



Understanding early human migrations: From archaeological data to statistical and multi-agent modelling

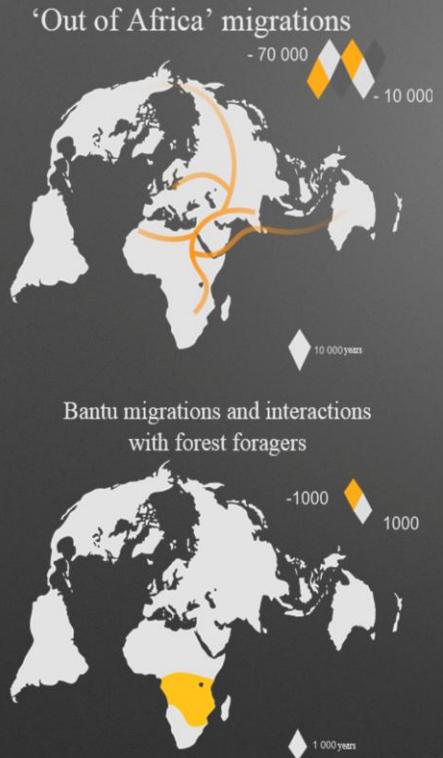
Christophe COUPE
Laboratoire Dynamique du Langage
CNRS / University of Lyon



The WSU Department of Anthropology and College of Arts and Sciences Pullman, November 10-14th, 2014

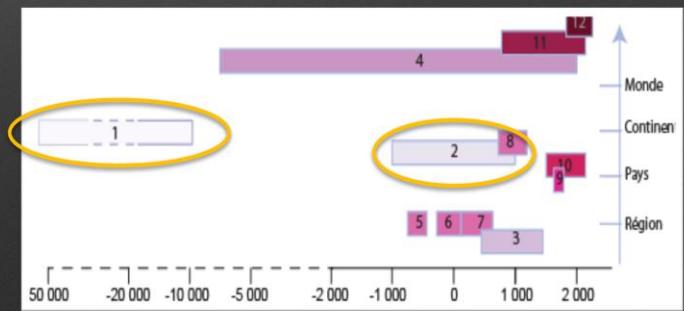
In this talk, we give details about two transitions of the « ANR Transmondyn ».

We highlight how integrating the methods and insights of several disciplines (archaeology, linguistics, computer modeling, statistics) can be helpful to better understand early human migrations.



Transmondyn
AGENCE NATIONALE DE LA RECHERCHE
ANR

Two early transitions related to *human migrations*, worldwide and in sub-Saharan Africa



The first two transitions are related to two major human migrations:

- The first migration out of the African continent by our direct ancestors *Homo sapiens* around 70,000 years ago
- The expansion of Bantu farmers in sub-Saharan Africa from 1,000 BCE, which led them to enter into contact with forest foragers living in the equatorial forest.

We are interested in the spatial and dynamic patterns of these migrations, as well as in their causal mechanisms. Given the great time depth and the related lack of knowledge about the populations involved, we do not focus on social aspects dealt with in other transitions, such as social structures of power or hierarchies.

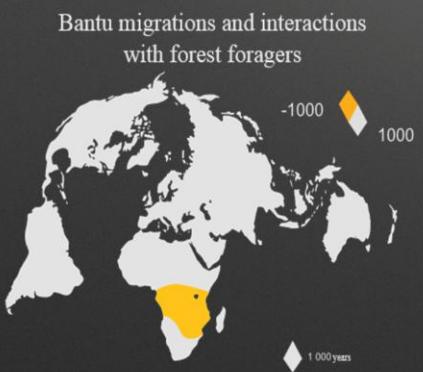


Out of Africa migrations

Before: Homo sapiens in Africa

Transition: Socio-cognitive evolutions

After: People migrating out of Africa



Bantu migrations

Before: Bantu cradle in Cameroon grasslands

Transition: Mastery of agriculture

After: Migrations, interactions with forested areas and forest foragers

We can decompose the two transitions under study into 3 phases: the transition itself, a (more or less) stable state before it, and a state after it.

For 'Out of Africa' migrations, socio-cognitive evolutions can be related to the emergence of our species: new cognitive skills and their behavioral consequences could explain that our species was able to quickly conquer a wide range of environments, and reach locations previously unvisited by human beings.

For Bantu migrations, the development of agriculture and the social changes it carried (demographic growth, need for cultivable lands, property of resources) led farmers to expand from their initial cradle in today's Cameroon.

Collaborations



The Transmondyn team



Jean-Marie Hombert
CNRS, DDL, Lyon

Work is conducted among the Transmondyn team, with closer collaboration with the following members: Jean-Marie Hombert, Florent Le Néchet, Hélène Matthian and Lena Sanders



French National Center for Scientific Research



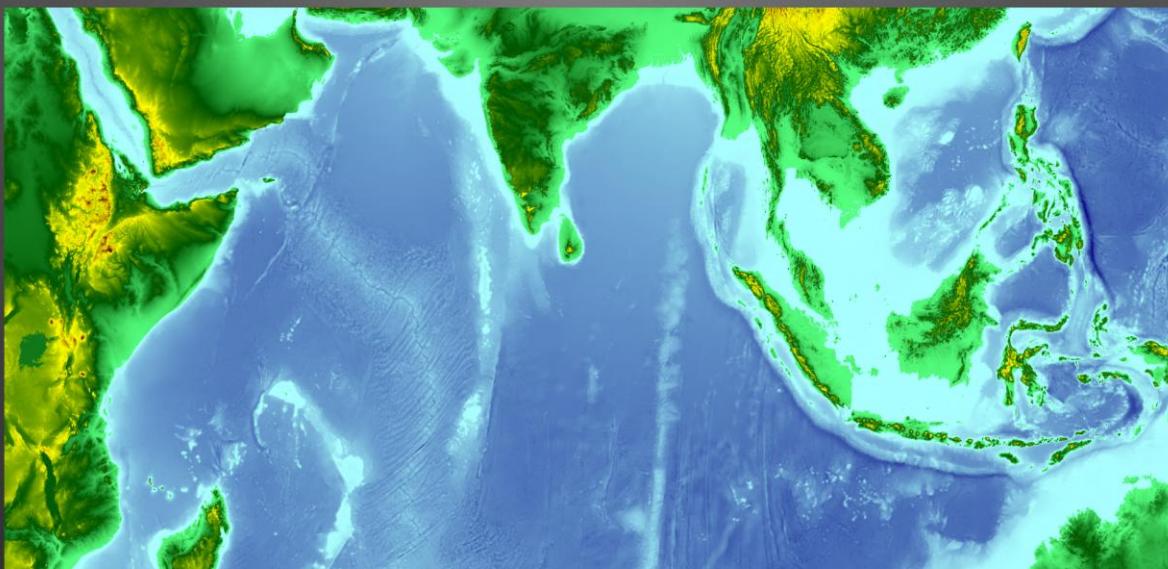
ASLAN Laboratory of Excellence

Issue

Only sparse archaeological traces of
ancient human migrations and activities
(behaviors, interactions etc.)

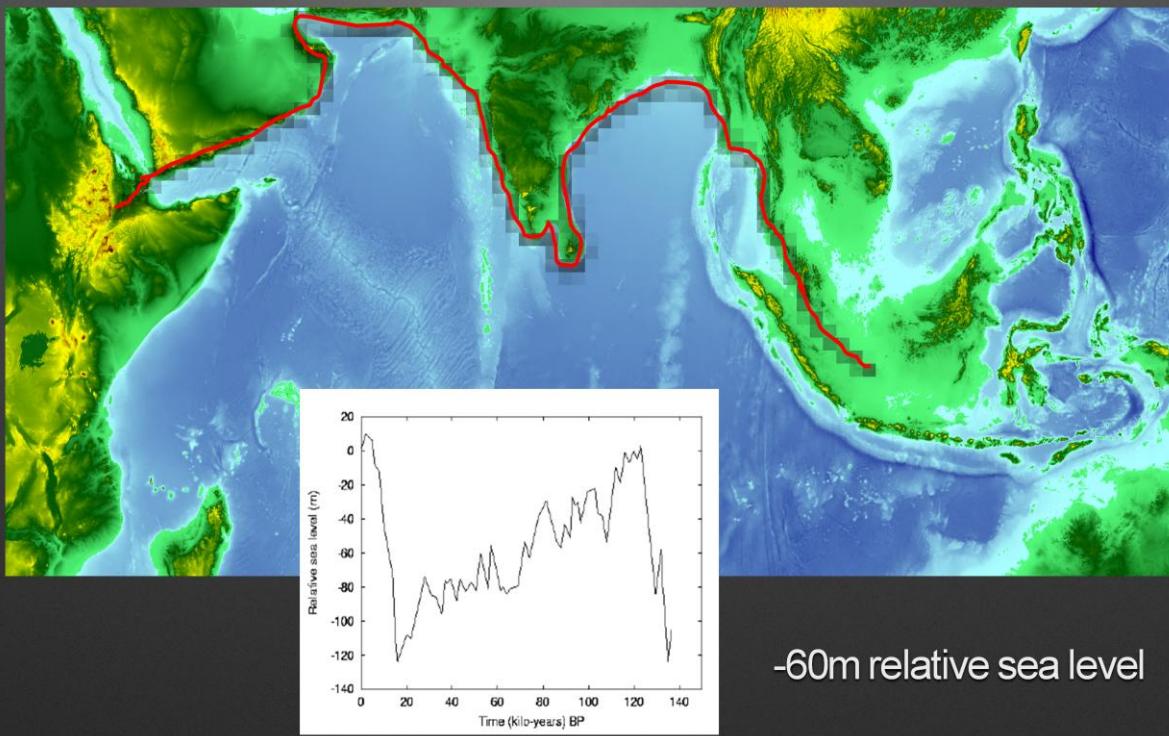
For both transitions, inquiries into what happened precisely and why it happened are constrained by the lack of data: no written document, sparse archaeological evidence. This explains why indirect approaches have their place in the investigation.

A Southern dispersal route?



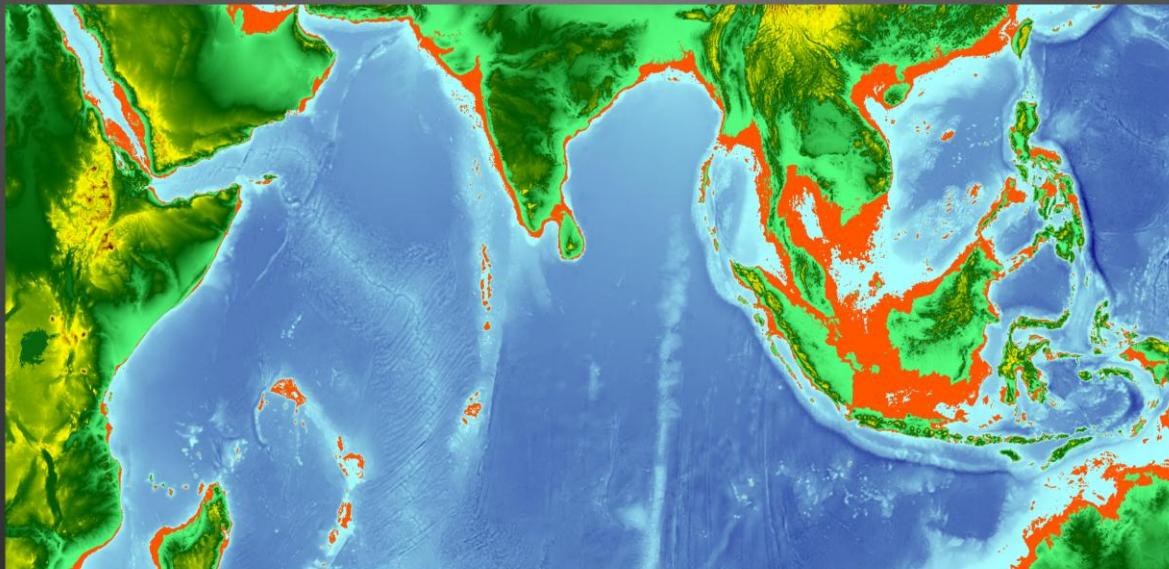
A telling example is the Southern dispersal route likely taken by *Homo sapiens* around 70,000 years ago. This migratory path runs from eastern Africa to Australia and New Guinea along the southern coasts of the Asian continent.

Southern Asia approx. 50 Ky B.P.



The evolution of sea levels during the last glaciation has to be taken into account: during colder periods of time, water from the ocean turns into massive ice packs at the poles, and the sea level decreases. Around 50,000 years ago, it was approximately 60 meters lower than today. As a consequence, migratory paths that were then close to the sea are now underwater. It is therefore very hard to find any material evidence of them. The Southern dispersal route is thus attested indirectly by modern human presence in Australia around 50,000 years ago, as well as by genetic distributions in today's populations.

An underwater southern route?



With modern cartographic tools, it is easy to highlight the differences between the current situation and the situation 50,000 years ago.

Only few archaeological traces of past human migrations



Contributions from other fields



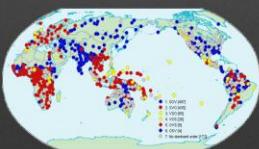
Considerations on early navigation



Understand the context and the facts



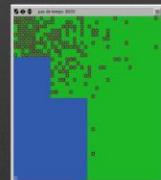
Spatial analysis of linguistic diversity



Analyze the consequences of hidden mechanisms



ABM of migrations



Simulate mechanisms and witness emergence

As previously said, indirect approaches have to be considered to compensate for the lack of material evidence.

In this presentation, we successively review three lines of research. Each domain sheds light on specific epistemological aspects:

- Understand the precise context and facts
- Analyze the consequences of hidden mechanisms, i.e. observe what they produced without necessarily deciphering them
- Simulate mechanisms and witness emerging phenomena to better understand causal mechanisms at play

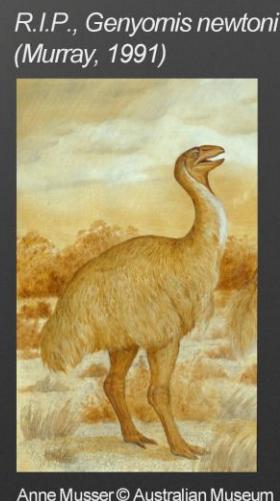
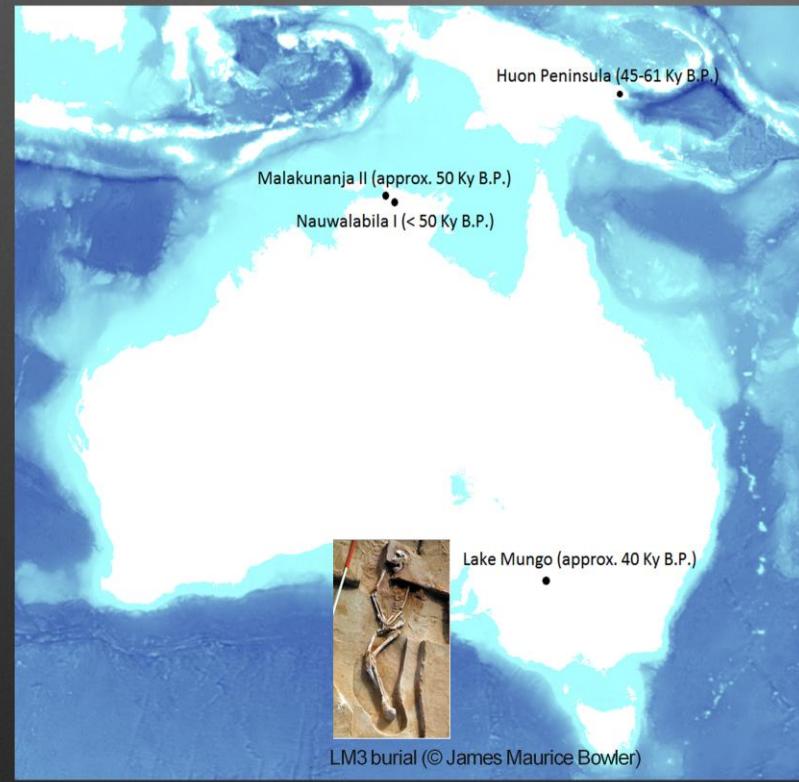
1. Considerations on early navigation



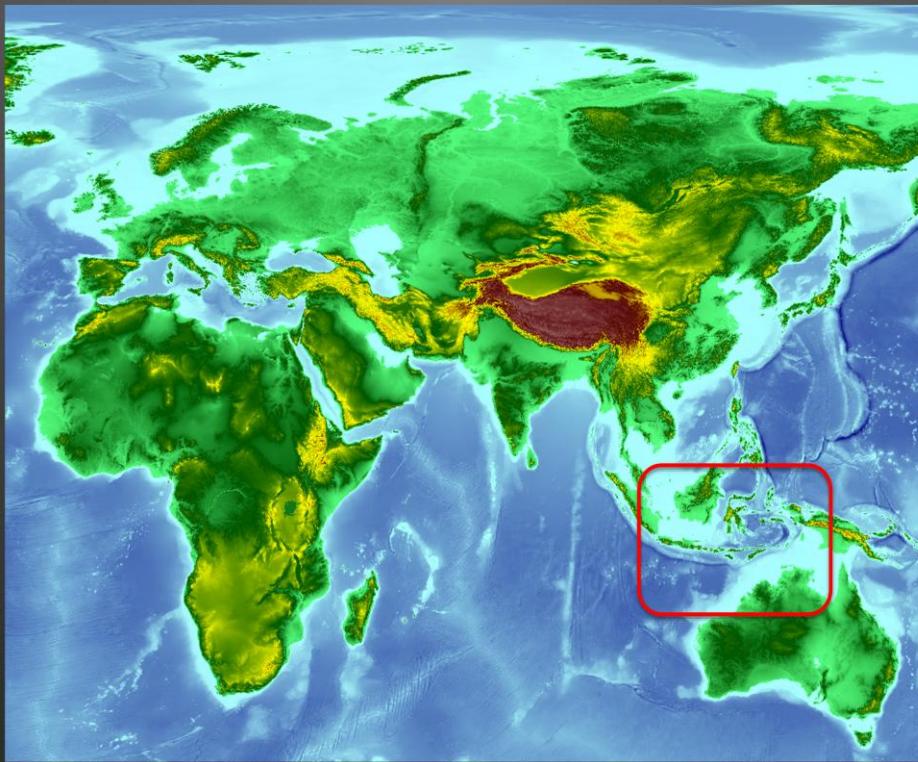
Worora youth on a mangrove tree raft (or 'kaloa'), George Water, Western Australia, 1916 (*Herbert Basedow*)

We first focus on the precise characterization of facts, taking early navigation as an example.

The colonization of Australia

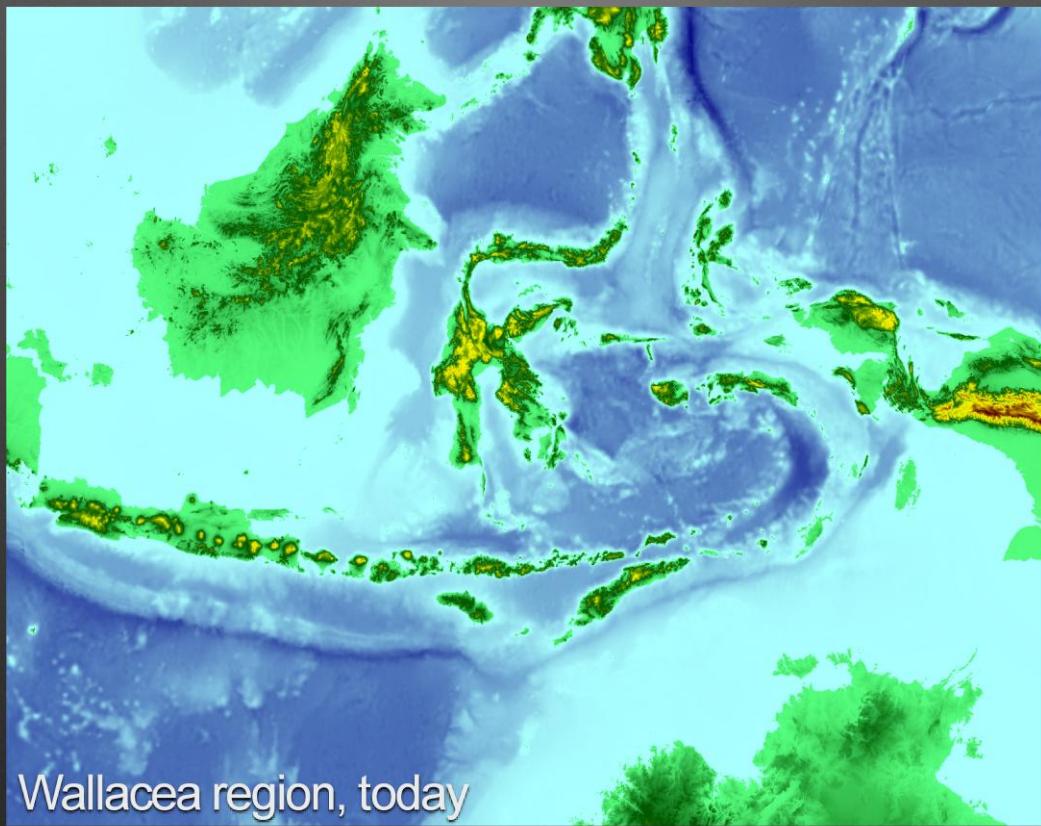


Several archaeological sites demonstrate that Australia was inhabited by modern humans from around 50,000 years ago. No other human beings had reached this land before. Other cues, such as the sudden disappearance of large animals, further prove human colonization.



To reach Australia from Africa, humans had to cross a region between peninsular south-east Asia and Australia/New Guinea, where Indonesia can be found today.

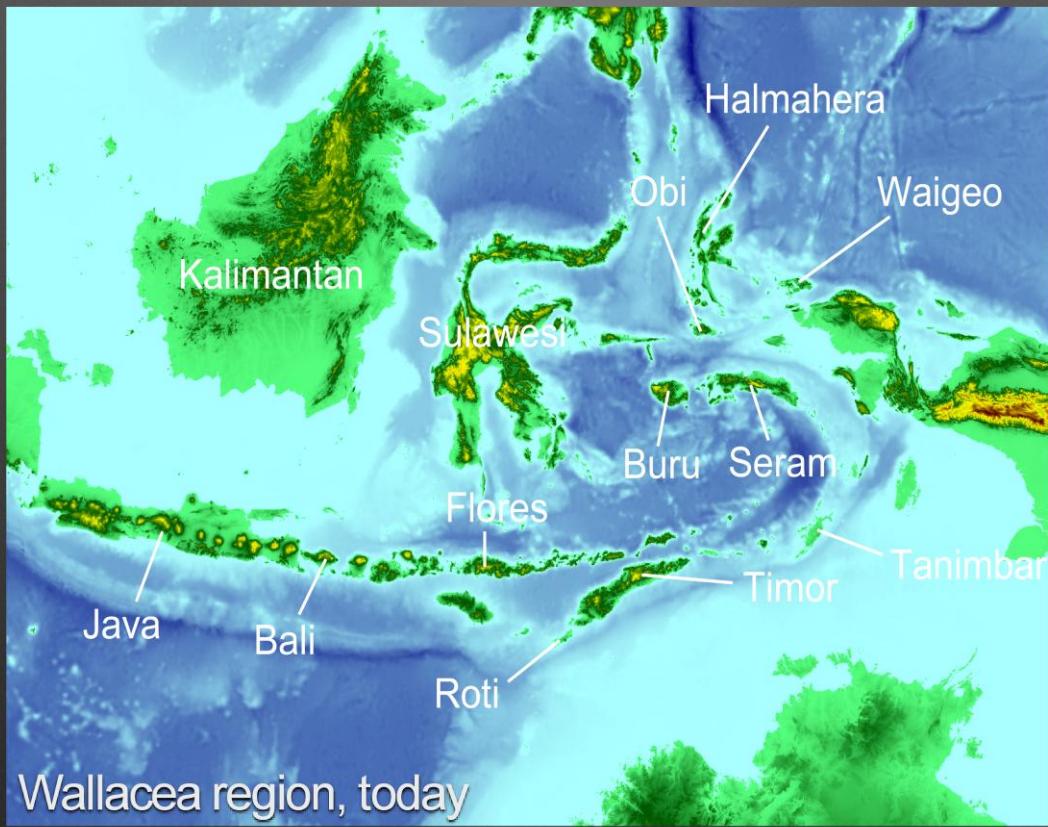
Possible routes across Wallacea



Wallacea region, today

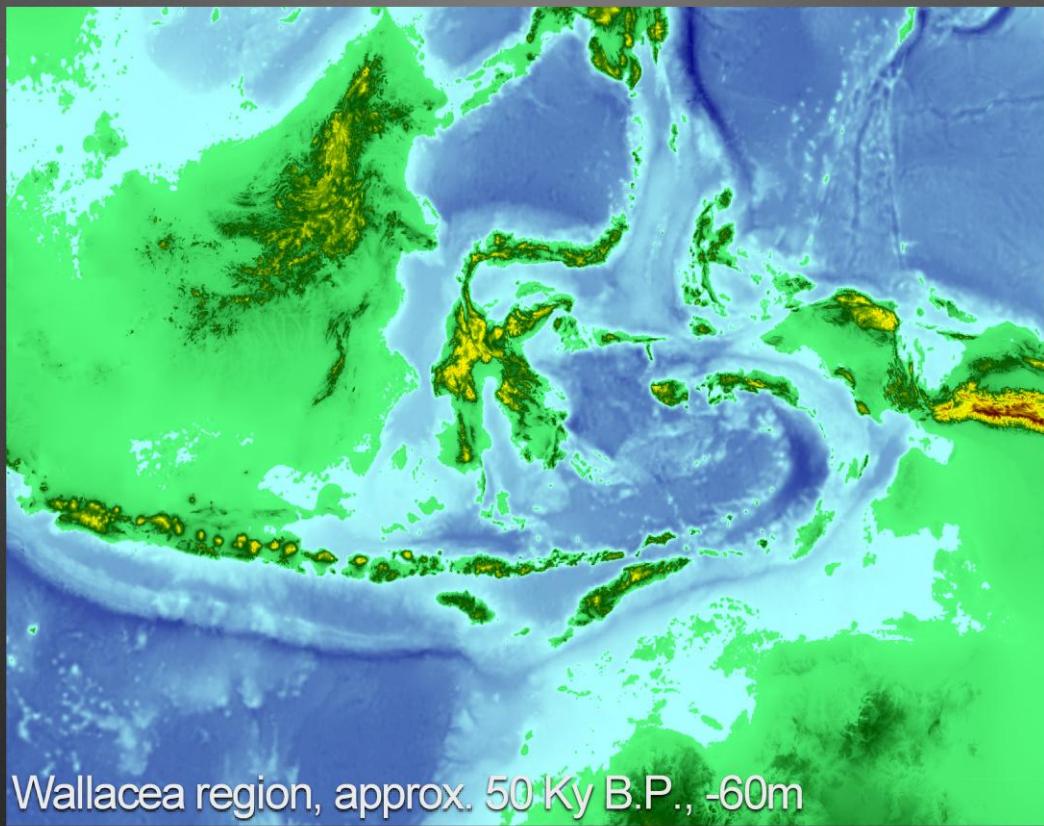
This is how the region, also known as the Wallacea region, looks today.

Possible routes across Wallacea



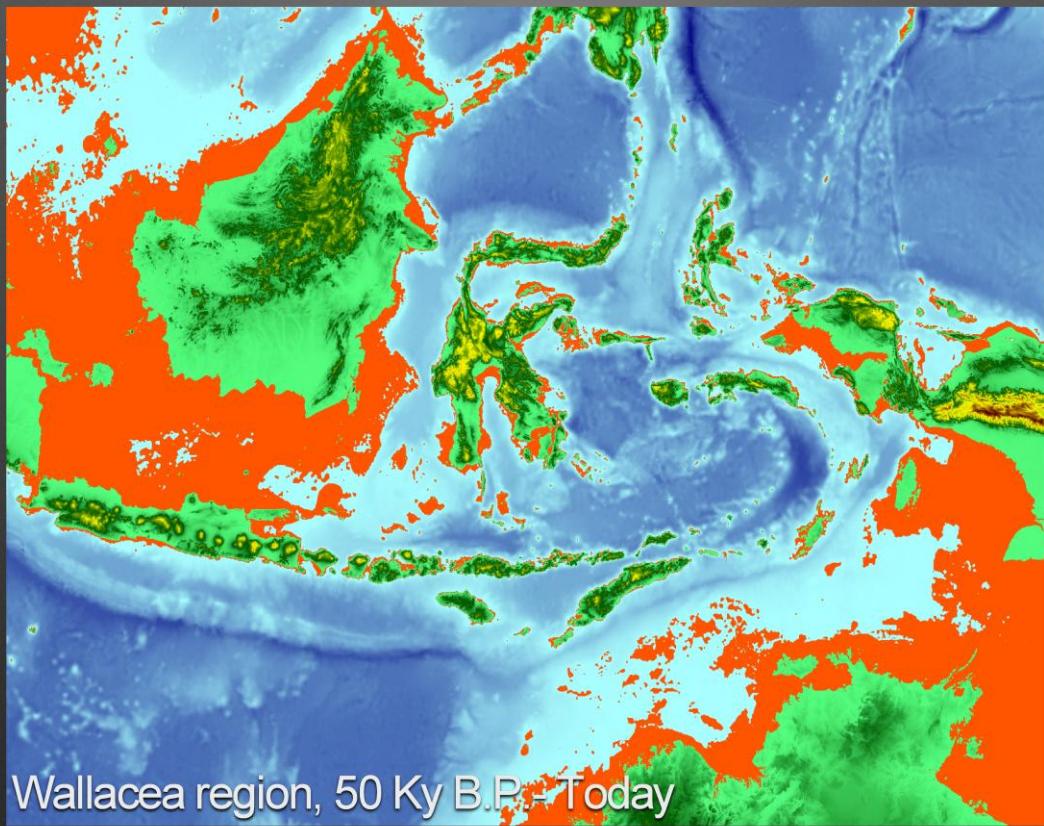
It can easily be noticed that it is impossible today to reach Australia or New Guinea on foot coming from southeast Asia. From Timor to Australia for example, several hundreds kilometers have to be crossed in open water.

Possible routes across Wallacea



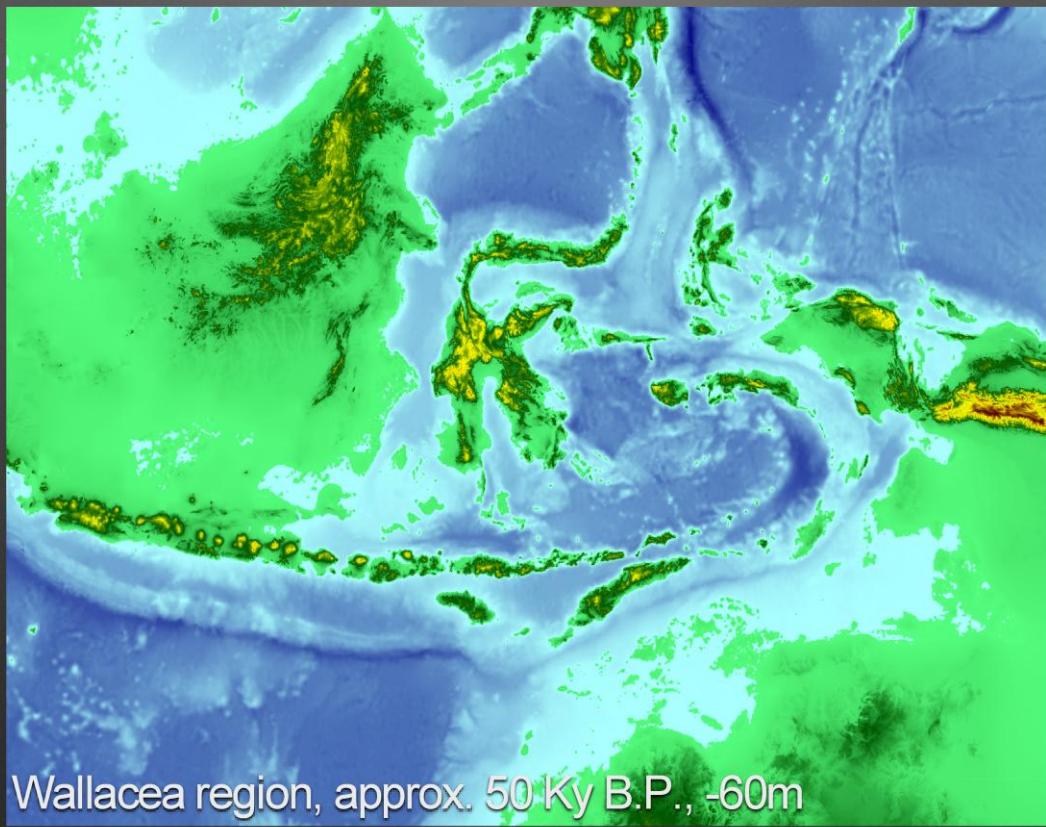
Taking into account the lower sea level 50,000 years ago, one can produce an approximate map of the Wallacea region at that time. Given the depth of the ocean floor in that region, large masses of land emerge. Many islands become connected together, as well as Australia with New Guinea.

Possible routes across Wallacea



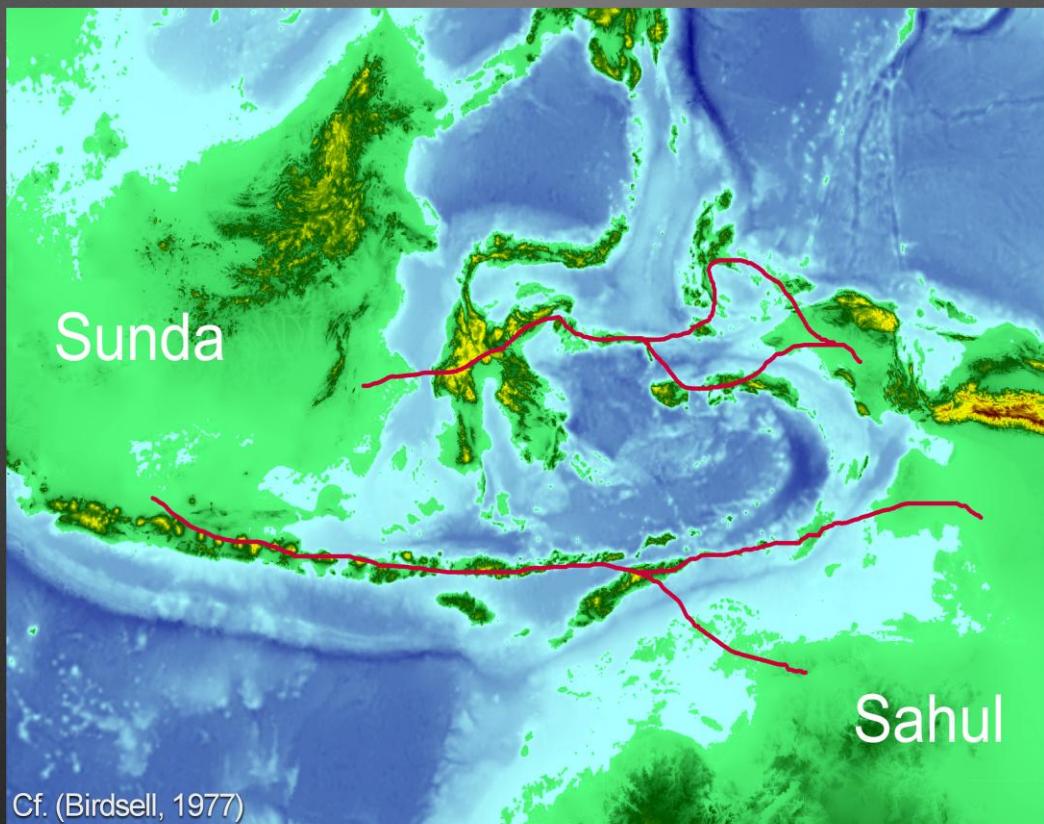
Once again, the difference between today's geography and the past situation can be highlighted.

Possible routes across Wallacea



Just as England could be reached from continental Europe on foot during the colder part of the last glacial period, one could have hypothesized that Australia could have been reached on foot during past colder episodes. However, paleo-geographic reconstructions stress that it was not the case. Our ancestors therefore had to cross the ocean to reach Australia, even if no remains of raft have ever been found.

Possible routes across Wallacea



An ‘archaeology of the invisible’ can however be performed, and various routes between so-called Sunda (a large piece of land on the west) and Sahul (Australia and New Guinea joint together) can be considered and compared in terms of various features: number of kilometers to be crossed on the water, width of target islands etc.)

It can be shown that in the best case, *Homo sapiens* had to cross at least 100 kilometers on the sea, following a northern route reaching New Guinea.

Accidental or intentional sea-crossings?

In the absence of clear evidence, compare the likelihood of competing hypotheses



Extreme situation: a woman pregnant with twins is carried away from Timor's shores to Australia (Calaby, 1976)



Collective project to reach distant lands, with strong motivation and planning
(Coupé & Hombert, 2005)

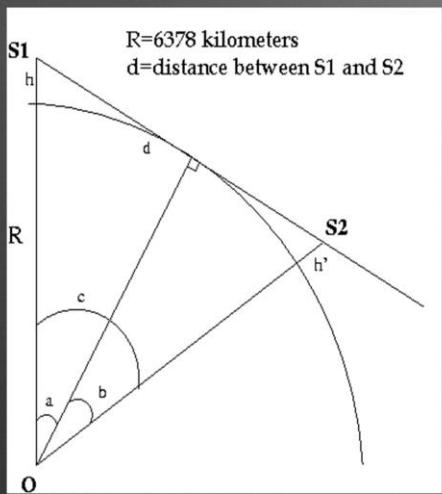
Relevant factors: winds, currents, cataclysms, visibility...

More globally, different hypotheses regarding the nature of the sea-crossings which took place in the region are competing: were these journeys accidental or intentional? A range of scenarios can be put forward, which fall between two extreme cases:

- A ‘purely accidental’ scenario, involving a few people being carried away by winds and currents while navigating close to one of today’s Indonesian islands (or Timor), successfully reaching Australia despite lack of resources and preparation, and further colonizing this new land. This seems highly unlikely, especially with respect to the initial number of individuals needed to successfully colonizing a new territory.
- A ‘fully intentional’ scenario, with human communities carefully planning their journey(s), planning food and water supply, leaving in groups etc. The motivations behind such projects can only be guessed.

To favor one scenario over the others on more solid grounds, factors like winds, currents or visibility have to be considered. Other facts should also be factored in, such as the Toba massive eruption on Sumatra island and its consequences 70,000 years ago, which could have forced people to move away from harsh living conditions.

Computing visibility of distant islands



Only the northern route offers visibility from each island to the next
→ Suggest a specific migratory route

$$d = \sqrt{2 \times \frac{4}{3} \times R \times (\sqrt{h} + \sqrt{h'})}$$

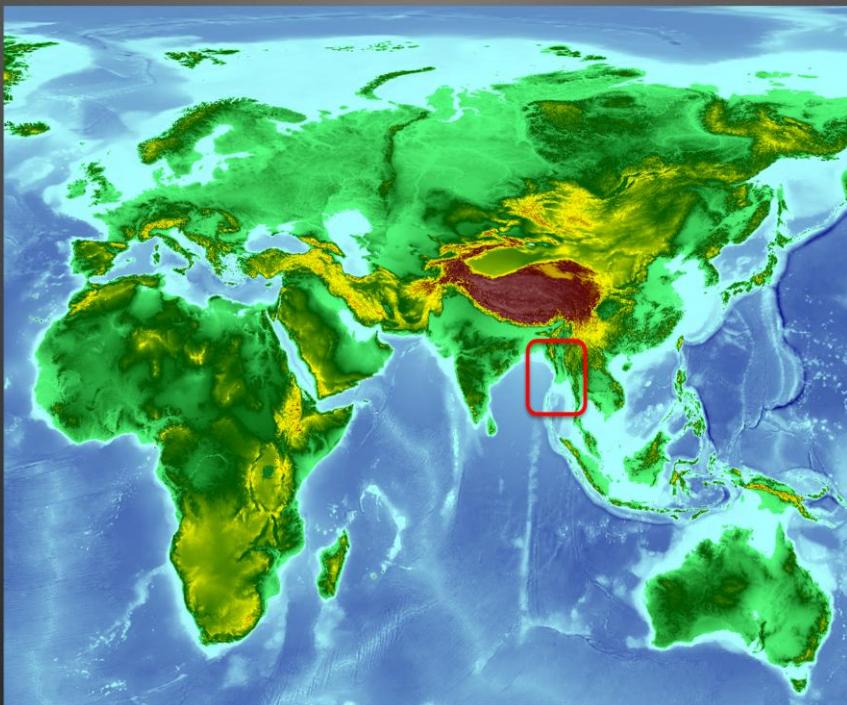
Earth radius

Accounting for refraction in atmospheric layers



We have stressed the potential role of the visibility of distant islands. It can be estimated with rather simple mathematical formulae. Computing visibility along the different possible routes leads to the conclusion that visibility from each island to the next is possible for one of the northern routes leading to today's New Guinea, but not for the southern routes ending in Australia. If one gives more credit to an intentional scenario of colonization, with decisions taken to reach islands that were visible from the starting points, a specific northern route then appears more likely.

Human presence on the Andaman Islands

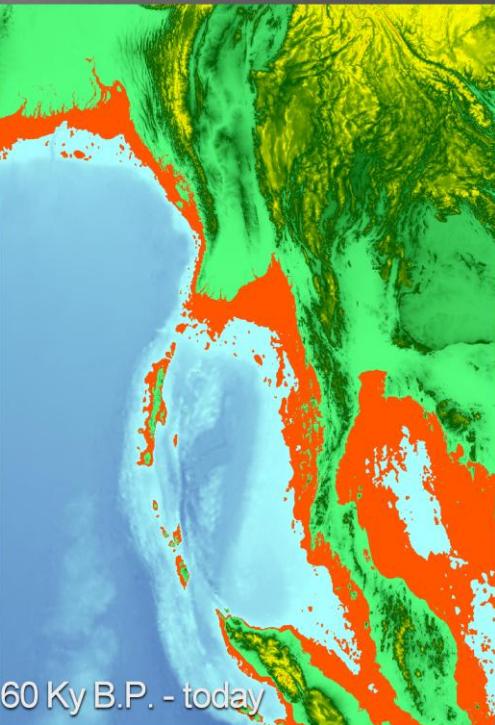


The hypothesis of intentional sea-crossings toward visible destinations can gain support from the study of another colonization event: the reaching of the Andaman islands, located west of the Burmese shores.

Human presence on the Andaman Islands



Sentinelese (©Photo
Source: oxfordjournals.org)



An early colonization of the (then visible) Andaman islands, followed by isolation?

(Coupé & Hombert, 2002)

The Andaman Islands are located along the Southern dispersal route which could have led early *Homo sapiens* from Africa to Australia. Paleo-geographic reconstructions suggest that these islands were visible from the continent 60,000 years ago, while they are not visible today.

Some of the Andaman Islands are inhabited today by human communities which have strongly refused any contact with the outside world, and whose genes suggest a very ancient presence on the island and a relationship with early *Homo sapiens* migrations out of Africa.

It is tempting to interpret the presence and features of these populations as the result of an ancient colonization of the Andaman islands, by people already skilled in navigation – around 60 km had to be crossed on the sea – and who could then see the islands from the continent. Later rise of the sea level would have ‘hidden’ the islands, and isolated their inhabitants for a very long time.

Such a scenario, implying a long-term development of navigation along the coast of southern Asia, supports the idea of further skillful and intentional sea-crossings in the Wallacea region.

Taking advantage of computational resource & power

Rely on viewshed algorithms to better estimate ‘area-wide’ visibility (not only of one location from another one)
(consider Earth curvature)

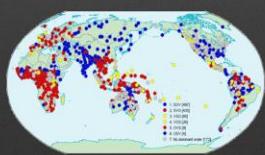
Take advantage of high-resolution datasets & GRASS-like tools (need for elevations above and below sea level)



To further assess hypotheses on ancient sea-crossings, one can today take advantage of available computational tools and data.

High-resolution topographic datasets and viewshed algorithms could especially provide more solid appreciation of visibility between islands.

2. Spatial analysis of linguistic diversity



We now illustrate how a migratory event may be partly revealed by distributional patterns resulting from it taking place in a more or less distant past.

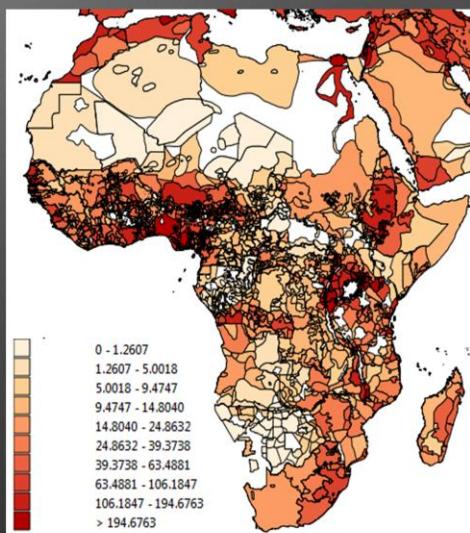
Genetic distributions are among the clearest signals of a past demographic events, but other non-biological features can also tell us about our ancestors' lives, such as present-day languages.

What does linguistic diversity mean?

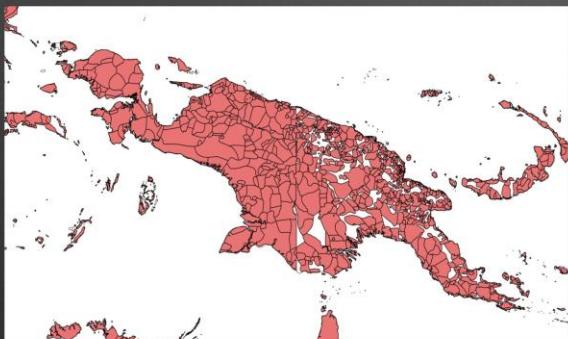
For the sake of clarity, we first introduce the extent of the concept of linguistic diversity.

'Geo-social' linguistic diversity

African languages: area & speaker density



New Guinea: 7.5 million people, more than 1,000 languages



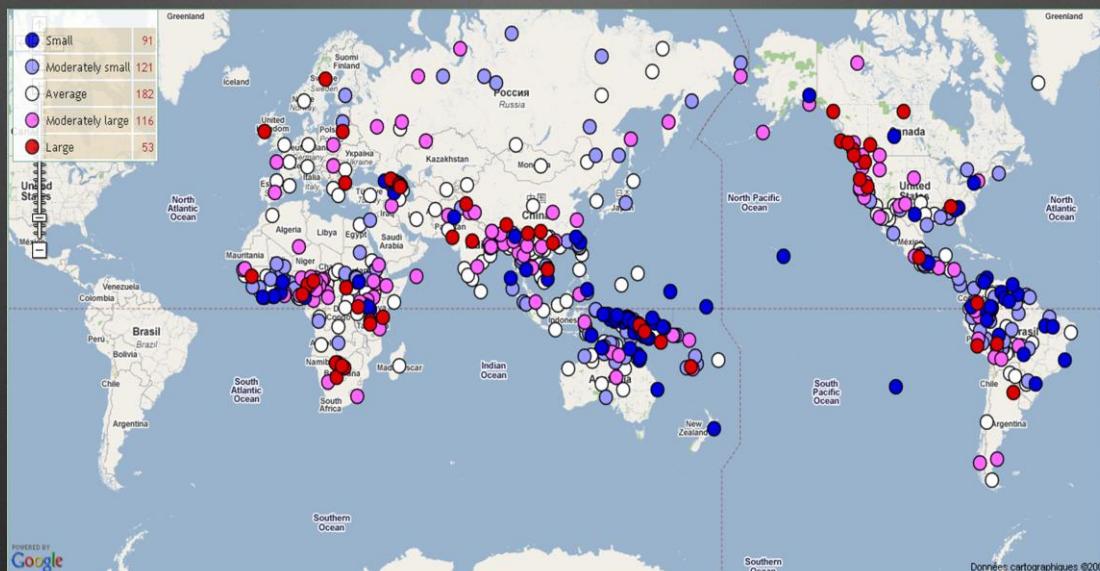
Diversity: number of speakers, area, language families and contact...

Linguistic diversity first relates to geographic and social aspects: how many persons speak French or Swahili? What is the size of the area where Japanese is spoken? How many other languages are spoken in the area of the Thai language? Etc.

'Structural' diversity

e.g. diversity of consonant systems

6–14	small
15–18	moderately small
22 ± 3	average
26–33	moderately large
34 or more	large

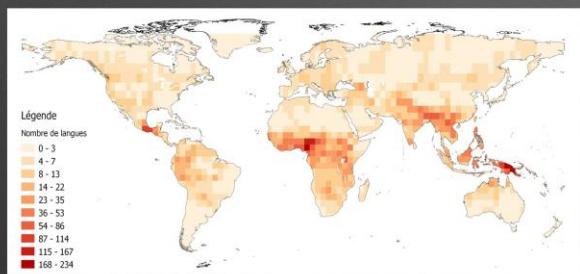


WALS, Ian Maddieson, 2005 (<http://wals.info/>)

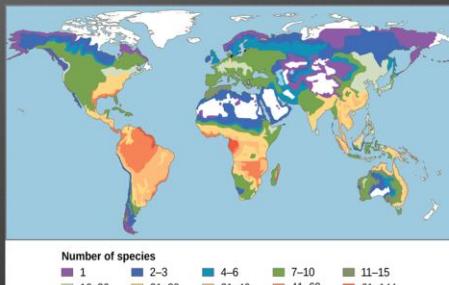
Linguistic diversity also refers to the variety of linguistic structures found in the languages of the world. These structures are found at the level of sounds, syllables, words or sentences.

Questions

Can we understand the causal mechanisms behind patterns of linguistic diversity?



Linguistic diversity

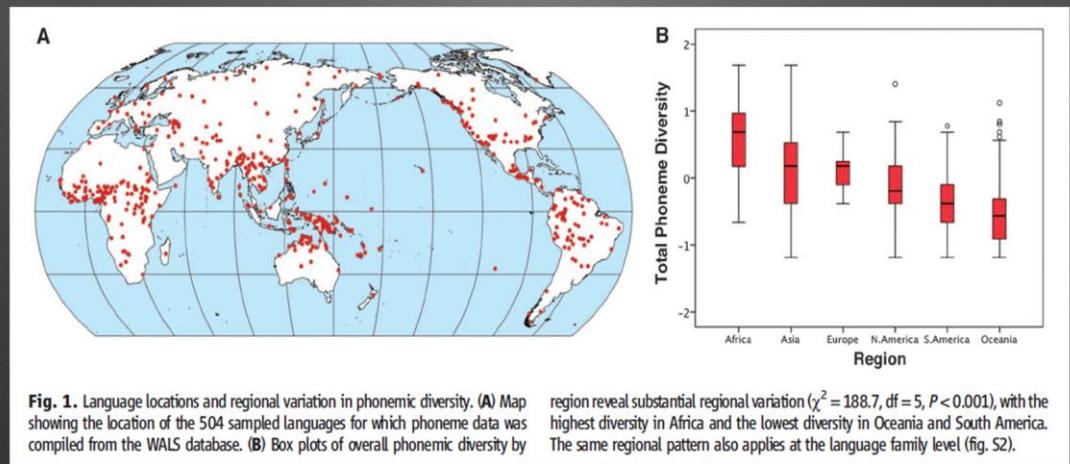


Biodiversity

Can linguistic diversity tell us something about human history and migrations?

Since we can observe regularities in the distribution of languages and of their features, we may be able to understand the different mechanisms that underlie these regularities. It is interesting to notice for example that hotspots of biodiversity match well with linguistic hotspots on the planet, and to wonder about similar mechanisms of diversification. Regarding our transitions, current linguistic diversity may relate to distant events.

Does phonetic diversity relate to ‘Out of Africa’ migrations?



31% of the variance of phonetic diversity predicted by a regression model with speaker population size and distance from central/southern Africa

(Atkinson, 2011)

Atkinson suggested that the distribution of phonetic diversity, which is at its highest in Africa and its lowest in South America and Oceania when measured by the number of phonemes a language has, is a signature of the Out of Africa migrations. Considering 2560 potential origin locations for today's populations in regression models, he found that phoneme inventory size was best predicted by speaker population size and distance from the origin location when this location was in central or southern Africa.

Factoring in environmental factors

Atkinson's study has been criticized on several grounds

A possibly overlooked issue is that environmental factors should be taken into account (to potentially avoid non-causal correlations)

(Atkinson, 2011; Jaeger et al., 2011; Coupé et al., 2012)

His approach has however been criticized, among others on statistical grounds: considering so many regression models indeed leads to a hugely inflated "Type I error", i.e. the best origin location identified by Atkinson may be purely accidental. His notion of phonemic 'founder effect' has also been contested.

We additionally suggest that such approaches should be more cautious of the possible influence of environmental factors. Left outside of such models, what could be interpreted as a causal relationship could indeed be only a side effect of an implicit mechanism, such as the one we describe in the next slide.

Nettle's study of ecological risk

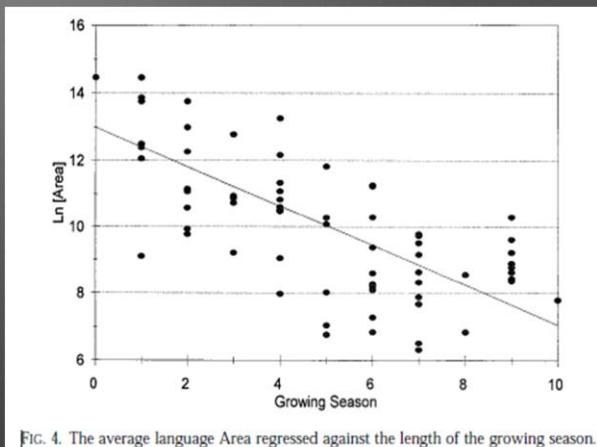
In Western Africa...

The shorter the growing season of plants

The higher the ecological and subsistence risks

The more geographically spread the networks of generalized exchange and mutual dependence

The less linguistic diversity

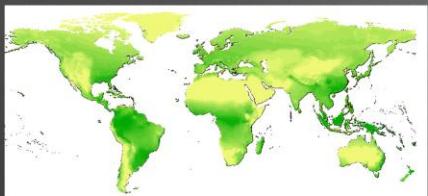


(Nettle, 1996)

Daniel Nettle has stressed how an ecological factor such as ecological risk – the risk for a farmer or a pastoralist to suffer from ecological events such as drought or flooding – could explain geographic patterns of “geo-social” linguistic diversity. This is a clear example of the influence of an environmental factor, which as others should be tested in models aiming at explaining geographic distributions. More explicitly, the pattern of phonemic diversity found by Atkinson could be explained by an environmental factor (such as elevation, humidity, forest coverage etc.) rather by an ancient migratory event.

Methodological aspects

Increasing availability of free high-resolution datasets for environmental, social and linguistic variables



Length of the growing period (FAO & IIASA, 2000)



Tree cover (DeFries et al., 2000)

A serious issue:

- Non-independence of statistical units: *grouping of languages into families and spatial distribution*

→ *Spatial regression models with fixed and random effects*

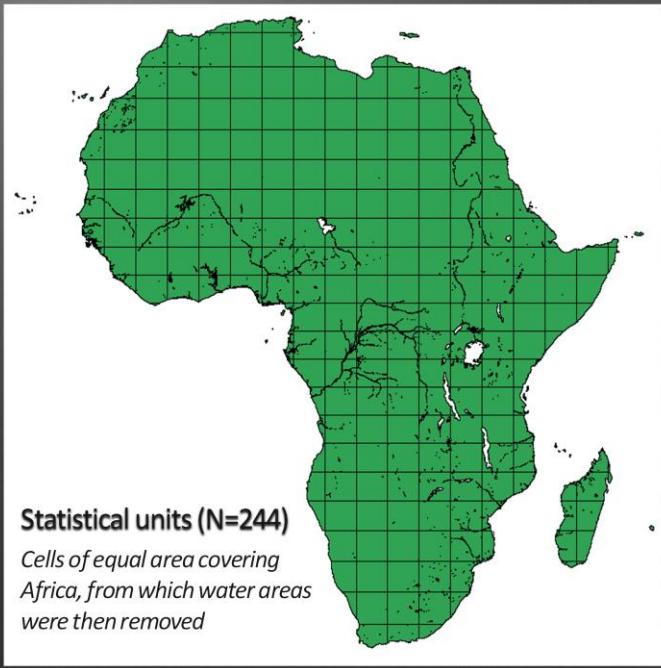
Developing statistical models with a spatial component is difficult, and various methodological challenges arise. At the same time, the advent of the “big data” era encourages us to apply such quantitative tools.

A particularly serious issue is to take into account into the models both non-independence of languages both in terms of linguistic families, and in terms of spatial proximity. The multicollinearity of the independant variables of the models is another problem to tackle.

Predicting the diversity of languages in Africa

In each cell:

- Mean values of environmental variables: ELV, RUG, SEA, WAT, TREE, HERB, LGP
- Population density (POP)
- Mean number of speakers & area per language
- Language density



We investigated to which extent the “geo-social” linguistic diversity of Africa could be statistically predicted by environmental variables. We divided the continent into regularly distributed geographic units (i.e. the statistical units of the model), and for each of them computed the average elevation, land rugosity, distance from the sea, distance from fresh water, tree coverage, herbaceous coverage and duration of growth period for plants. We also computed the average population density. We tried to predict sociolinguistic variables such as the mean number of speakers of a language found to be spoken in the unit, its average area and the total number of languages spoken in the unit.

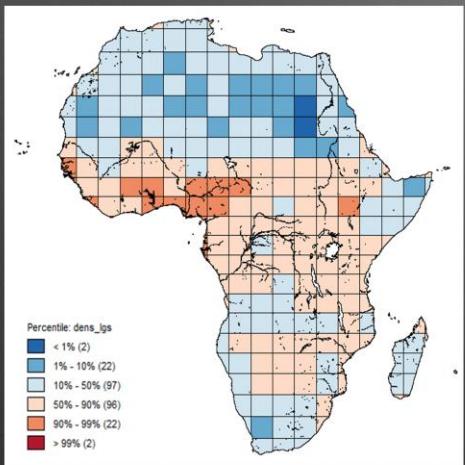
Example of regression model: predicting the density of languages

	OLS	Spatial error model
ELV	-0.27	-0.30
RUG	0.15	0.19
WAT	0.29	0.045
SEA	0.076	-0.19
LGP	0.15	0.046
HERB	0.46	0.29
TREE	0.34	0.27
POP	0.25	0.11
λ		0.80
Adjusted R ²	0.59	
Log likelihood	-229.9	-187.3
Akaike info criterion	477.9	392.7
Schwarz criterion	509.2	424.0

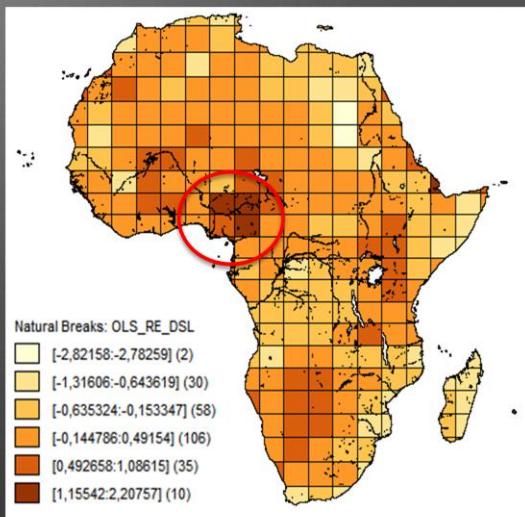
→ The density of languages in a cell is well predicted by environmental and social variables

We found that some of the sociolinguistic parameters could be well predicted by the independent variables, in particular by models taking the spatial relationships of the statistical units into account. The variance of the density of languages in geographic units could be well explained by a linear combination of several of the environmental predictors.

What is not predicted by environmental factors...



Density of languages, percentile map



Residuals of the OLS model

A signature of the cradle of Bantu migrations?

The deeper the history in a location, the more diversity

We also paid attention to the residuals of the non-spatial models. This amounted to identifying the locations in Africa where linguistic diversity was higher or lower than what was predicted by the relationship established at the continental level with environmental factors. The highest positive divergence was found to be located at the putative origin of Bantu languages. Given that, for languages as for genes, regions with the deepest history of human presence show the highest diversity, it therefore seemed that signals of ancient Bantu migrations could be found in today's distribution of languages, given a precise statistical treatment.

3. Computational modelling of migrations



Finally, we highlight how computational models may complement the previous approaches.

A different approach

Previous attempts try to :

- Ascertain the facts and build hypotheses (sea-crossings)
- Study the resulting structures of causal mechanisms of migrations: co-variations of variables (linguistic diversity)

→ *Supplement these approaches with multi-agent modelling*

- Implement plausible mechanisms and observe the emergence of spatial (and temporal) structures

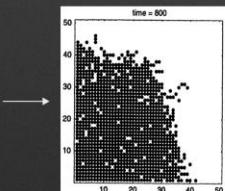
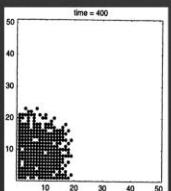
The reproduction ‘in silico’ of past phenomena such as human migrations, with identification of key parameters and study of their impact, is a powerful way to bring out causal mechanisms. Emergence of specific patterns on the basis of plausible mechanisms, with self-organized structures that would have eluded other modeling attempts, is one of the key advantages of computational models.

Human Migration and Environment: the HUME model

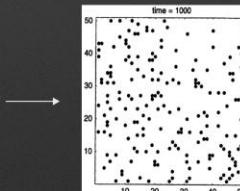
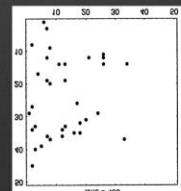
- An interaction between **modelers, geographers** (Lena Sanders, Florent Le Néchet, Hélène Matthian) and **(paleo)linguists** (Jean-Marie Hombert & Christophe Coupé)
- A model derived from Young's model (2002)

A single model giving rise to different forms of colonization according to the values of i) demographic growth rate, ii) mobility rate

High growth rate and low migration rate → **wave front**



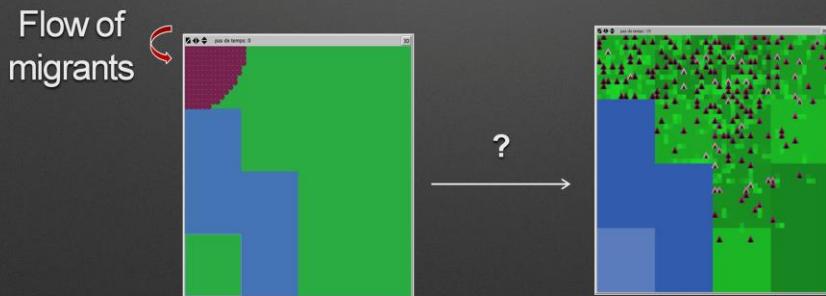
Slow growth rate and rapid migration → **Diffuse colonization** of whole space, low density



We are developing a model named HUME, which results from interactions between “thematicians” and “modelers”. It derives from a previous model designed by Young, and which aimed at identifying the factors behind various forms of spatial colonization.

Aim of the model

- Study the case of ‘Out of Africa’ migrations
- Model the transition from unoccupied to occupied land in a **stylized space**



Modelling driving forces underlying the *spatial* pattern of land colonization

The HUME model addresses the “Out of Africa” migrations (but see the end of this presentation for extension to Bantu migrations).

It simplifies what occurred in reality long ago by adopting a stylized space in which a transition from empty to occupied land takes place through the migration of a number of human groups entering the space from one of its corners.

We should emphasize that the model does not focus on the socio-cognitive evolutions behind the migrations, but rather on the parameters which result in specific migratory patterns across the land, this adopting Young’s perspective.

Modeling choices (1/2)

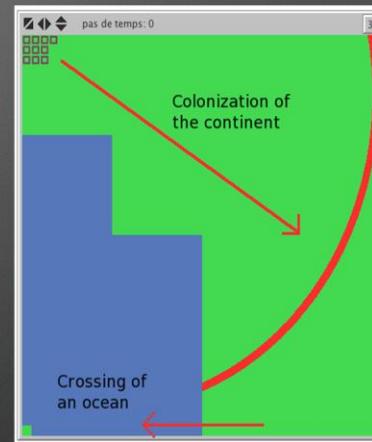
- A **group-centered agent-based modeling** (with respect to cellular automata, with density of occupation)
- **Heterogeneous area and dynamic resources** to be exploited by the agents (cells with evolving carrying capacity)
- Mobile groups who can exploit the cells, **divide** and **disappear**
- **Interactions** between agents and **capacity to innovate** to exploit the environment more efficiently

A number of choices had to be made in order to obtain a fully specified model. One of the advantages of computational models is indeed to force their designers to provide an algorithmic description for every aspects of the phenomenon under study.

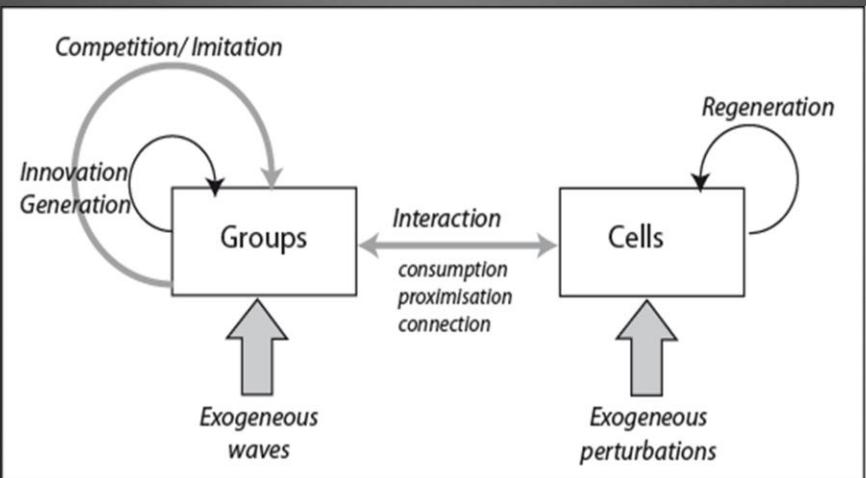
Here, we for example adopted a multi-agent framework rather than a cellular automata, and chose to forget individuals to equate agents to human groups, as our ancestors were living in groups of around 25 to 50 individuals scattered in vast regions of the planet. We also identified that heterogeneous land cells in terms of resources to be exploited by groups would be a significant factor. When other modelers like Hazelwood and Davison or Parisi designed specific non-homogeneous distributions in their models, ours were stochastic to reflecting climatic and environmental changes during the period of colonization.

Modeling choices (2/2)

- 52*52 cells grid
- Resources initially uniformly distributed
- An island, difficult to reach from the continent
- 10 waves of 10 groups arriving in an empty area (tap)



Patterns of interaction



Interaction between groups and between groups and environment

Different stylized mechanisms among groups/agents as well as between them and the environment were designed:

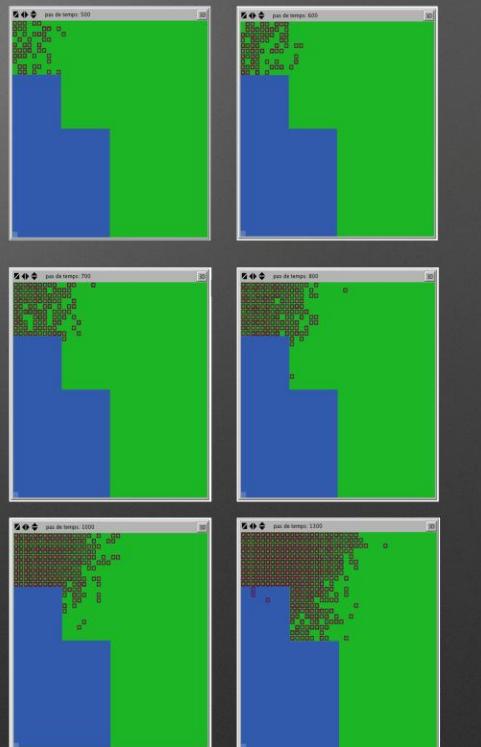
- Groups can consume resources of the cell they occupy, which slowly regenerates and can also vary according to exogenous perturbations. Accumulated energy is consumed during moves across the lands
- Groups can innovate and become more efficient at exploiting resources, but can also learn from groups occupying the same cell as them.

Adding the building blocks

Migration without reason



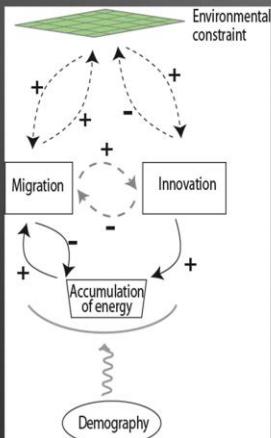
Young A: low mobility + high birth rate



We followed a compositional approach, in the sense that we started from simple models with only the most necessary components to simulate a demographic expansion, before gradually adding other “building blocks” capturing phenomena such as accumulation of energy, innovation or exogenous perturbations. This was helpful to focus on the specific consequences of each component of the model.

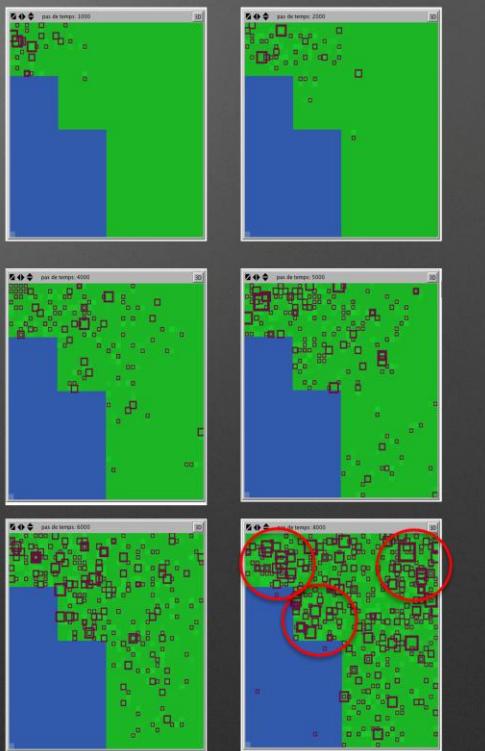
Adding the building blocks

Adaptation & innovation



interactions
group/environment
group/group
group

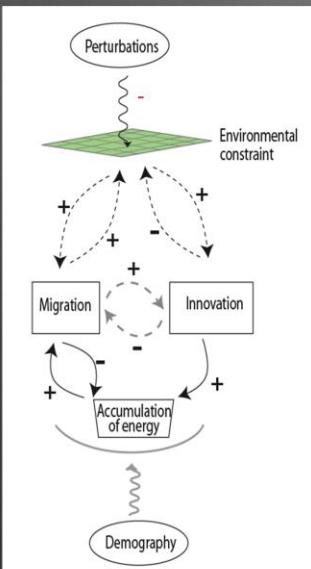
Adding innovation



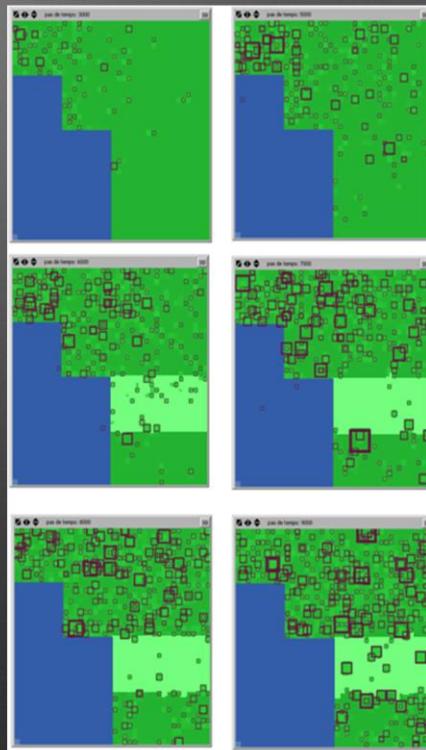
Adding innovations and better exploitation of the environment leads to higher local densities of population, related to the success of a group and its descendants.

Adding the building blocks

Adaptation & innovation



Adding a simple perturbation



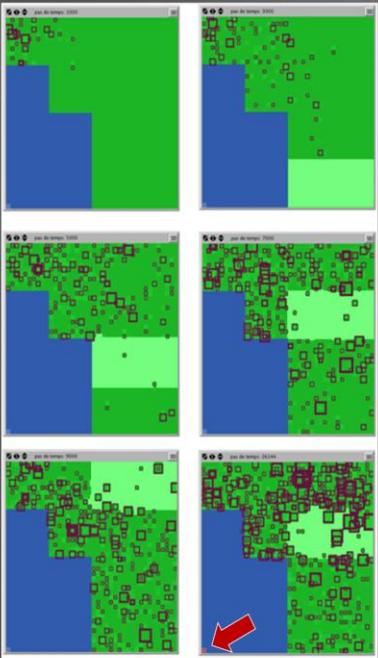
Adding exogenous perturbations allows observing how human groups react to them and global changes in spatial patterns of colonization.

How to measure success?

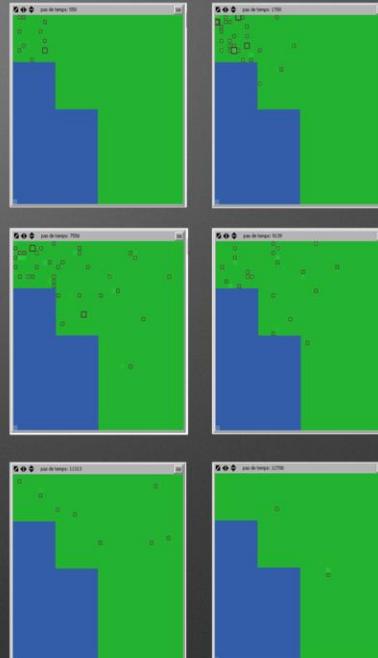
- *Survival of groups* (number of groups at time $t = 20000$)
- *Colonization of the whole space* (average distance of groups from the starting point)
- *Crossing of the ocean* (number of groups on the island at time $t = 20000$)

Since exploring real migratory routes (such as the previously mentioned Southern dispersal route) cannot be expected from a stylized model, we focused on the parameters potentially explaining the success of a colonization event. This relates to the question why our direct ancestors managed to reach lands other human species had failed to inhabit. We focused on three measures related to temporal and spatial dimensions of the colonization. Reaching an island in the stylized space after crossing an ocean where no resources could be collected was a reference to the colonization of Australia.

Diversity of outcomes



Successful colonization

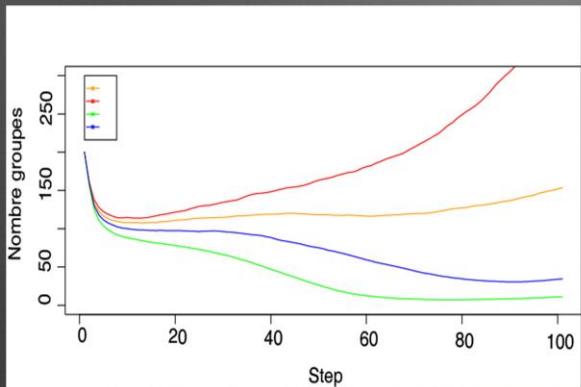


Failed colonization

Also: survival without colonization, colonization without crossing the ocean

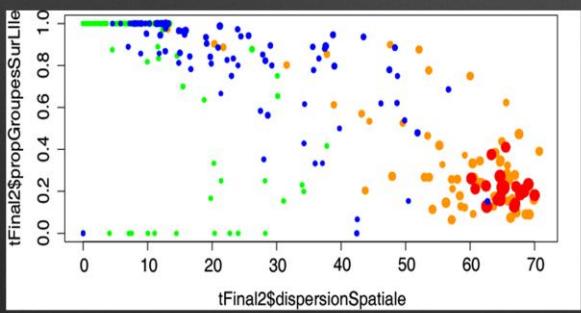
Given the stochasticity of the model, a fixed set of parameters for moves, consumption, innovation etc. led to a variety of outcomes: successful colonization with sea-crossing, successful colonization without sea-crossing to the island, disappearance of the human population etc.

Results



Study of 300 replications

- Variety of outcomes for colonization
- Number of groups over time
- Automatic clustering shows 4 main patterns



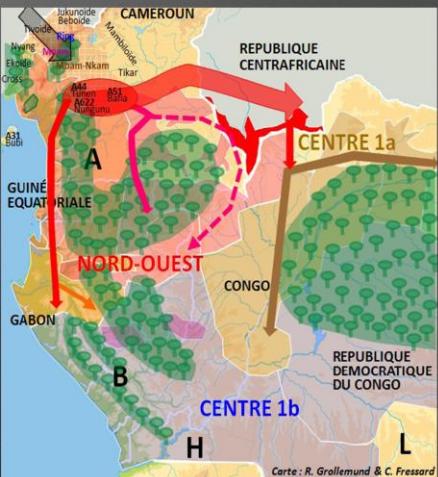
A number of replications were studied and their results could be classified in different ways. When considering the number of groups over time, spatial dispersion and successful sea-crossings, automatic clustering of the simulation outcomes showed that 4 distinct pattern could be differentiated.

Perspectives

Perform a sensitivity analysis to explore the condition of success

Develop the model

1. Track the genealogy of agents
2. Add linguistic markers to compare outputs of the model to today's distribution of languages
3. Adding new agents, and study interactions with them along migratory routes (Bantu migrants meeting forest foragers)



Study the crossing or bypassing of the equatorial forest by Bantu migrants

A sensitivity analysis has yet to be performed to better understand the conditions of colonization success.

At the same time, the HUME model is now further developed to address the case of Bantu migrations. While this migratory event share grounds with the "Out of Africa" migrations, the interactions between Bantu farmers and forest foragers have to be specifically accounted for. Our target in the so-called HUME-BIP model (BIP: Bantu interacting with Pygmies) is to understand whether interactions with Pygmies led to crossing the forest rather than bypassing it. How the obstacle was dealt with by Bantu farmers is indeed an unresolved question today.

The model can also be complemented in different ways, for example either by tracking the genealogy of groups or by adding linguistic markers to them to compare real linguistic distributions to those produced by the simulations.

Conclusions

- Complementary approaches, with different epistemological statuses
- Environment and climate play a central role
- Lack of palaeoclimatic data (especially for models)
- Causal factors and motors: socio-cognitive evolutions and exploitation of resources
- Compare linguistic/genetic distributions as model outputs with real distributions

Different approaches can be simultaneously considered to study the migrations of the first two transitions of the ANR Transmondyn. In all of them, environment and climate are very significant. Especially 70,000 years ago, at a time where social structures and technological developments were quite different from today, adapting to the environment and its changes was of primary importance.

Along these lines, a strong limitation to the previous models is the lack of paleoclimatic datasets. Considering real climatic changes rather than artificial perturbations would move the simulations closer to reality. Despite this limit however, various other directions can be taken, which can help us to better understand past human migrations.