A predictive model for Lake Chad total surface water area using remotely sensed and modeled hydrological and meteorological parameters and multivariate regression analysis

Frederick Policelli [a,d], Alfred Hubbard [b,c], Hahn Chul Jung [a,c], Ben Zaitchik [d], Charles Ichoku [e]

[a] NASA Goddard Space Flight Center
8800 Greenbelt Rd
Code 617, Hydrological Sciences Lab
Greenbelt, MD, 20771, USA

[b] NASA Goddard Space Flight Center
8800 Greenbelt Rd
Code 618, Biospheric Sciences Lab
Greenbelt, MD, 20771, USA

[c] Science Systems and Applications, Inc.
10210 Greenbelt Rd
Lanham, MD, 20706, USA

[d] Johns Hopkins University
Dept. of Earth and Planetary Sciences

3400 N. Charles St.
Baltimore, MD 21218, USA

[e] Howard University
Dept. of Interdisciplinary Studies
1840 7th Street, NW
Washington, DC 20001, USA

* Correspondence: frederick.s.policelli@nasa.gov; Tel.: +01-301-614-6573

Abstract

Lake Chad is an endorheic lake in west-central Africa at the southern edge of the Sahara Desert. The lake, which is well known for its dramatic decrease in surface area during the 1970s and 1980s, experiences an annual flood resulting in a maximum total surface water area generally during February or March, though sometimes earlier or later. People along the shores of Lake Chad make their living fishing, farming, and raising livestock and have a vested interest in knowing when and how extensive the annual flooding will be, particularly those practicing recession farming in which the fertile ground of previously flooded area is used for planting crops. In this study, the authors investigate the relationship between lake and basin parameters, including rainfall, basin evapotranspiration, lake evapotranspiration, lake elevation, total surface water area, and the previous yearôs total sur season month (except November) linking total surface water area to the other parameters. The resulting equations allow the user to estimate the December average monthly total surface water area of the lake in late November, and to make the estimates for January to May in early December. Based on the results of a Leave One Out Cross Validation analysis, the equations for lake area are estimated to have an average absolute error ranging from 5.3 percent (for February estimates) to 7.6 percent (for May estimates).

[1]Keywords: surface water area, Lake Chad, precipitation, evapotranspiration, lake height

1. Introduction

Lake Chad is a shallow, endorheic lake in the Sahel region of west-central Africa shared by Chad, Nigeria, Niger, and Cameroon. There have been numerous hydrological models developed for Lake Chad. Each of these has limitations relative to the statistical model for lake area we present here.

Bader et al. (2011) developed a hydrological model that simulates the water level in the northern pool, southern pool, and archipelago using riverine and direct rainfall inputs to the lake. It also estimates total water area for each of the pools. According to (Lemoalle et al., 2012), the model results correspond well with satellite measurements of northern pool surface water area and with satellite measurements of the total surface water area from 1980 onward, though these

---

Abbreviations: Inter-Tropical Convergence Zone (ITCZ), above sea level (ASL), Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS), Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System (FLDAS), Land Surface Temperature (LST), evapotranspiration (ET), Leave One Out Cross Validation (LOOCV), precipitation (P). For reading regression results: Pws is total south basin wet season precipitation, ETws is total south basin wet season evapotranspiration, LakeETws is wet season lake ET percent variation from the 1988 to 2016 wet season mean, H is November lake height variation from the 1993-2002 mean, A is the average surface water area for the given month, A- is the previous year's average surface water area for the given month.

results are not quantified. One disadvantage of this model is that it requires input of the Chari River and Komodougu-Yobe River discharges, data that is not publicly available.

Coe and Foley (2001) report the results of a hydrological model of Lake Chad. They describe a model with ñgood agreement with the inferre[...] 1954 to 1967. The ñinferred lake areaò is not a d[...], but rather is derived from a relationship between lake area and lake level. It is important to note that the hydrology of Lake Chad changed significantly in the 1970ôs [...] ñSmall Lake Chadò and the model, based on a coarse (10 km resolution) digital elevation model, may not be adequate to define the lake after this transition.

Gao et al. (2011) developed a hydrological model of Lake Chad. They compared images of the lake extent from the model with images derived from remote sensing. Three image pairs were shown, for October 31, 1963; December 25, 1972; and January 31, 1987. Each pair of images from the two earlier dates (before the transi[...] ñSmall Lake Chadò) looked quite similar, though no numerical value was provided. The model-observation pair for the post transition period did not look very similar. This raises the question of the utility of the model for producing lake area in the current, post transition period.

Lemoalle (2004) developed a crude expression for Lake Chad surface area based on a simplified water balance model and described [...] assumes no seepage from the lake and requires knowledge of the streamflow to the lake, which as previously noted is not publicly available. Delclaux et al. (2008) developed a hydrological model of the Lake Chad Basin, however the results they presented included streamflow and elevation, but not surface area.

Coe and Birkett (2004) used upstream measurement of river height along with in-situ stream flow and gauge height to estimate river discharge 500 km downstream and wet season height of Lake Chad, greater than 600 km downstream. Their method, though, clearly relies on hard-to-obtain in-situ measurements and does not include lake area.

The first objective of this paper is to assemble and examine a set of satellite- and model-based data for the southern Lake Chad Basin relevant to developing a statistical model for the area of the lake during its flooding season. This data set includes time series of satellite- and gauge-based precipitation and modeled ET for the southern part of the basin, modeled lake ET data, satellite-based lake elevation data, and satellite-based estimated lake total surface water area. Given the limitations of the existing models described above, the second objective is to develop a predictive statistical model for total lake surface water area using regression methods on the data set. The regression method includes backward elimination variable selection and a Leave One Out Cross Validation (LOOCV) analysis to optimize the resulting statistical model.

2 Study Area

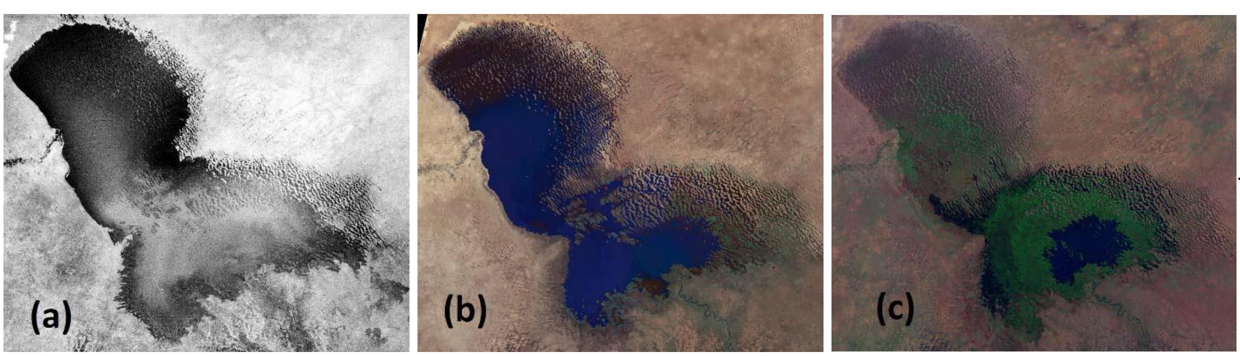
The Lake Chad basin is approximately 2.5 million square kilometers, about eight percent of the African continent, and the largest endorheic basin in the world (Gao et al., 2011). Lake Chad is the terminal lake of this basin. The northern part of the basin lies within the Sahara and does not generate runoff that reaches Lake Chad (Delclaux et al., 2008).

For this reason, we work with the southern part of the basin (figure 1).

Fig. 1 Southern Lake Chad Basin (white), lake, and major rivers (blue)

The lakeôs average depth varies between 1.5 a Any change in lake volume translates to a substantial change in lake shoreline and area Fund, no date, https://www.worldwildlife.org/ecoregions/at0904).

In the 1960ôs the lakeôs area was on the order -1980s its area was reported to be about one tenth of that size (Grove, 1996), though it is not clear if that includes flooded vegetation. If one includes is flooded estimated at close to 14,700 sq. km (Policelli et al., 2018). Figure 2 shows the evolution of Lake Chad from the time of the earliest space-based images of the lake.



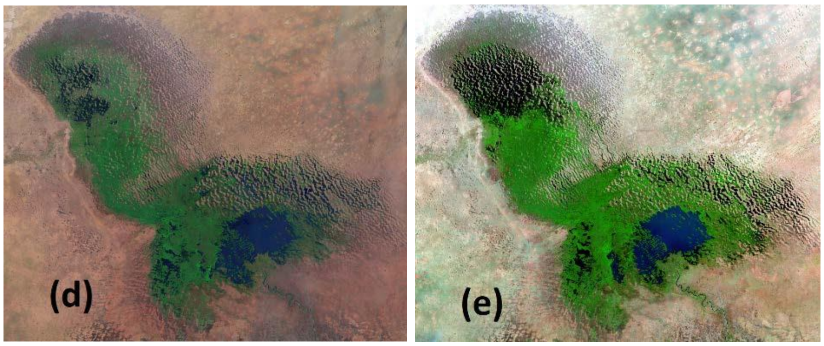


Fig. 2 Evolution of Lake Chad. Optical imagery from (a) 1963, Argon satellite (b) 1973, Landsat 1 (c) 1987, Landsat 5 (d) 2003, Landsat 7 (e) 2013, Landsat 8. Images provided by U.S. Geological Survey, Department of the Interior/USGS

Below about 280 m ASL, the lake separates into a southern pool and a northern pool divided by ñthe Great Barrierò, and the southern pool se sandy islands (Lemoalle, 2004).

The population of the lake shore is around 2 million (Magrin, 2016) and the people make a living through a combination of fishing, farming, and raising livestock (Sarch and Birkett, 2000). ñRecession farmingò is an important method of plant in the enriched soils following each yearôs floo the hydrology, it is difficult to provide farmers with information on the timing of the floods and a sense of how large the flood is going to be in any given year. This can be a serious problem for farmers who grow crops near the lake shore and periodically lose crops to flooding (Okpara et al., 2016).

According to (Sarch and Birkett, 2000), the south-west shore of Lake Chad begins in mid-January to late February. It seems reasonable to conclude that predictions of lake surface area made in late November or early December could be used for agricultural decision making for locations with similar start dates.

There are a number of dams and irrigation schemes in the river basins that drain into Lake Chad, such as those of the Chari-Logone and Komadougu-Yobe river systems. However, at the large scale, they do not amount to a substantial revision of the natural seasonal hydrological patterns, which are largely determined by the West African Monsoon and the position of the ITCZ (Birkett, 2000).

The surface of Lake Chad is not flat and level; the barriers to flow from the southern pool to the northern pool and archipelago result in the water in these areas frequently being at different elevations. Additionally, the local winds and flow of water from the Chari River into the lake (addressed further in the Discussion section) combined with the complex shape of the lake lead to an evolving lake surface (Carmouze et al., 1983). Lake surface elevation time series data used in this study refer to satellite-based measurements made in the relatively small open water portion of the lake in the southern pool.

Figure 3 presents areas of the lake at or below selected elevations. The topographic data used to create figure 3 is from a 1 arc-second (~30m) resolution Digital Elevation Model (DEM) from the NASA Shuttle Radar Topography Mission (SRTM). The blue areas would correspond to lake levels (limited by the accuracy of the DEM) if the lake surface were flat and the lake filled

7.6%, while [illegible] the average May [illegible] 13.1% [illegible] context for Lake Chad [illegible] Using the average May lake area of 13,363 sq. km, these percentages are equivalent to 1017 sq. km, 1751 sq. km, and 2085 sq. km respectively.

5. Discussion

The lake elevation data show that the maximum average monthly elevation typically occurs in November (though on occasion in October, and once during our record in January). It is curious therefore that the maximum area typically though not always occurs in February or March. The reason for this seems to be the complex meteorology and hydrology of Lake Chad. The movement of the water in Lake Chad is influenced by both the winds and the Chari-Logone water supply. Monsoon winds drive the displacement of the southern waters toward the north, and movement begins around the northeastern end of the Great Barrier in June, at the end of the low water. The Chari-Logone flood waters begin in August and provide half of their water in October and November when the northeasterly wind known as the Harmattan drives the water back toward the southern pool. This is also when the satellite radar altimeters (which collect data over the southern pool) typically record the highest levels. During the peak of the riverine flooding, water again reaches the northern pool and also spreads into the archipelago from the south basin. Following the end of the movement of water to the north pool in January, there is a general spreading of water in the southern pool to the periphery until April. (Carmouze et al., 1983). These movements result in a complex and changing lake surface topography, and are likely the reason we find poor correlation between lake elevation and lake area.

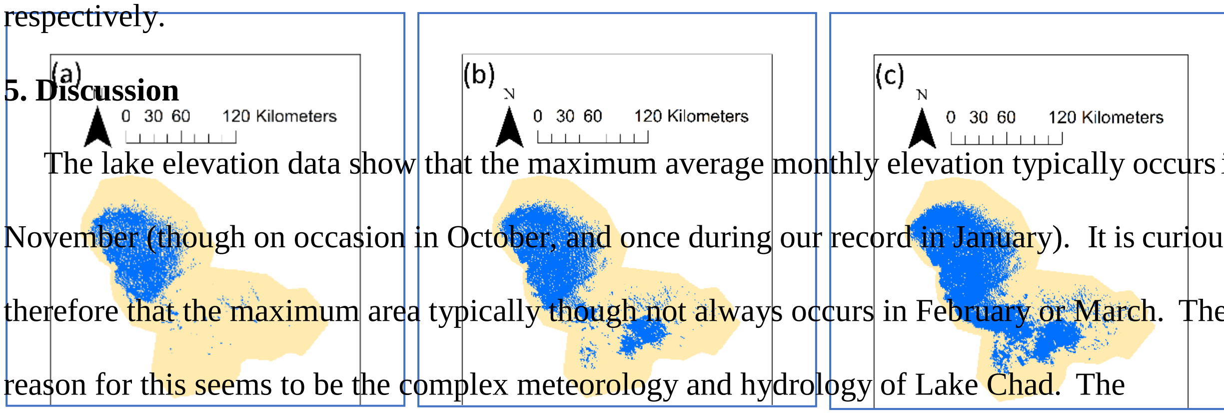


Fig. 3 Blue areas have elevations at or below (a) 279m ASL, (b) 280m ASL, (c) 281m ASL

According to (Leblanc et al., 2011) most of the flooded area of Lake Chad is covered by aquatic vegetation including rooted and floating plants. This area is not readily measured with optical remote sensing, but must be accounted for to get an accurate estimate of total surface water area for the lake (Policelli et al., 2018).

Lake Chad receives 90-96% of its water from the Chari-Logone River system (Zhu et al., 2017a), with the remaining coming from smaller tributaries and direct rainfall. Most of this water arrives from such a distance that the peak lake level and lake area occur months after the rainfalls that produce them. The delay is due to the slow runoff and routing of flood water to Lake Chad from the southern portion of the basin where precipitation rates are highest (Leblanc et al., 2011). It is the reason that the Lake Chad peak level and area take place in the dry season.

The main limitation of the model we present here is that it is not a physically-based model and may not perform well for conditions outside of those in the database we have developed. For example, we are not able to do regressions and provide predictions for wet season months because we do not have surface water area data for those conditions (Policelli et al., 2018). Also, the performance may degrade if it is used outside of the range of areas for which the regression models were established (8,700 sq. km ï16,800 sq. km) or when the hydrology of the lake changes substantially , such as when ñsmall Lake Chadò tr at about 18,000 sq. km. There is no fixed date at which the model becomes unusable. However, if it is used for operational forecasting, it would be wise to regularly update the model with new data as it becomes available.

6. Conclusions

We have built a record of remote sensing data and model products (precipitation, evapotranspiration, lake height, and lake area) for the Lake Chad Basin and used this record to run a correlation analysis and a regression analysis. From the correlation analysis (see Appendix), we have found (1) the highest correlation between basin evaporation and total lake surface water area is 0.43 and occurs with a seven month lag, (2) the highest correlation between lake height anomaly and total lake surface water area is 0.57 and occurs with a four month lag, (3) the highest correlation between precipitation and total lake surface water area is 0.39 and occurs with a seven month lag, (4) the highest correlation between percent of 1988-2016 average lake ET and the lake total surface water area is 0.65 at zero lag time, and (5) there is a correlation

The predictions using all of the equations derived from regression in this study can be made with only remotely sensed data and model outputs; no in-situ data is required. Any improvements in the measurement of the parameters we use in this analysis would likely improve the desired end result ï the prediction of the Lake Chad surface area in time to be used for agricultural decisions.

Acknowledgements

Support for this research was provided by the NASA Goddard Space Flight Centerôs Academic Investment for Mission Success (AIMS) Program. Support for this research was also provided by NASA under its Research Opportunities in Space and Earth Sciences (ROSES)-2013 Interdisciplinary Studies (IDS) Program (solicitation NNH13ZDA001N-IDS) through the Radiation Sciences Program. We appreciate the efforts of providers of the data products used for this study from various satellite sensors and models, particularly lake imagery from USGS, lake height data from USDA, CHIRPS precipitation data from U.C. Santa Barbara, evapotranspiration data from the FEWS NET Land Data Assimilation System (FLDAS), the southern Lake Chad Basin shapefile from WWF, meteorological data from NOAA NCEP-DOE, and Lake Chad surface water area from (Leblanc et al., 2011).

Role of the funding Sources: The funding sources had no role in the study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Author Contributions: Frederick Policelli and Ben Zaitchik conceived and designed the experiments; Frederick Policelli and Alfred Hubbard performed the experiments; Frederick Policelli, Hahn Chul Jung, and Ben Zaitchik analyzed the data; Charles ... helped conceive the experiments and provided domain knowledge; Frederick Policelli wrote the paper with contributions from all co-authors.

Declarations of Interest: None

References

Austin, P.C., Steyerberg, E.W., 2015. The number of subjects per variable required in linear regression analyses. J. Clin. Epidemiol. 68, 627–636. https://doi.org/10.1016/j.jclinepi.2014.12.014

Bader, J.-C., Lemoalle, J., Leblanc, M., 2011. Modèle hydrologique du lac Tchad. Hydrol. Sci. Journal – Journal Sci. Hydrol. 56, 411–425.

Badr, H.S., Dezfuli, A.K., Zaitchik, B.F., Peters-Lidard, C.D., 2016. Regionalizing Africa: Patterns of Precipitation Variability in Observations and Global Climate Models. J. Clim. 29, 9027–9043. https://doi.org/10.1175/JCLI-D-16-0182.1

Birkett, C.M., 2000. Synergistic remote sensing of Lake Chad: Variability of basin inundation. ... 72, 218–230.

Busker, T., de Roo, A., Gelati, E., Schwatke, C., Adamovic, M., Bisselink, B., Pekel, J.-F., Cottam, A., 2018. A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry. Hydrol. Earth Syst. Sci. Discuss. 1–32. https://doi.org/10.5194/hess-2018-21

Carmouze, J.-P., Durant, J.R., Lévêque, C., 1983. Lake Chad: ecology and productivity of a shallow tropical ecosystem. Springer Science & Business Media.

Coe, M.T., Birkett, C.M., 2004. Calculation of river discharge and prediction of lake height from satellite radar altimetry: Example for the Lake Chad basin. Water Resour. Res. 40, W10205. https://doi.org/10.1029/2003WR002543

Coe, M.T., Foley, J.A., 2001. Human and natural impacts on the water resources of the Lake Chad basin. J. Geophys. Res. Atmospheres 106, 3349–3356. https://doi.org/10.1029/2000JD900587

Delclaux, F., Le Coz, M., Coe, M., Favreau, G., Ngounou Gatcha, B., 2008. Confronting models with observations for evaluating hydrological change in the lake chad basin, Africa, in: XIIIth World Water Congress. URL: https://www.iwra.org/member/congress/resource/abs355_article.pdf accessed 7/18/2018

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations ð a new environmental record for monitoring extremes. Sci. Data 2, sdata201566. https://doi.org/10.1038/sdata.2015.66

Gao, H., Bohn, T.J., Podest, E., McDonald, K.C., Lettenmaier, D.P., 2011. On the causes of the shrinking of Lake Chad. Environ. Res. Lett. 6, 034021.

Grove, A.T., 1996. African river discharges and lake levels in the twentieth century. Limnol. Climatol. Paleoclimatology East Afr. Lakes Gordon Breach Neth. 95 ï100.

Harrell, F.E., Lee, K.L., Mark, D.B., 1996 Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors. Stat. Med. 15, 361 ï387. https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4

Lake Chad flooded savanna", World Wildlife Fund, no date, https://www.worldwildlife.org/ecoregions/at0904 accessed 7/18/2018

Leblanc, M., Lemoalle, J., Bader, J.-C., Tweed, S., Mofor, L., 2011. Thermal remote sensing of

and water security applications. Sci. Data 4, 170012. https://doi.org/10.1038/sdata.2017.12

Lehner, B., Grill G., 2013. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. Hydrological Processes, 27(15): 2171-2186.

Lehner, B., Verdin, K., Jarvis, A., 2008. New Global Hydrography Derived From Spaceborne Elevation Data. Eos Trans. Am. Geophys. Union 89, 93–94. https://doi.org/10.1029/2008EO100001

Lemoalle, J., 2004. Lake Chad: A Changing Environment. In: Nihoul J.C.J., Zavialov P.O., Micklin P.P. (eds) Dying and Dead Seas Climatic Versus Anthropic Causes. NATO Science Series: IV: Earth and Environmental Sciences, vol 36. Springer, Dordrecht

Lemoalle, J., Bader, J.-C., Leblanc, M., Sedick, A., 2012. Recent changes in Lake Chad: Observations, simulations and management options (1973–2011). Glob. Planet. Change 80, 247–254.

McNally, A., Arsenault, K., Kumar, S., Shukla, S., Peterson, P., Wang, S., Funk, C., Peters-Lidard, C.D., Verdin, J.P., 2017. A land data assimilation system for sub-Saharan Africa food

[illegible] under flooded vegetation: New observations of [illegible] patterns for the [illegible] Lake Chad. J. Hydrol. 404, 87–98. https://doi.org/10.1016/j.jhydrol.2011.04.023

Morton, F.I., 1986. Practical Estimates of Lake Evaporation. J. Clim. Appl. Meteorol. 25, 371–387. https://doi.org/10.1175/1520-0450(1986)025<0371:PEOLE>2.0.CO;2

Okpara, U.T., Stringer, L.C., Dougill, A.J., 2016. Lake drying and livelihood dynamics in Lake Chad: Unravelling the mechanisms, contexts and responses. Ambio 45, 781–795. https://doi.org/10.1007/s13280-016-0805-6

Policelli, F., Hubbard, A., Jung, H.C., Zaitchik, B., Ichoku, C., 2018. Lake Chad Total Surface Water Area as Derived from Land Surface Temperature and Radar Remote Sensing Data. Remote Sens. 10, 252. https://doi.org/10.3390/rs10020252

Ricko, M., Carton, J.A., Birkett, C.M., Cretaux, J.-F., 2012. Intercomparison and validation of continental water level products derived from satellite radar altimetry. J. Appl. Remote Sens. 6, 061710. https://doi.org/10.1117/1.JRS.6.061710

Sarch, M.-T., Birkett, C., 2000. Fishing and farming at Lake Chad: Responses to lake-level fluctuations. Geogr. J. 166, 156–172.

Satellite Radar Altimetry: Global Reservoir and Lake Elevation Database [WWW Document], no date URL https://www.pecad.fas.usda.gov/cropexplorer/global_reservoir/ (accessed 10.17.17).

Schneider, Udo; Becker, Andreas; Finger, Peter; Meyer-Christoffer, Anja; Rudolf, Bruno; Ziese, Markus (2015): GPCC Full Data Monthly Product Version 7.0 at 0.5 |: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data. DOI: 10.5676/DWD GPCC/FD M V7 050

Senay, G.B., Bohms, S., Singh, R.K., Gowda, P.H., Velpuri, N.M., Alemu, H., Verdin, J.P., 2013 Operational Evapotranspiration Mapping Using Remote Sensing and Weather Datasets: A New Parameterization for the SSEB Approach. JAWRA J. Am. Water Resour. Assoc. 49, 577 ±591. https://doi.org/10.1111/jawr.12057

Zhu, W., Yan, J., Jia, S., 2017a. Monitoring Recent Fluctuations of the Southern Pool of Lake Chad Using Multiple Remote Sensing Data: Implications for Water Balance Analysis. Remote Sens. 9, 1032. https://doi.org/10.3390/rs9101032

Appendix

We examined the correlation between P ±ET for the southern Lake Chad Basin and the lake elevation variation relative to 1993-2002 for 1992 to 2017, and found the maximum correlation of 0.69 at four months lag time between these variables. Next, we examined the correlation between the P ±ET for the southern part of the basin and the lake surface area and found the maximum correlation of 0.37 at eight months lag time. We also examined the correlation between the lake elevation variation and the lake surface water area and found the maximum correlation of 0.57 at four months lag time. We found a maximum correlation of 0.43 for ET vs WKH ODNH¶V VXUIDFH ZDWHU DUHD DW PRQWKV area DQG D FR YHUVXV WKH SUHYLRXV \HDU¶V VXUIDFH ZDWHU memory DUHD IRU W RI WKH SUHYLRXV system. \HDU¶V DUHD LQ WKH

To determine the value added by the ET data to our analysis, we examined the correlation between precipitation (without subtracting ET) for the southern Lake Chad Basin and the lake elevation and found an increase in the maximum correlation to 0.80 at four months lag time. We found the correlation between the percentage of the 1988-2016 average lake ET and the lake total surface water area to be 0.65 at zero lag time. We also examined the correlation between precipitation and total lake surface area and found an increase in the maximum correlation to 0.39 at seven months lag time. This is the same lag time as for the maximum correlation between ET and total lake surface area and represents the time it takes for much of the net precipitation to make its way from run off in the southernmost part of the Lake Chad basin, to flowing through the Chari-Logone River system, to reaching the lake and causing an increase in the lake area. A surprising result of the correlation analysis is that the use of FLDAS ET data in the analysis to produce (P-ET) causes a small decrease in the correlation numbers relative to what is achieved with precipitation alone. Note however that ET is somewhat more closely correlated with total lake surface water area than is precipitation (.43 vs. .39, both at 7 months lag time). Precipitation and ET have a maximum correlation coefficient of 0.93 with a one month delay between the two.