**Construction of the lakes and soils database and uncertainties associated with Late Pliocene lakes and soils.**

Late Pliocene lake and soil data was compiled from published literature (Supplementary table 1). This supplementary section will seek to explain the full databasing methodology, outline possible uncertainties and document specific uncertainties associated with the Late Pliocene mega-lakes.

**Database construction methodology**

Adding a new site to the database first involves producing a latitude and longitude of the locality. This does not present a problem even with older literature. Further to this, it is considered that continental drift from the Late Pliocene to the present day is minimal (Salzmann et al., 2008) and we therefore do not Palaeo-rotate localities.

In the Lakes and Soils database the absolute maximum age range has been recorded, taking into account all published errors and the published dating technique has also been recorded (Supplementary Table 1). By providing the method used by the original authors to date the site, it provides a qualitative means of assessing the accuracy of the dating. For example it is more likely that a radiometrically dated locality will have a more tightly constrained age than one dated through stratigraphic correlation. However, we would strongly advise caution about simply using this information from Supplementary Table 1, to discount or select data without reading the original publications. In an effort to provide maps suitable for future time-slice focussed palaeoclimate modelling and in the absence of the dating accuracy required for such maps, we have provided maps designed for a warm-humid (wet-lakes) and a cold-arid (dry-lakes) climate regime.

To make the soils database internally consistent we opted to use the USDA soil classification guide (Soil Survey Staff, 1999). This is a simple classification scheme and each of the soil orders (with the exception of Inceptisols and Entisols) can be associated with specific vegetation types (which facilitated producing a global distribution map using the Salzmann et al. (2008) data – model hybrid vegetation reconstruction as an additional data source) (Soil Survey Staff, 1999). The USDA soil classification scheme is a hierarchical classification with 12 soil orders, each of which can be further classified into suborders, groups, subgroups, families and series (Soil Survey Staff, 1999). The majority of paleosol localities were already reported in the USDA soil classification scheme, whereas a minority could be easily assigned to a soil order based on the original author’s identification or through their soil horizon descriptions.

Late Pliocene lakes required a size and surface area of the water body to allow a map suitable for palaeoclimate modelling to be produced. By either using numbers provided in the published text (e.g. Dodson and Ramrath, 2001), extracted from scaled diagrams of the palaeoenvironment (e.g. Tiercelin, 1986), calculated from scaled maps of reconstructed lake extent (e.g. Drake et al., 2008), calculated from scaled geological outcrop maps (e.g. Yeniyol, 2012) or through a combination of these. It is worth noting at this stage that if a lake extent could not be confidently calculated from the published literature it was not included in the database. An example of this would be the lake deposits reported from around the palaeo-Yukon River in north-west North America (Matthews et al., 2003; Pound et al., In Prep.). To facilitate mapping Late Pliocene mega-lakes, a north, south, east and west latitude – longitude point was recorded, to give a geographical indication of the lakes extent. This enabled the size and location to be more accurately translated into the gridded maps for the modelling study. As well as the size of the lakes, published details such as chemistry, the location of inflows and outflows, type of lake (e.g. evaporitic, mesotrophic etc.) and any evidence for orbital controls and specific events (see supplementary table 1).

Developing a dry-lakes scenario map required additional information to be taken into account. This ranged from utilising the published distribution of geological outcrops showing a shift from lacustrine to fluvial or sub-aerial sediments within the Late Pliocene (e.g. Schuster et al., 2009), the distribution of evaporite deposits (e.g. Sáez et al., 1999) or the original author’s palaeoenvironmental interpretation (e.g. Salama, 1987). The original author’s interpretation of a lakes response to climates provided the most rigorous data for generating the dry-lakes scenario. This either meant that the original author clearly stated that the lake was reduced to a certain size, or that it would have been absent. In the absence of a statement from the original author on how a lake would have responded to a drier climate we used the geological evidence presented in the paper. For example, if a Late Pliocene geological outcrop containing one or multiple layers of fluvial or sub-aerial sediments was within the reconstructed boundaries of a lake then at some point in the Late Pliocene the lake must have been smaller. These were then used as marker points to recalculate the lake dimensions and hence the surface area. This incorporates an element of uncertainty and with continued work more accurate information may be published on how many of these lakes responded to changes in Pliocene climate.

Each lake in the database represents a different sedimentary basin, the vast majority of which have no direct link to any of the other basins. As has been previously discussed in Peters and O’Brien (2001), our current geological understanding of these lakes means we cannot conclusively state that any of these lakes definitely co-occurred with any other.

**Specific uncertainties of the Late Pliocene mega-lakes**

In the manuscript we highlighted the specific uncertainties surrounding Mega-lake Zaire as it has the greatest uncertainty, in the following supplementary section we shall detail the assumptions and uncertainties of the other mega-lakes.

Mega-lake Fazzan

We calculated the extent of mega-lake Fazzan using the Late Pliocene regional reconstruction presented in Drake et al. (2008). This reconstruction is for humid periods in the Late Pliocene (Drake et al., 2008). In the absence of any published evidence on arid periods during the Late Pliocene we reduced the extent of mega-lake Fazzan by 50% for our dry-lakes scenario. It is possible that mega-lake Fazzan was absent during arid periods of the Late Pliocene as it was during arid periods of the Pleistocene, but there is no evidence to show this was the case (Drake et al., 2008).

Mega-lake Chad

For our wet-lake scenario we calculated the extent of mega-lake Chad based on figure 1 in Schuster et al. (2009) and as our calculated spatial extent lies within the previously calculated estimates (Ghienne et al., 2002; Leblanc et al., 2006; Schuster et al., 2005; 2009), although it is towards the upper estimate of Ghienne et al. (2002), we feel it is an appropriate estimate. To generate the extent of mega-lake Chad for our dry-lake scenario we used the geological column of Koro-Toro (Schuster et al., 2009), dated in Otero et al. (2010) as an approximate edge of the lake. An assumption based on the presence of fossil roots thought to represent plants growing in a seasonally wet and dry climate and the variability of the geology, from clay rich sandstone to lacustrine sediments, when compared to the massive pelites and diatomites further up the geological column. By locating Koro Toro on figure 1 in Schuster et al. (2009) we were able to use this as an approximate edge to the lake and recalculate the lakes extent for our dry-lake scenario, using the shaded relief image of the Chad Basin.

The greatest degree of uncertainty in the reconstruction of mega-lake Chad in the Late Pliocene is the assumption that it is comparable to the Holocene mega-lake Chad. Published geomorphological and geological evidence does not specify a size of mega-lake Chad during the Late Pliocene only confirming its existence (Otero et al., 2010; Schuster et al., 2009). In a recent modelling study on the Late Pliocene mega-lake Chad Contoux et al. (2013) used the lake extent of 350000 km2, which was presented in Schuster et al. (2009) as the extent of a mid-Holocene mega-lake Chad. Using the methodology described in the preceding paragraph we have calculated a larger extent (446760 km2), which is close to the maximum mid-Holocene size suggested by Ghienne et al. (2002) and comparable to the reconstruction presented in Goudie (2005). Published estimates for the maximum extent during the mid-Holocene range from 340400 – 448000 km2 (Ghienne et al., 2002; Drake and Bristow, 2006; Leblanc et al., 2006; Schuster et al., 2005; 2009). Griffin (2006) produced a reconstruction of a “Neogene” Lake Chad (using the present day 400m contour), which had an approximate spatial extent of 830000 km2. In both our present study and the work presented in Contoux et al. (2013) we make the same assumption: that the Late Pliocene mega-lake Chad could have been of comparable size to the mid-Holocene extent.

Mega-lake Sudd

Our reconstruction of mega-lake Sudd is based upon data presented in Salama (1987) and Berry and Whiteman (1968). A large lake is thought to have inhabited the Sudd Basin during humid periods of the Tertiary and Salama (1987) presents evidence to show deposition has been active in all parts of the basin since 18 Ma. It is thought to have been relatively shallow and so prone to considerable shrinking during arid periods (Salama, 1987). As the Sudd Basin is within the East African Rift System it is not only subjected to climate fluctuations, rifting and tilting will have influenced the size and duration of any lake (Goudie, 2005). We have attempted to represent this in our reconstructions by having a lake at maximum extent in the wet-lake scenario and almost absent in the dry-lake scenario. However, the presence of a lake is controversial with supporters including Lawson (1927), Ball (1939) and Salama (1987), whilst Berry and Whiteman (1968) dispute the existence of a lake.

Mega-lakes Okavango and Makgadikgadi

There is large uncertainty in the reconstruction of these lakes as they are based on the distribution of pre-Holocene alluvium presented in Ringrose et al. (2002; 2005), the known impoundment of the two systems during the Late Pliocene (Moore and Larkin, 2001; Partridge and Maud, 2000), the geomorphic chronology in Helgren (1984) and the acknowledgement of “Tertiary wet periods” by Butzer (1984). For the dry-lakes scenario and in the absence of any published data, we reduce the surface area by 50% to keep our methodology consistent with that used for mega-lake Fazzan and in light of the Pleistocene fluctuations in depositional type (Ringrose et al., 2005).

Mega-lake Eyre

The Lake Eyre basin is reconstructed in Alley (1998) as containing shallow ephemeral lakes that produced evaporite deposits, whilst Martin (1990; 2006) reports that the basin would have contained significant wetlands based on abnormally high Cyperaceae pollen values. Simon-CoinÇon et al. (1996) report that Lake Eyre would have stood higher due to the positioning of shallow slope glacis erosional features and extensive ?mid-Pliocene fluvial - lacustrine sediments. The seemingly contradictory evidence is likely to have been deposited at different geological times and the current dating evidence cannot successfully disentangle them (Alley, 1998). During a general trend of increasing aridity periods of warm – wet climate would have increased the lake size and facilitated the development of the wetlands (Alley, 1998; Benbow et al., 1995; Martin, 2006). Despite the dating uncertainties associated with the basin, it is known that the first sand dunes present were formed after the Late Pleistocene (Wopfner and Twidale, 1988) and the geographically widespread semi-arid fluvial Tirari Formation is older than 3.9 - 3.4 Ma (Tedford et al., 1992). This indicates that current levels of aridity were not reached during the Pliocene. The high degree of uncertainty around the timing of events during the Pliocene means that the extent of Lake Eyre in our reconstruction is based on the assumption that the higher water table associated with the geomorphological features of Simon-CoinÇon et al. (1996), the wetlands of Martin (1990) and extensive Late Pliocene lacustrine sediments in the Callabonna sub-basin (Alley, 1998) would have created a much larger lake, during wetter intervals in the Late Pliocene. With this evidence we used reconstructions presented in Alley, 1998), geological maps in Callen et al. (1998) and Simon-CoinÇon et al. (1996) to produce our wet-lake and dry-lake reconstructions. Due to the high level of assumptions, lack of definitive evidence and poor dating we consider mega-lake Eyre to have a high degree of uncertainty in our reconstructions.

Glenn’s Ferry lake

The reconstruction of Glenn’s Ferry lake in southern Idaho, USA is based on the Palaeo-drainage maps reported in Smith (1981), the occurrence of similar ostracod faunas in two Glenn’s Ferry Formation localities that are nearly 300 km apart (Swain, 1986a;b) and the literature review provided in Thompson (1992). A large deep lake is known to have existed in this region from around 6 Ma, and was finally drained at approximately 2 Ma when the Hells Canyon was formed (Malde, 1991). Prior to this draining event regressions of the lake extent are known to have occurred, but the lake was never absent due to the continuation of the fish fauna (Smith et al., 1982). We consider our wet-lakes reconstruction to have very little uncertainty, but our dry-lakes extent assumes a 50% reduction (in line with our treatment of Fazzan, Okavango and Makgadikgadi) so this has a higher degree of uncertainty.

Alley, N.F., 1998. Cainozoic stratigraphy, palaeoenvironments and geological evolution of the Lake Eyre Basin. Palaeogeography, Palaeoclimatology, Palaeoecology 144, 239-263.

Ball, J., 1939. Contributions to the Geography of the Nile. Government Press, Cairo, 308 pp.

Benbow, M.C., Alley, N.F., Callen, R.A., Greenwood, D.R., 1995. Geological History and Palaeoclimate. In: Drexel, J.F., Preiss, W.V. (Eds.), The Geology of South Australia. Volume 2. The Phanerozoic. South Aust. Dep. Mines Energy Bull. 54, 208–218.

Berry, L., Whiteman, A.J., 1968. The Nile in the Sudan. The Geographical Journal 134, 1-37.

Butzer, K.W., 1984. Archeogeology and Quaternary environment in the interior of southern Africa. In Southern African Prehistory and Paleoenvironments, Klein RG (ed.). Balkema: Rotterdam; 1–64.

Callen, R.A., Alley, N.F., Greenwood, D.R., 1998. Lake Eyre Basin. In: Drexel, J.F., Preiss, W.V. (Eds.), Bulletin 54: The geology of South Australia Volume 2 The Phanerozoic. Mines and Energy South Australia, pp. 188-194.

Chang, H., Ao, H., An, Z., Fang, X., Song, Y., Qiang, X., 2012. Magnetostratigraphy of the Suerkuli Basin indicates Pliocene (3.2 Ma) activity of the middle Altyn Tagh Fault, northern Tibetan Plateau. Journal of Asian Earth Sciences 44, 169-175.

Contoux, C., Jost, A., Ramstein, G., Sepulchre, P., Krinner, G., and Schuster, M.: Megalake Chad impact on climate and vegetation during the late Pliocene and the mid-Holocene, Clim. Past, 9, 1417-1430, doi:10.5194/cp-9-1417-2013, 2013.

Dodson, J.R., Ramrath, A., 2001. An Upper Pliocene lacustrine environmental record from south-Western Australia - preliminary results. Palaeogeography, Palaeoclimatology, Palaeoecology 167, 309-320.

Drake, N., Bristow, C., 2006. Shorelines in the Sahara: geomorphological evidence for an enhanced monsoon from palaeolake Megachad. The Holocene 16, 901-911.

Drake, N.A., El-Hawat, A.S., Turner, P., Armitage, S.J., Salem, M.J., White, K.H., McLaren, S., 2008. Palaeohydrology of the Fazzan Basin and surrounding regions: The last 7 million years. Palaeogeography, Palaeoclimatology, Palaeoecology 263, 131-145.

Ghienne, J.-F.o., Schuster, M., Bernard, A., Duringer, P., Brunet, M., 2002. The Holocene giant Lake Chad revealed by digital elevation models. Quaternary International 87, 81-85.

Goudie, A.S., 2005. The drainage of Africa since the Cretaceous. Geomorphology 67, 437-456.

Griffin, D.L., 2006. The late Neogene Sahabi rivers of the Sahara and their climatic and environmental implications for the Chad Basin. Journal of the Geological Society, London 163, 905-921.

Lawson, A.C., 1927. The valley of the Nile. University of California Chronicle No. 29.

Leblanc, M., Favreau, G., Maley, J., Nazoumou, Y., Leduc, C., Stagnitti, F., van Oevelen, P.J., Delclaux, F., Lemoalle, J., 2006. Reconstruction of Megalake Chad using Shuttle Radar Topographic Mission data. Palaeogeography, Palaeoclimatology, Palaeoecology 239, 16-27.

Malde, H.E., 1991, Quaternary geology and structural history of the Snake River Plain, Idaho and Oregon, in Morrison, R.B., ed., Quaternary nonglacial geology; Conterminous U.S.: The Geology of North America v. K-2, 251-281.

Martin, H.A., 1990. The palynology of the Namba formation in the Wooltana-1 bore, Callabonna Basin (Lake Frome), south Australia, and its relevance to miocene grasslands in central Australia. Alcheringa: An Australasian Journal of Palaeontology 14, 247-255.

Martin, H.A., 2006. Cenozoic climatic change and the development of the arid vegetation in Australia. Journal of Arid Environments 66, 533-563.

Matthews Jr. JV, Westgate JA, Ovenden L, Carter LD, Fouch T, 2003. Stratigraphy, fossils, and age of sediments at the upper pit of the Lost Chicken gold mine: new information on the late Pliocene environment of east central Alaska. Quaternary Research 60, 9-18.

Moore, A.E., Larkin, P., 2001. Drainage evolution in south-central Africa since the breakup of Gondwana. S. Afr. J. Geol. 104, 47 – 68.

Otero, O., Pinton, A., Mackaye, H.T., Likius, A., Vignaud, P., Brunet, M., 2010. The early/late Pliocene ichthyofauna from Koro-Toro, Eastern Djurab, Chad. Geobios 43, 241-251.

Partridge, T.C., Maud, R.R., 2000. The Cenozoic of southern Africa. Oxford University Press, New York (406 pp.)

Peters, C.R., O'Brien, E.M., 2001. Palaeo-lake Congo: Implications for Africa's Late Cenozoic climate - some unanswered questions. Palaeoecology of Africa 27, 11-18.

Pound, M.J., Lowther, R.I., Peakall, J., Chapman, R.J., Salzmann, U. Palynology, palaeoenvironment and age of the gold bearing Late Pliocene White Channel Gravels from Bonanza Creek, Yukon Territory. Manuscript in preparation.

Ringrose, S., Huntsman-Mapila, P., Basira Kampunzu, A., Downey, W., Coetzee, S., Vink, B., Matheson, W., Vanderpost, C., 2005. Sedimentological and geochemical evidence for palaeo-environmental change in the Makgadikgadi subbasin, in relation to the MOZ rift depression, Botswana. Palaeogeography, Palaeoclimatology, Palaeoecology 217, 265-287.

Ringrose, S., Kampunzu, A.B., Vink, W., Matheson, W., Downe, W.S., 2002. Origin and palaeo-environments of calcareous sediments in the Moshaweng dry valley, southeast Botswana. Earth Surface Processes and Landforms 27, 591-611.

Sáez, A., Cabrera, L., Jensen, A., Chong, G., 1999. Late Neogene lacustrine record and palaeogeography in the Quillagua–Llamara basin, Central Andean fore-arc (northern Chile). Palaeogeography, Palaeoclimatology, Palaeoecology 151, 5-37.

Salama, R.B., 1987. The evolution of the River Nile. The buried saline rift lakes in Sudan - I. Bahr El Arab Rift, the Sudd buried saline lake. Journal of African Earth Sciences 6, 899-913.

Salzmann U, Haywood A.M., Lunt D.J., Valdes P.J., Hill D.J., 2008. A new global biome reconstruction and data-model comparison for the Middle Pliocene. Global Ecology and Biogeography 17, 432-447.

Schuster, M., Duringer, P., Ghienne, J.-F., Roquin, C., Sepulchre, P., Moussa, A., Lebatard, A.-E., Mackaye, H.T., Likius, A., Vignaud, P., Brunet, M., 2009. Chad Basin: Paleoenvironments of the Sahara since the Late Miocene. Comptes Rendus Geoscience 341, 603-611.

Schuster, M., Roquin, C., Duringer, P., Brunet, M., Caugy, M., Fontugne, M., Taïsso Mackaye, H., Vignaud, P., Ghienne, J.-F., 2005. Holocene Lake Mega-Chad palaeoshorelines from space. Quaternary Science Reviews 24, 1821-1827.

Simon-CoinÇon, R., Milnes, A.R., Thiry, M., Wright, M.J., 1996. Evolution of landscapes in northern South Australia in relation to the distribution and formation of silcretes. Journal of the Geological Society 153, 467-480.

Smith, G.R., 1981. Late Cenozoic Freshwater Fishes of North America. Annual Review of Ecology and Systematics 12, 163-193.

Smith, G.R., Swirydczuk, K., Kimmel, P.G., and Wilkinson, B.H., 1982, Fish biostratigraphy of late Miocene to Pleistocene sediments of the western Snake River Plain, Idaho, in Bonnichsen, B. and Breckenridge, R.M., eds., Cenozoic Geology of Idaho: Idaho Bureau of Mines and Geology Bulletin v. 26, 519-541.

Swain, F.M., 1986a. Late Cenozoic non-marine ostracoda from southern Idaho. Revista Española de Micropaleontologia 18, 85-114.

Swain, F.M., 1986b. Freshwater ostracoda from the Cache Valley Formation (Pliocene) of southeastern Idaho. Revista Española de Micropaleontologia 18, 363-385.

Tedford, RH; Wells, RT and Barghoorn, SF. Tirari formation and contained faunas, Pliocene of the Lake Eyre basin, South Australia [online]. Beagle: Records of the Museums and Art Galleries of the Northern Territory, The, Vol. 9, Dec 1992: 173-193.

Thompson, R.S., 1992. Palynological data from a 989-FT (301-M) core of Pliocene and Early Pleistocene sediments from Bruneau, Idaho. USGS Open File Report 92-713, 1-28.

Tiercelin, J.J., 1986. The Pliocene Hadar Formation, Afar depression of Ethiopia. Geological Society, London, Special Publications 25, 221-240.

Wopfner, H., Twidale, C., 1988. Formation and age of desert dunes in the Lake Eyre depocentres in central Australia. Geologische Rundschau 77, 815-834.

Yeniyol, M., 2012. Geology and mineralogy of a sepiolite-palygorskite occurrence from SW Eskisehir (Turkey). Clay Minerals 47, 93-104.