Use and abuse of detrital zircon U-Pb geochronology—A case from the Río Orinoco delta, eastern Venezuela

Mauricio Ibañez-Mejia1*, Alex Pullen2, Martin Pepper3, Franco Urbani4, Gourab Ghoshal5, and Juan C. Ibañez-Mejia6
1Department of Earth and Environmental Sciences, University of Rochester, Rochester, New York 14611, USA
2Department of Environmental Engineering and Earth Sciences, Clemson University, Clemson, South Carolina 29634, USA
3Department of Geosciences, The University of Arizona, Tucson, Arizona 85721, USA
4Escuela de Geociencias, Universidad Central de Venezuela, Caracas 1053, Venezuela
5Department of Physics and Astronomy, University of Rochester, Rochester, New York 14611, USA
6I. Physikalisches Institut, Universität zu Köln, 50937 Cologne, Germany

ABSTRACT
Advancements in mass spectrometry methods over the past two decades have allowed for rapid measurement of (U-Th)/Pb isotopes for geochronologic applications, a tool that has deeply influenced the way sediment provenance and paleo-tectonic reconstructions are approached. Geochronology-based studies of sediment provenance typically rely on dating n ≈ 100–150 single detrital zircon crystals from individual samples, where the sample age distributions are used to make inferences about the parent age distributions, make qualitative geologic interpretations, and/or perform quantitative intersample comparisons. Most efforts to quantitatively compare detrital zircon age spectra make use of non-parametric dissimilarity statistics. Here, we use laser ablation–inductively coupled plasma–mass spectrometry (LA-ICP-MS) U-Pb detrital zircon moderate-n (n ≈ 100) and large-n (n ≈ 1000) results from unconsolidated fluvial sediments of the Río Orinoco delta, eastern Venezuela, to highlight the concealed pitfalls of making geological interpretations based on quantitative comparisons of U-Pb age distributions alone. Three samples analyzed at large n, selected from contrastingly different mean sediment grain sizes along the active channel of the Río Orinoco, yield large dissimilarities amongst their age spectra, resulting in the misleading conclusion that these were likely not sourced from the same parent distribution. We demonstrate that statistically significant differences amongst detrital zircon samples derived from the same (integrated) source region can be introduced by the dynamics of sediment transport, which may in turn lead to erroneous geologic interpretations arising from the inaccurate assumptions that, at present, condition the quantitative treatment of detrital zircon data.

INTRODUCTION
“The fascinating impressiveness of rigorous mathematical analysis, with its atmosphere of precision and elegance, should not blind us to the defects of the premises that condition the whole process.”—Thomas C. Chamberlin (Chamberlin, 1899).

Most sediment provenance studies that apply quantitative detrital zircon U-Pb geochronology rely on the assumption that direct intersample comparisons using dissimilarity metrics allows for determination of similar and/or distinct sediment sources (e.g., Saylor and Sundell, 2016; Vermeesch, 2018, and references therein). There are, however, two inconvenient and commonly ignored issues in geochronology-based sediment provenance: (1) detrital zircon U-Pb ages can cluster by grain size (Sircombe et al., 2001; Garzanti et al., 2009; Lawrence et al., 2011); and (2) grain-size sorting is strongly influenced by the dynamics of sediment transport (e.g., Udden, 1914; Hajek et al., 2010). The size, shape, and density of detrital minerals are the primary physical properties that determine the settling and deposition of detritus entrained within active currents (e.g., Schuiling et al., 1985; Malusà et al., 2016). Thus, even if sediments are derived from the same catchment area, sedimentary rocks containing detrital zircon crystals with vastly different physical properties may appear to have distinguishable provenance characteristics based on comparison of detrital zircon age distributions alone.

To determine the provenance of sediments, geologists are often tasked with making geological interpretations based on comparison of samples collected tens to thousands of kilometers apart, sometimes across terrane boundaries and modern or ancient ocean basins (e.g., Rainbird et al., 1992). One presumption that has led workers astray in detrital mineral geochronology is the belief that if a sufficient number of age observations are made in a sample, then the measured distribution of ages will approach the “true” distribution present in a given sedimentary system with sufficient accuracy to make geologically sound interpretations (e.g., Pullen et al., 2014). This premise remains ignorant of how detrital minerals fractionate within sedimentary environments, and also presumes that biases introduced through sample selection and processing are negligible (Sircombe and Stern, 2002; Sláma and Kössler, 2012).

Here, we demonstrate the fractionation of detrital zircon age distributions within the natural fluvial system of the Río (River) Orinoco delta, eastern Venezuela (Fig. 1), and illustrate the issues with current quantitative approaches to detrital zircon age comparisons. We show that single samples, even if investigated at large numbers of single detrital zircon crystals (n ≈ 1000), are insufficient to adequately characterize the complete spectrum of detrital zircon ages available to a fluvial system from any given sample. In turn, our results demonstrate that physico-mechanical biasing (e.g., size and morphological sorting) can influence the distribution of ages found amongst samples deposited in contrasting energy environments within the same sedimentary system, thus affecting the accuracy of quantitative provenance interpretations built exclusively upon detrital sediments.

*E-mail: ibaneznm@rochester.edu

Zircon U-Pb dates. These findings are used to (1) point out pitfalls of current approaches and identify where the greatest gains can be made in the future, and (2) remind practitioners of this technique that “statistically significant” is not necessarily the same as “geologically significant” (e.g., Vermeesch, 2009).

GEOLOGIC SETTING AND DETRITAL ZIRCON GEOCHRONOLOGY

Uranium-lead (U-Pb) dates were measured for eight unconsolidated sediment samples from an active delta channel of the Río Orinoco (Fig. 1), the fourth-largest river on Earth by discharge volume of water. Zircon U-Pb dates were determined by laser ablation–inductively coupled plasma–mass spectrometry (LA-ICP-MS) following methods outlined in Pullen et al. (2014) (see the GSA Data Repository for results and analytical methods). The laser beam diameter for these experiments was 20 μm and thus defined the analytical limit on the zircon crystal size that could be dated. Samples were collected over a relatively small area (~26 km²) across the same channel, compared to the overall magnitude of the river length (~2.1 × 10⁶ km) and drainage basin area (~8.8 × 10⁸ km²). This narrowly defined sampling was employed to focus on the channel-perpendicular fractionation of detrital zircon ages (i.e., differences in zircon age distributions occurring along the “cross-sectional energy profile” of the channel), as opposed to the fractionation of detrital zircon age populations that would occur along the thousands-of-kilometers-long path of the river (i.e., channel-parallel fractionation). A moderate number of dates (n = 100) were determined for each of the eight samples (Fig. 2A). Three of these samples were selected for large-n dating based on their contrasting mean grain sizes (Figs. 2B and 2C); these samples include an overbank deposit (sample RORDVN05, mean zircon size = 75 ± 18 μm), a mid-channel sandy bar deposit (sample RORDVN02, mean zircon size = 130 ± 30 μm), and a coarse-grained mid-channel bed load sample (sample RORDVN03, mean zircon size = 220 ± 45 μm; Figs. 1 and 2B).

RESULTS AND DISCUSSION

The sediment load of a continental-scale river system like the Río Orinoco is expected to include detrital zircon crystals of multiple definable age components (or modes). Major age components from a sample (e.g., fractional abundance f ≥ 0.15) have a reasonably high probability of being recognized if n = 100–120 individual crystals are dated (Vermeesch, 2004; Andersen, 2005). Nevertheless, the relative proportions of those major age components at this level of observation are unlikely to be determined with accuracy, nor could that accuracy be assessed (Pullen et al., 2014). Additionally, there is a high probability that lower-abundance (e.g., f < 0.1), albeit potentially diagnostic, age modes would remain unobserved or poorly represented with moderate-n studies (n = 100).

Moderate-n U-Pb dating of detrital zircon crystals from all eight samples collected from the Río Orinoco shows the presence of multiple (>10) age modes within all samples, with three of these (i.e., ca. 1000, ca. 1400, and ca. 1500 Ma) representing >10 age modes within each sample (Andersen, 2005). Furthermore, because of the large intersample variability, it is similarly unrealistic to accurately determine the “true” distribution of zircon ages available to the bulk sedimentary system, or address whether the complete age spectrum could be determined from the analysis of a single sample even if a large number of single-grain observations were made (see discussion below).

To quantitatively assess the correlations that may exist among samples and/or evaluate the likelihood that a given set of samples could be drawn from the same parent distribution (i.e., the “null” hypothesis), most studies make use of non-parametric similarity-dissimilarity metrics for intersample comparisons (see Vermeesch, 2018, GSA Data Repository item 2013838, U-Pb geochronology methods, analytical results, PDF and KDE figures for all samples, dissimilarity matrices, and KS test results from bootstrap subsamples, is available online at http://www.geosociety.org/datarepository/2018/ or on request from editing@geosociety.org.)

GSA Data Repository item 2013838, U-Pb geochronology methods, analytical results, PDF and KDE figures for all samples, dissimilarity matrices, and KS test results from bootstrap subsamples, is available online at http://www.geosociety.org/datarepository/2018/ or on request from editing@geosociety.org.

Figure 1. A: Composite satellite image of Río Orinoco delta, eastern Venezuela (base images from Google Earth). Yellow box indicates location of panel B. B: Sampling localities across Río Orinoco delta; white circles and stars denote samples investigated using n = 100 and n = 1000 single detrital zircon crystals, respectively.

Figure 2. Cumulative density functions of samples analyzed in this study and graphical representation of upper-half portion of dissimilarity matrices showing results of Kolmogorov-Smirnov (KS) test and probability density function (PDF) cross-correlation (R²) values. Numbers bordering the matrices are sample numbers. A: Results of eight samples analyzed using n = 100 single detrital zircon crystals. B: Representative transmitted light microscope images (left) and scanning electron microscope–backscattered electron images (right) of zircon crystals from three samples investigated at n = 1000; mean short-axis and long-axis lengths of zircons found in each sample are shown below each panel. C: Results of three samples analyzed at n = 1000. See Data Repository (see footnote 1) for PDFs and kernel density estimate plots of all samples, as well as complete dissimilarity matrices.
and references therein). Traditionally, most researchers apply a two-sided Kolmogorov-Smirnov (KS) or Kuiper test to the cumulative density functions (CDF), and use the test’s D- and/or p-values to quantify dissimilarity and/or to reject (or fail to reject) the null hypothesis, respectively (e.g., Berry et al., 2001; Lawrence et al., 2011; Laskowski et al., 2013). Applying a two-sided KS test to our moderate- to Orinoco data (Fig. 2A) indicates, at 95% confidence, that only about half of the possible sample pairs fail to reject the null hypothesis. This is counter to the geological observations conditioning this experiment, as failure to reject the null hypothesis is the a priori expectation based on the knowledge that all analyzed sediments were collected from the same channel of this active drainage basin (Fig. 1; see also Bonich et al., 2017).

Recent studies have suggested that the KS and Kuiper tests are poorly suited as quantitative descriptors of dissimilarity at low observation numbers (i.e., <1000 for KS and <475 for Kuiper), and instead have suggested that cross-correlation coefficients (R² values) of probability density functions (PDFs) might be better suited for this purpose at moderate-n values, even if they are not suitable for hypothesis testing (Saylor and Sundell, 2016). Application of the PDF cross-correlation test to our moderate-n data set also reveals, as the KS test did, the differences amongst the measured spectra, as 86% of all possible intersample pairs (n = 24) yield R² values <0.60, and 36% of pairs (n = 10) yield R² values <0.40 (Fig. 2A); nevertheless, due to the low sampling numbers, these coefficients from PDF cross-correlations cannot reliably discriminate identical from different populations (Saylor and Sundell, 2016).

The statistical tests described above using moderate-n results have inherently low precision (due to random sampling; Andersen, 2005), but application to samples with a larger number (n = 1000) of single-crystal U-Pb observations can be considered statistically accurate (Saylor and Sundell, 2016). Application of the KS test to the large-n data set yields large dissimilarities and indicates that the null hypothesis can in all cases be rejected with >99.99% confidence (i.e., p-values <1 x 10⁻⁴). On the other hand, PDF cross-correlation coefficients are 0.30 and 0.38 for comparison of the large-grained sample with the small- and medium-grained samples, respectively, whereas comparison of the medium- and fine-grained samples yields a coefficient of ~0.76 (Fig. 2C). Low R² values of 0.30 and 0.38 indicate that, statistically speaking, intersample comparisons between the large- and the small- and medium-grained samples both confidently allow rejecting the (null) hypothesis that these were derived from the same parent distribution. The R² of ~0.76, on the other hand, would be indicative of mixing from different sources, but cannot confidently be used to reject or fail to reject the hypothesis that these sediments were derived from the same source or parent distribution (see Saylor and Sundell, 2016, their figure 12, for interpretation).

Although the tests described above—particularly for our large-n data set—can be considered statistically accurate, they are unlikely to result in geologically accurate conclusions; many researchers would interpret these statistical observations to indicate derivation from different sources or “parent” distributions. D-values in the KS and Kuiper tests and R² values in PDF cross-correlations are strongly controlled by the presence or absence and the abundance of particular age modes that compose the spectra, both of which can be strongly controlled by natural sorting processes and are thus not necessarily indicative of different provenance, source areas, or tectonic phenomena (e.g., Fig. 2C). The calculated dissimilarities amongst samples as described above are arguably not the result of derivation from different source areas (i.e., we know they derive from the same modern drainage basin), but instead arise from the systematic sorting and biasing experienced by detrital zircon crystals during transport. We conclude, then, that the critical disadvantage that any generalized dissimilarity approach has for comparing detrital mineral geochronologic data lies in that such tests remain ignorant to, and uninformed by, the underlying geology.

To further illustrate the limitations that traditional n = 100 geochronologic data sets have at accurately representing (graphically and quantitatively) detrital zircon provenance, we performed a sensitivity test using random sampling methods. Starting with the measured large-n data sets for each of our three measured samples (Fig. 2C), 100 subsamples of n = 100 dates were drawn from each of these “parent” distributions using bootstrap without substitution. Shown in multidimensional scaling space using KS D-values as dissimilarity estimator (KS-MDS; see Vermeesch, 2018), these random subsamples plot as “clouds” around their respective n = 1000 “parents” (Fig. 3A). The “parent” age distribution for our large-grained sample is so dissimilar to that of the medium- and small-grained samples that their mutual random subsamples do not overlap in KS-MDS space.

On the other hand, random subsamples of the medium- and fine-grained samples show a greater degree of overlap; this implies that there is greater probability that random combinations of n = 100 subsamples from samples RORDVN05 and RORDVN02 could yield KS distances that fail to reject the null hypothesis and thus result in statistical type II errors (or false positives). In other words, even though we know from the KS test using n = 1000 data that the sample distributions of samples RORDVN05 and RORDVN02 are statistically distinct (Fig. 2C), if different researchers take samples from the exact same locations and analyze equivalent samples at n = 100, there is a large probability that they would arrive at geologically contradictory conclusions. To illustrate this, Figure 3B shows the distribution of D- (red curves) and p-values (blue curves) that would result from comparison of all random subsamples amongst our three large-n data sets, plotted as kernel density estimates (KDEs) of all possible combinations of two-tailed KS tests amongst subsamples of two given large-n samples. Whereas comparison amongst the n = 100 random subsamples of sample RORDVN03 with respect to RORDVN02 and RORDVN05 results

![Figure 3](https://pubs.geoscienceworld.org/gsa/geology/article-pdf/46/11/1019/4527714/1019.pdf)
in low probabilities of type II errors (i.e., $p$-values only fail to reject the null hypothesis in 3%–9% of the cases, at 95% confidence), subsample comparisons amongst samples RORDVN02 and RORDVN05 would result in type II errors in 83% of the instances (at 95% confidence).

The experiments shown here pose the uncomfortable problem: If the questions we formulate in such a simple experiment from an active fluvial system fail to be accurately answered using current data evaluation tools, what confidence do we have in our quantitative detrital zircon provenance interpretations from the ancient geological record? It is becoming increasingly clear that in order to resolve this, simply dating more crystals per sample is not the only answer. Although larger values of $n$ are evidently fundamentally different grain sizes within that system will better approximate the complete zircon age provenance and ensuing geological processes.

ACKNOWLEDGMENTS

We thank D. Mendi for assistance in the field; N. Perez, M. Eddy, and L. Grohn for comments to an earlier version of the manuscript; A. Satkoski and S. Samson for constructive reviews; and J. Schmitt for editorial handling. This research was funded by the U.S. National Science Foundation (grants 1545859 and 1649254), the University of Rochester (New York, USA), and Clemson University (South Carolina, USA).

REFERENCES CITED


Printed in USA