**Some GeoData-Mining goals of the extensive *Earth-history* data sets that have been compiled at Purdue– PhD program concepts for "Andy" Abdullah Khan Zehady** *(based on discussions and readings over the past two years with Professor Ogg)*

**Background**

Many researchers of Earth’s history have suggested that there are major upheavals in the past with uncertain cause-effects (asteroid impacts, massive volcanic eruptions, mass extinctions, swings in global climate and sea-level, disruptions of the planetary carbon cycle, etc.). How to test these speculations? Therefore, the NSF EarthCube workshop on stratigraphy (Chen and Budd, NSF white paper) emphasized the need for “*Analyzing the sedimentary record to identify and understand forcing factors, feedbacks and tipping points, and the resultant impact on the deep-time Earth system. Develop a deeper understanding of the interplay between life, the physical and chemical environment in Earth’s past*.” which had the research challenge of “*Development of geochronological tools that provide more precise and accurate timing of events in Earth’s history*” including “*data discovery, coupling of diverse datasets, and ability to search/query both observational and interpretative data*” and “*visualization software for stratigraphic columns, timescales and other data (biostratigraphic, chronological, geochemical, etc.).”.*

These type of databases and tools envisioned by “EarthCube” projects were partly begun at Purdue when Dr. Ogg was coordinator (2008-2012) of the *Subcommission for Stratigraphic Databases* for the *International Commission on Stratigraphy*, and have been progressively enlarged during the scope of the Geologic Timescale 2012 and 2016 community compilations (Gradstein et al., 2012; Ogg et al., 2016) and their joint projects with the Integrated Ocean Drilling Program, Smithsonian, British Geological Survey, Geoscience Australia, Chinese National Stratigraphic Commission, and other groups.

Most of the published interpretations of cause-effect correlations among geological-climate-biological-chemical records are based on focusing on a local “curious” event, then looking in selected literature for “hypothesis testing” to compile a “global picture” based on only a very few selected actual geological studies. One speculative example is the widely cited correlation of “*Large igneous provinces and mass extinctions*” by Paul Wignall, which omitted many major volcanic eruptions and the actual record of faunal diversity in different groups.

Alternatively or in addition, there may be major periodic cycles in Earth’s history. Some have suggested a 62-million-year “disaster cycle” in the evolution of life (e.g., the widely publicized “Nemesis” papers by Adrian Melott, which have even been partly based on our Purdue databases), while others have suggested multi-million-year changes in Earth’s climate due to orbital-eccentricity changes (e.g., “*Long-period astronomical cycles*” by Masayuki Ikeda, 2013; and ca. 5-myr-“orbiton” frequency papers by the late geophysicist, Robert Matthews of Brown Univ.). Major trends in rock types and evolution (or other Earth events) is another area of speculation (e.g., the NSF-supported “MacroStratigraphy” website by Peters).

But all of these studies are based on a rather subjective examination of the geological record and partial “fitting data to a hypothesis”. What is needed is a “mining” and realistic statistical evaluation of Earth history data to see when there are synchronous anomalies, when there are long-term oscillations, what are the main trends, and what are statistically-significant changes in rates.

The *TimeScale Creator* databases have been progressively enlarged over the past ten years, and now should enable testing of some of these speculations, plus selecting geodata-mining to find other potential anomalies, periodicities and significant trends. An important feature that is lacking in all other datasets (e.g., Peters 2011; Ikeda 2013; Melott; etc.) is that NONE have been carefully enhanced and calibrated to the same time scale! That is a major defect in past “data-mining” when the datasets being compared are using different standards.

Our geo-data sets compiled and standardized to the same time scale over the past decade currently include the evolution of life (ranges of about 30,000 major marine life forms spanning the entire animal evolution through 600 million years), significant new features in evolution of different types (“zone” boundaries), reversals of the Earth’s magnetic field, swings in carbon cycle and temperature, and changes in rock types for nearly half of the Earth’s major basins. For the past five years, we have worked to calibrate these to the same time scale, including uncertainties.

There is no comparable data set for all types of features and changes in our planet’s history; and our goal is to use this growing suite to test hypotheses that have been proposed by other publications. In addition, we hope that processing of these datasets might yield other insights to be further investigated; because many major scientific discoveries are the result of observations for other purposes or unanticipated developments, rather than from hypothesis testing.

**Some potential questions (and hypothesis testing)**

1. Hypothesis proposed by others to be tested: “***The Earth has had semi-periodic episodes of unusual surface/biological change.***” Method: Simple “*how abundant are events with time*” – Load ALL the databases (ca. 500,000 entries at this point), then exclude the “curve” type (because those points are only hand-entered to capture the published trends). *How many events are occurring in each 1-million-year window for the past 600 million years?* This will be a fascinating output: (a) what intervals are relative upheavals in Earth’s history and which intervals are anomalously “quiet”, (b) do these oscillations support the “Nemesis” or other periodic catastrophe models (ca. 20 to 60 million-year periods if one does an evolutive spectra analysis?), (c) are there suggestions of the postulated long-term (ca. 2 to 5 myr) astronomical cycles, and (d) which peaks in events correlate with peaks in volcanic activity, continental collisions and/or major asteroid impacts? What happens if we run 2 or 5 million year sliding windows instead? What would be the projected main changes for the next million years (our own future) based on any periodicities and trends?
2. Hypothesis proposed by others to be tested: “***Pulses of biological evolution occur simultaneously with global changes in sediment facies***.” Method: Proportion of Sediment types in basins: “*what are main trends with time and with space*”, “*do these support the suggested changes of sediments corresponding to evolution (Peters 2011)*”, “*do we see a correlation with major interpreted climate shifts and continental collisions*” (which are sets within our databases), “*are there any significant correlations with other datasets for sliding window-intervals of time*”. Essentially, we wish to get a first-hand view of what might be the main factors affecting Earth’s rock record. Sliding windows will be one method; but, in contrast to #1, the rock record is continuous, therefore we’d extract the sediment types in each column and determine the regional and global proportions of carbonate, clay, sand and other major types. In turn, these will help guide understanding of why our main oil-gas-coal deposits are concentrated in certain time periods.
3. Hypothesis proposed by others to be tested: “***Rates of evolution are correlated to Rates of geochemical and sea-level change.***” Method: Rates of change in Earth’s geochemical cycles: We have compiled and re-scaled records for temperature, sea-level (long/short trends), carbon cycle, and strontium (a measure of spreading rates of oceans and continental erosion). No one has really examined *whether RATES OF CHANGE of these are interconnected*, but only a visual comparison of main trends and “excursions”. *Is there any correlation between such rate changes and the frequency of volcanic eruptions, continental collisions, evolutionary pulses* (from #1 above) *or rock record shifts* (from #2 above). What would this imply about the response of Earth’s processes (rates) to our current human interference?
4. Hypothesis proposed by others to be tested: “***A major factor in the rise and fall of human civilization in different continents is climate change .***” Method: Human history: One of our data sets is the record of all major civilizations, dynasties and other cultural shifts for all continents for the past 10,000 years. Are there any “times of high frequency of events” in these datasets? We also have ice-core records of temperature and carbon-dioxide levels, plus volcanic eruption histories. Can one identify whether natural processes and climate change (and rates of these changes, not just shifts) are playing an important role in human-civilization upheavals and trends? For example, anomalous global cooling events and resulting reduced regional rainfalls have been implicated in both the rise and the collapse of the Egyptian pyramid building dynasty; but are there other trends/features that are subtle but above the noise level?

These are just some preliminary data-mining questions. Once we have a system to apply a variety of analytical techniques to the entire set of databases (the ca. 300 “internal” suite, plus the ca. 20 separate multi-column ones that contain up to 500 or more columns in some instances = ca. 500,000 events/datapoints) in a flexible system for different types of data mining, then we can begin to run other cross-correlations to reveal other aspects and test other hypotheses.

I’ve only included a brief sample from recent publications on related speculations; and several later publications have referenced these interesting “what is governing Earth’s history?” ideas as if they are fact. But, no one had rigorously tested them, because they have not had adequate large datasets, such as we’ve compiled.

**Sample references:**

Chan, M., and Budd, D.A., 2014. Executive summary: EarthCube sedimentary geology workshop. NSF white paper.

Ernst, R.E., and Youbi, N., 2017. How large igneous provinces affect global climate, sometimes cause mass extinctions, and represent natural markers in the geological record. *Palaeogeography, Palaeoclimatology, Palaeoclimatology*, 478: 30-52.

Gradstein, F.M, Ogg, J.G., Schmitz, M.D., Ogg, G.M. (coordinators), 2012. *The Geologic Time Scale 2012*. Boston, USA: Elsevier, 2 volumes plus chart, 1176 pp.

Ikeda, M., and Tada, R., 2013. Large period astronomical cycles from the Triassic to Jurassic bedded chert sequence (Inuyama, Japan); Geologic evidences for the chaotic behavior of solar planets. *Earth Planets Space*, 65: 1-10.

Matthews, R.K. and C. Frohlich 2002. Maximum flooding surface and sequence boundaries: Comparisons between observation and orbital forcing in the Cretaceous and Jurassic (65-190 Ma). *GeoArabia*, 7, no. 3, p. 503-538.

Matthews, R.K. and M.I. Al-Husseini 2010. Orbital-forcing glacio-eustasy: A sequence-stratigraphic time scale. *GeoArabia*, 15, no. 3: 155-167.

McGowan, A.J., and Smith, A.B. (eds), 2011. Comparing the geological and fossil records: implications for biodiversity studies. *Geol. Soc. London Spec. Publ*., 358.

Melott, A.L., and Bambach, R.K. 2010. Nemesis reconsidered. *Paleobiology*. arXiv:1001.3449.

Ogg, J.G., and Deconinck, J.F. (2013). Chemostratigraphy, Magnetostratigraphy, Chronology, Palaeoenvironments and Correlations: Overview. *Ciéncias de Terra* (UNL), 18: 69-72.

Ogg, J.G., Ogg, G.M., and Gradstein, F.M., 2016. *The Concise Geologic Time Scale 2016*. Elsevier, ca. 234 pp.

Peters, S.E., and Hein, N.A., 2011. Macrostratigraphy and macroevolution in marine environments: testing the common-cause hypothesis. *Geol. Soc. London Spec. Publ*, 358: 95-104.

Wignall, P.B., 2001. Large igneous provinces and mass extinctions. *Earth-Science Reviews*, 53: 1-33.