Data 1SM

Summary of the research methods that supported the several research areas comprised in the 'Paleoclimate Project', which is the first-ever large-scale survey undertaken by the Instituto Tecnológico Vale in ULs of the Carajás region.

For geomorphological mapping of the lake basins, worldview-2 multispectral satellite imagery with 2 m resolution and LiDAR (Light Detection and Ranging) data were integrated to produce a digital elevation model (DEM) using PCI Geomatica 15. This dataset was used for interpretation of the morphological features and real-time navigation during field surveys, which were supported by a Differential Global Positioning System (DGPS).

The structural features were analyzed concerning their strike, dip, dip direction, length and kinematic in the weathered rocks exposed along the lake borders and hill slopes. The collected data were analyzed by stereographic projection in domains and spatial 3D blocks, and subsequently, integrated in larger views. The main lithotypes and soil samples were described and collected according to standard references [1, 9].

Bathymetric data and seismic profiling were conducted in the active lakes using a 200 kHz Raytheon DE719E and StrataBox™ SyQuest Echo Sounders, respectively. The bathymetric maps and digital terrain model were designed for planimetric map with 1 m isobaths of interval and 3D view to allow a precise visualization of the lake morphology [21]. Based on these maps, 145 surficial (0-10 cm) sediment samples were collected using a Van Veen Grab sampler. In addition, eleven sediment cores were collected using a Livingstone-type drive rod piston corer (operated over an adapted AIRE Leopard Cataraft) and a Russian peat borer on active and filled lakes, respectively (Table S1; Figure S1). Surface and core samples were described according to the classification system of lacustrine sediments of the Global Lake Drilling Program [22] and facies analysis [23].

Major and minor elements in sediments, soils and crusts were determined in pulverized samples by inductively coupled plasma optical emission spectrometry (ICP-OES) and X-ray fluorescence spectrometry (XRF), while trace elements by inductively coupled plasma optical mass spectrometry (ICP-MS). Quality control analysis and statistical treatment of the data were carried out following [24]. The geochemical dataset used in this study is comprised of surficial sediments from 5 active lakes and their catchment materials such as ferruginous and Al-enriched ferruginous crusts and soils, which were taken from [5, 9, 25], and core sediments from 11 inactive lakes, which were taken from [16, 17, 24, 26, 27]. A detailed description of sample type, sampling locations, chemical analysis, and quality control information of these dataset can be found in the respective references. This dataset was used for descriptive statistics (minimum, maximum, mean, median), normality test, Analysis of Variance (ANOVA) test, and multivariate analysis such as Principal Component Analysis (PCA). Prior to this statistical analysis, elements with > 70% of censored data were removed, and for elements having small proportion of censored data, the values below the lower detection limit (DL) were substituted by DL/2. The descriptive statistics (Table S3) show a high variability between mean and median for some elements, which is an indication of non-normal distribution. This can be supported by their higher skewness, kurtosis, and standard deviation values. The Kolmogorov-Smirnov (K-S) and Shapiro-Wilk's (S-W) test also indicated the non-

normal distribution of some elements as their as the values of α was < 0.05. Due to this issue, a nonparametric Spearman rank sum method was applied to understand the relationship between variables. Subsequently, PCA was carried out on centered log-ratio (clr) transformed data to reduce the non-normality behavior and closure issues in the dataset. Mann-Whitney U test, a non-parametric version of the ANOVA was performed to test significant differences of mean values of chemical elements in sediments among lakes. The null hypothesis for this test is that the mean elemental concentrations between lakes are not significantly different at a level of significance $p > 0.05$. All statistical analyses were carried out using SPSS software (version 18.0) and R statistical software.

The main vegetation types were defined during field observation and DEM interpretation (for detailed information about sampling method and palynological analysis, see standard reference [17]. Fertile material was incorporate in our pollen database (PALIITV). Thus, it helps to calibrate all pollen data of this work. In addition, artificial pollen traps installed along the drainage basin of the Serra Norte Lake allowed the evaluation of the annual pollen signal of very small canga areas surrounded by HETF and SDF [28]. Modern pollen rain data were also based on surficial sediment samples from active ULs [29, 30].

For $\delta^{13}C$ and $\delta^{15}N$ analysis, sediment samples were determined by an elemental analyzer coupled to a Continuous Flow Isotopic Ratio Mass Spectrometer (CFIRMS). Total sulphur (TS) and total organic carbon (TOC) were measured using a LECO CS-300 combustion analyzer (for more detailed information about sample preparation and data treatment, see [5, 16, 17, 25].

Around $1cm^3$ of sediment samples (from both surficial and core sediment) was collected at intervals of around 2.5 to 5 m depth for palynological analysis, which was based on classical techniques for palynomorphs extraction including acetolysis [31]. Identification were conducted by morphological comparison with PALIITV.

One hundred twenty-four bulk samples of about 1-2 g each were collected for radiocarbon dating (Table S2), which was performed by an Accelerator Mass Spectrometry (AMS). The age-depth model was made based on Bayesian accumulation (Bacon) histories for lake and peat deposits [32], using R as an interface and Intcal20 calibration dataset [33].

For limnological study, water samples were collected (surface, middle and bottom of the lake) in both rainy and dry seasons between 2013 and 2016 using Van Dorn water sampler. Sample collection and analysis of various inorganic, organic and bacteriological parameters were carried out by using Bureau Veritas analytical facilities following referred guidelines [34, 35]. In-site measurement of physico-chemical parameters were carried out using Horiba W-20XD multi sensor probe.