Biodiversty accross scales

Research proposal by Juan Escamilla

==Pete:== My intention here is to give the necessary tools and motivations for the work disregarding the overall motivation of why it's important to study biodiversity and why its change through scale and space; which will be included in the introduction.

==I think this can be used in the methodology==

0.1 Motivations

A model is an abstraction of reality (Smith and Smith, 2007). With the appropriate treatment, it can be embedded into a well founded mathematical theory. Mathematical theories rely on the fact that are formal systems and are developed with precision and rigour to be consistent and complete ==Need good reference==. Translating the real object into a mathematical object allows the inheritance of its properties, consequences and consistencies to analyse, explain and expect behaviours. Always limited by the context and scope bounded by the mathematical theory and the abstraction process.

0.2 Foundations

In order to give the model a state of formality, it is necessary to build a clear framework based on assumptions, hypotheses and axioms. In this work, an *assumption* is considered to be any limiting fact that depends on the reality of the problem. For instance, any issue concerning the quality of the input data. E.g. *bias in sampling, erroneous classification, non existent data, etc.*

An *axiom* is any proposition that cannot be proof because of: i) it is *self-evident* (e.g. life in Earth has a spatial component) or ii) there is a scientific theory[1] that supports it as valid and for the purposes of this work it is not necessary to proof it. E.g. *All organisms have a common ancestor.*

---

[1]In this case the Theory of Evolution

A *hypothesis* is any proposition that will be proven in this work. The hypothesis will make use of the axioms, assumptions, other hypothesis and results from other works, namely references.

0.2.1 Research questions

This section is still an outline.

One of the oldest questions in ecology has been: *How does composition of communities occur?* This question has been tried to be answered describing the community in terms of the diversity of organisms that inhabit an area and historically the unit of choice for measuring it has been the Species (Wilson, 2010). Pavoine and Bonsall (2011) demonstrated how species, trait and phylogenetic diversity can be combined together from large to local spatial scales to reveal the historical, deterministic and stochastic processes that impact the compositions of local communities. These structures had revealed essential details that might act simultaneously in the assembly of species communities. Details that can serve for answering questions about the distribution of the phylogenetic structures through different biomes; The fragmentation and symmetry of these structures through scale and space; and the loss of species and phylogenetic diversity driven by human impacts.

Therefore the questions addressed in this work are:

- How does biodiversity, together with its structure differ from biome to biome? Are there similarities between taxonomic levels?

- How do bio-climatic factors affect the distribution and structure of biodiversity through biomes, space and spatial resolution?

- Is there any fragmentation (or continuity) thresholds within the taxonomic structure?

- How do human populations affect biodiversity and which taxonomic level is most influenced by?

0.2.2 Formalisms

In order to answer these main research questions it is necessary to develop a common language without ambiguities and to the maximum extent as possible with mathematical formalism.

0.2.2.1 Axioms

Axiom 0.1 (Common Ancestor)**.** *Every two living beings (organisms) in Earth are phylogenetically related. It means that there is a relationship of ancestry. This relationship implies that in the past they belonged to the same species.*

This axiom guarantees that every pair of organisms are comparable in the evolutionary sense. The former axiom can be changed to satisfy other type of relationship between organisms only when this relationship makes the organisms comparable as well.

Axiom 0.2 (Uniqueness of Common Ancestor)**.** *There is only one most recent common ancestor for every two organisms.*

This axiom implies that it is not possible to have two or more immediate common ancestors. In this sense speciation, the evolutionary process in which a new species arise, is a bifurcation from a pre-existant specie. Also this axiom prevents ambiguity in lineages. i.e A species has one and only one linage (taxonomic chain).

Axiom 0.3 (Last Universal Ancestor (LUA))**.** *There is a universal ancestor of all living species in Earth. This is the most recent organism from which all living organisms in Earth descends[2].*

This axiom states that there is a minimal element in the set of all organisms and as such every organism in Earth is related to this primordial living being.

Axiom 0.4 (Taxonomic relationship)**.** *The hierarchical ordering of:* kingdom, phylum, class, order, family ,genus *and* species *is based on the* Natural Method *of biologic classification (systematic). Its main objective is to preserve ancestry. Therefore a taxonomic relationship is a relationship of ancestry that groups organisms with likelihood of being descendant from a common ancestor.*

This is the link between the systematic taxonomy and the evolutionary relationships. The classification is a continuous process performed by experts in different fields and it is always subject to modifications.

Axiom 0.5. *Life happens in Earth and has a spatial component.* ==*(You've used a better phrase but I don't remember)*==

This is the link between biology and spatial analysis.

Axiom 0.6 (Life is conspicuous)**.** *Life is conspicuous in all the surface of the Earth. Meaning that any area (with reasonable size) on the surface of the Earth contains a non-empty set of living beings.*

---

[2] It is estimated that this organism could had lived in the Paleoarchean Era (3.8 -3.5 billion years ago) (Glansdorff et al., 2008).

This is a strong assumption for the model. Its main purpose is to ease the analysis and model's develop by supposing that the *living phenomenon* is continuous in the surface of the Earth. A more interesting but complex to formulate (from the mathematical point of view) is the use of a probabilistic space <mark>pending for results in first round</mark> See (Amari and Nagaoka, 2000) Information geometry SUPER INTERESTING!! .

0.2.2.2 Assumptions

These assumptions are some of the many that could be involved in the modelling. e.g. stationarity when correlating with bioclimatic variables. <mark>This section will be prone to modifications until the end of the results.</mark>

Assumptions for the GBIF-data:

- The GBIF database is valid for the axioms proposed in the above section. Therefore the data extracted from it can be used in the model.

- The classification (taxonomy) of occurrences in the database is accurate and preserve phylogenetic relationships (common ancestry) [This is a hard assumption because there is not enough evidence to support this for all described species. Nevertheless, the higher the level the more likely to be true but depends a lot on each group.]

- The analysis will not discriminate synonym species. Different names for the same specie will be considered as different species.

0.2.2.3 First definitions

Definition 0.7 (Biologic Species). : Two organisms that in natural conditions are able to reproduce and have fertile offspring are members of the same class. This class is called biologic specie.

Definition 0.8. Let $V(G)$ be a set and $E(G) \subseteq V(G) \times V(G)$. A graph G is duple given by $(V(G), E(G))$. $V(G)$ is the set of vertices of the graph and $E(G)$ is the set of edges.

Definition 0.9. Let G be a graph. G' is a subgraph of G ($G' \subseteq G$) if and only if $V(G') \subseteq V(G)$ and $E(G') \subseteq E(G)$.

Definition 0.10. If for every $u, v \in V(G)$ there exist a path that connects them then G is say to be connected. If that path is unique for every u, v then G is acyclic (without cycles).

Definition 0.11. A tree T is a connected and non-cyclic graph.

Definition 0.12. Let T be a tree. A subtree T' is a subgraph of T such that is also a tree.

Every tree follows a hierarchy given by a partially order set. The Taxonomic Tree is a tree given by the hierarchy of taxonomic relationship.

0.2.2.4 First theoretical results

Lemma 0.13. *There is a unique Taxonomic Tree of all life on Earth. This tree is called The Tree of Life.*

Proof. All organisms have Common Ancestor. Because of this is possible to build taxonomic relationships based on this comparison. The Uniqueness of this common ancestor and the existence of LUA implies that: i) there is just one path that connects any pair of species (vertices) and ii) the graph is connected. □

Lemma 0.14 (Local Tree)**.** *For any area in Earth it is possible to derive a unique Taxonomic Tree.*

Proof. Because Life is Conspicuous it is possible to find organisms in any place. By the axioms of Common Ancestor and Taxonomic Relationship it is possible to build a taxonomic hierarchy between the group of organisms within that place. Because Axiom of LUA there is only one tree that represents these taxonomic /ancestry relationships. □

Proposition 0.15. *For a given area[3] in Earth, the taxonomic tree derived from it is a subtree of the Tree of Life.*

Proof. Let T be the Tree of Life and $T(A)$ the local tree in the area A. $A \subseteq Earth$. $T(A)$ is a tree because of lemma 1.14. $T(A)$ is based on the same taxonomy given by the species in A (which are leaves in the tree) therefore all the edges of $T(A)$ are in T. The species in A is a subset of all the species in the $Earth$ otherwise the $Earth$ would not be the $Earth$ and there exist another greater set that could be called $Earth$. □

Corollary 0.16. *If $A = Earth$ then $T(A) = $ Tree of Life.*

Proof. Let $A = Earth$. This implies that all species in A are in $Earth$ and vice versa. $V(T(Earth)) = V(Treeof Life)$ and the taxonomic chain (path) of $V(T(Earth))$ is the same as in $V(Treeof Life)$ because it is unique. Therefore $Treeof Life = T(Earth)$ □

---

[3]Any open set contained in the surface Earth. Earth can be considered as a compact surface embedded in $\mathbb{R}^3$

0.2.3 Methods needed

This is just an outline There are several steps needed to answer the main questions.

- ~~Translate occurrences to a tree-structure~~

- Use tree handling libraries (current)

- Generate a global grid with different spatial resolutions minimum 4km or 1km but it's too heavy.

- Develop a method for counting occurrences within a grid cell (boolean, count and proportion) (current).

- Develop a method for measuring connectivity at grid-cell level.
  How about an algorithm for percolation?

- Look for a good-standarized-highly-referenced biomes' map.

- Implement the diversity indexes proposed in Pavoine and Bonsall (2011)

- Load bio-climatic data.

- Geostatistics on bio-clim, point statistics and machine learning coupled with the biodiversity distances.

- Relate human activity.

Desirable analysis:

- Colocation networks at different taxonomic levels.

References

Amari, S. and Nagaoka, H. (2000). *Methods of information geometry.* Translations of mathematical monographs,. American Mathematical Society, Providence, RI. 00059362 Joho kika no hoho. English Shun-ichi Amari, Hiroshi Nagaoka ; [translated from the Japanese by Daishi Harada]. ill. ; cm. Includes bibliographical references (p. 187-202) and index.

Glansdorff, N., Xu, Y., and Labedan, B. (2008). The last universal common ancestor: emergence, constitution and genetic legacy of an elusive forerunner. *Biol Direct*, 3:29. Glansdorff, Nicolas Xu, Ying Labedan, Bernard eng Research Support, Non-U.S. Gov't Review England 2008/07/11 09:00 Biol Direct. 2008 Jul 9;3:29. doi: 10.1186/1745-6150-3-29.

Pavoine, S. and Bonsall, M. B. (2011). Measuring biodiversity to explain community assembly: a unified approach. *Biol Rev Camb Philos Soc*, 86(4):792–812. Pavoine, S Bonsall, M B eng Research Support, Non-U.S. Gov't Review England 2010/12/16 06:00 Biol Rev Camb Philos Soc. 2011 Nov;86(4):792-812. doi: 10.1111/j.1469-185X.2010.00171.x. Epub 2010 Dec 14.

Smith, J. U. and Smith, P. (2007). *Introduction to environmental modelling.* Oxford University Press, Oxford ; New York. 2007296741 GBA685447 Jo Smith and Pete Smith. Environmental modelling ill. ; 25 cm. Includes bibliographical references and index.

Wilson, E. O. (2010). *The diversity of life.* Questions of science. Belknap Press of Harvard University Press, Cambridge, Mass., 1st harvard university press pbk. edition. 2011420458 Edward O. Wilson. ill., maps ; 24 cm. With a new preface, dated 20 May 2010. Questions of science.