ich noch andere diese Arbeit weder in der vorliegenden noch in einer mehr oder weniger abgewandelten Form als Leistungsnachweise in einer anderen Veranstaltung bereits verwendet haben oder noch verwenden werden.

Die „Richtlinie zur Sicherung guter wissenschaftlicher Praxis für Studierende an der Universität Potsdam (Plagiatsrichtlinie) - Vom 20. Oktober 2010", im Internet unter http://uni-potsdam.de/ambek/ambek2011/1/Seite7.pdf, ist mir bekannt.

Es handelt sich bei dieser Arbeit um meinen ersten/zweiten Versuch.

_____

Ort, Datum   Unterschrift