In press in Bernard Laks (ed.) Origin and Evolution of Languages: Approaches, Models, Paradigms. London,Equinox.

Simulating the expansion of farming and the differentiation of European languages

Domenico Parisi, Francesco Antinucci, Francesco Natale, Federico Cecconi
Institute of Cognitive Sciences and Technologies, National Research Council

1. Simulations as theories

Theories in science are traditionally expressed using either everyday language or mathematical equations, with sometimes the help of visual tools such as pictures and flow charts. Many phenomena of human behavior and human societies are too complicated to be captured by mathematical equations while verbally formulated theories tend to be vague and with little empirical content and pictures and flow charts only capture static properties but not the dynamics of mechanisms and processes. Computer simulations are a new way of expressing scientific theories that may help solve the problem of how to formulate theories in the social sciences. Simulations are theories expressed as computer programs. When the program runs in the computer, the results of the simulation are the empirical predictions derived from the theory incorporated in the simulation. For those who are interested in the phenomena of human behavior and societies simulations offer three important advantages. First, simulations force the researcher to make his or her theory explicit, consistent, precise, and complete, because otherwise the theory/simulation will not run in the computer or will not generate the expected results. Secondly, theories expressed as simulations are necessarily rich in empirical content in that, by manipulating the conditions and variables of the simulation and the value of the parameters, much as one does in a laboratory experiment, the researcher can generate a rich and detailed set of simulation results which, as already noted, are the empirical predictions of the theory incorporated in the simulation. Of course, the final test of a simulation is the demonstrated correspondence between the simulation results and the actual empirical or historical facts. But computer simulations can amplify and strenghten the dialogue between theory and empirical facts which is critical for scientific progress. A third advantage of simulations for the study of human behavior and societies is that simulations can reproduce the manner in which phenomena develop, their dynamics in space and time, which is a critical requirement given the historical character of many human phenomena.

Cellular automata [1] [2] are an useful tool for doing simulations of processes of geographical expansion and of cultural/linguistic change. A grid of cells is superimposed on a specific geographical region, with each cell covering an area comprising a given number of square kilometers. Each cell is assigned the same set of properties with a specific value for each property which can vary from cell to cell. These properties may represent the geographical characteristics or the resources present in the area covered by the cell, the number of people living in the cell, their culture or language, their political or technological expansion potential, and so on. Furthermore, the model includes a set of rules for modifying, in a succession of update cycles, some or all of the properties of each cell as a function of the properties of the neighboring cells - for example, the eight cell surrounding a square cell. This produces change in space and time in the region covered by the cellular automaton that may help us understand actual historical changes.

In this chapter we will first briefly present a simulation based on a cellular automaton that tries to reproduce (explain) the expansion of the Neo-Assyrian empire in the Ancient Near East but we will dedicate most of the chapter to describing a simulation model of the expansion of farming in Ancient Europe with possible links to the origin and differentiation of European languages. In both cases our main goal is methodological, i.e., to illustrate the potential of simulations as a research tool in the social and historical sciences, not to claim the validity of any particular model which may seem to be supported by the simulations.

2. The expansion of the Assyrian empire

Consider the geographical region of the ancient Near East and the phenomenon of the expansion of the Assyrian empire which took place in that region from the XIV to the VII century before Christ. The entire region is divided up into exagonal cells of 50 sq. km [3]. At the beginning of the simulation only a single cell in the Upper Tigris is occupied (controlled) by the Assyrians. All the cells occupied by the Assyrians have a property called “expansion potential” with a value that, locally, can go from 1 (maximum expansion potential) to 0 (no expansion potential). The value of the expansion potential for the initial cell is 1. This means that in the next cycle the Assyrians will expand to the 6 neighboring cells with probability of 100%. However, the expansion potential of the 6 newly occupied cells will not be 1 but it will only be 0.98. The rule is that the expansion potential of a newly occupied cell is the same as that of the occupying cell with a decrement of 0.02. This decrement captures the logistic costs of maintaining transportation and communication links between subsequenly occupied cells and the point of origin of the Assyrian espansion, i.e., the initial Assyrian cell. These costs increase with the distance of each Assyrian cell from the center of the empire on the Upper Tigris. The expansion potential of a cell determines the probability that the cell will occupy a new, previously unoccupied, cell. As we have said, an expansion potential of 1 signifies a probability of expansion of 100%. A value of 0.98 a 98% probability, a value of 0.10 a 10% probability, etc. The progressive decrement in the expansion potential of progressively more distant cells signifies that the expansion process will sponteaneously arrive to an end when the logistic costs of maintaining a large empire become too large.

But the model includes other factors. Each cell has two additional properties, each with its own local value: the geographical penetrability of the cell and its political penetrability. Geographical penetrability means that the nature of the terrain (mountains, plains, rivers, desert, etc.) makes the expansion to a particular unoccupied cell more or less probable. Cells with mountains, desert areas, large rivers, have a low geopraphical penetrability index. Cells covering easily accessible terrain have a higher index. Political penetrability refers to the presence of other peoples already occupying a given cell, and the value of this property reflects the demographic density and the level of political and military organization of these peoples. Numerous and well organized people already living in a given cell reduce the political penetrability of the cell. Hence, the probability that a given Assyrian cell will expand to a new, previously unoccupied, cell is a function not only of the expansion potential of the Assyrian cell but also of the two indices of geographical and political penetrability of the cell to be occupied.

Another aspect of the model is that the expansion potential of an Assyrian cell is not only a function of the distance of the cell from the center of the Assyrian empire, reflecting increasing logistic costs, but it also depends on natural and, possibly, other types of resources (e.g., artefacts and other resources constituting the "wealth" of the people originally occupying the cell) which are present in the territory covered by the cell. The expansion potential of an Assyrian cell does not only decrease with the distance from the center of the empire but it can also increase because the Assyrians acquire these resources. Hence, every cell is assigned a further quantitative index which reflects the quantity of resources present in the cell and this index influences the expansion potential of the Assyrians when they happen to occupy the cell.

There are two versions of the model: monocentric and policentric. In the monocentric model only the Assyrians expand. The presence of other peoples in the region (Babilonians, Egyptians, Hittites, Aramaic, etc.) is only passive and is reflected in the political penetrability index and, in some cases, in the resource index of the cells to be occupied by the Assyrians. In the policentric model all the peoples of the ancient Near East may expand like the Assyrians - or fail to expand. Each people is represented by an initial cell in a specific geographical region and the cell is assigned a certain expansion potential, which may vary from 1 to 0 depending on the people. All the cells still have a political penetrability index but this index is not assigned once and for all at the beginning of the simulation, as in the monocentric model, but it changes dynamically during the historical process the simulation is trying to reproduce. If a cell is “empty”, i.e., it is not occupied by anyone, the index has its maximum value: 1. If the cell happens to be occupied by some people, its current political penetrability index reflects the local expansion potential of the people occupying the cell. When two peoples occupy neighboring cells and therefore each people tries to expand to the cell occupied by the other people, the people that “wins” is the one with a greater local expansion potential.

To run a simulation using the monocentric model the researcher initially defines three maps: the map of the geographical penetrability of the cells, the map of their political penetrability, the map of the resources present in each cell. To run the policentric model, the map of political penetrability is omitted but one has to assign to each people of the region an expansion potential index for their initial cell. Furthermore, the cellular automata model has to be temporally calibrated. The simulation runs in cycles. In each cycle all the cells are updated, i.e., their properties that need to be modified are modified. The temporal calibration of the model means that one has to decide how many years are covered by each update cycle.

Then the simulations begins. The initial year for the expansion of the Assyrian empire is 1350 B.C. Its end is the end of the VII century.

A basic result of the simulation is how the size and shape of the Assyrian empire changes at various successive dates during the period considered. The simulation generates a succession of maps of the (simulated) Assyrian empire at various dates which can be compared with the succession of historical maps. A more global but quantitative result is how the total size of the empire changes in each of the successive dates. The resulting curve can be compared with the corresponding historical curve. In fact, using the monocentric model the simulation generates a curve of historical changes in the size of the Assyrian empire which closely matches the actual historical data.

Two aspects of the actual historical phenomena are not captured by the simulation. First, it is known that, while the total size of the Assyrian empire tends generally to increase in the period considered, a couple of times during the period there is contraction rather than expansion of the territory controlled by the Assyrians. And, second, at the end of the period considered there is a rather sudden collapse of the entire empire which disappears as a unified political entity in only a few years towards the end of the VII century. Both phenomena are not, and could not, be replicated in the simulation we have described. The first phenomenon is likely to be due to contingent factors such as succession crises or ascent to power of inadequate emperors. The terminal collapse tends to be interpreted as either due to external factors such as the emergence of new powerful neighbors or to internal reasons such as an excessive growth of logistic and administrative costs for maintaining such a large empire. Contingent factors that are known to have existed historically can be incorporated in the simulation which, contrary to what one might think, is not restricted to systematic and general mechanisms and rules but can also take into consideration in its calculations contingent factors and causes. On the other hand, the collapse of the empire at the end of the VII century for internal causes might be predicted by the simulation by incorporating in the simulation a model of the organization of the Assyrian empire, and in particular of the relationship between its size and costs. But, independently of how one might obtain a better match between the simulation results and the actual historical phenomena in this particular case, the general lesson to be drawn from the simulation is that mismatches are always possible and that they may be useful for modifying the model and for better understanding the actual explanatory power and limitations of the ideas in terms of which we try to account for historical phenomena. Furthermore, a mismatch between the simulation results and the historical data in our possession can be useful if the simulation results can induce us to look for new, additional, data.

3. The expansion of farming in Europe

We now turn to the simulation of the expansion of farming in Europe starting from West Anatolia nine thousand years ago and the possibly associated process of differentiation of European languages.

The entire territory of Europe is divided up into a grid of square cells of around 70 square kilometers [4]. (Adjustments are made for the spherical shape of the Earth.) Each cell is assigned four properties that determine the potential farming productivity of the cell: (1)

extension and height of mountains; (2) river runoff (annual average); (3) rainfall (annual average); (4) nature of terrain with specific reference to its suitability for farming. These four properties are put together to determine a single index of the Carrying Capacity of the cell, i.e., of the number of farming people that can live in the cell. Each cell is also assigned a demographic index measuring the actual number of farming people currently living in the cell. At the beginning of the farming expansion process in the IX millennium before present only one cell in South West Anatolia has a value of 400 (people) for this index. All the other cells have a zero index. This means that farmers only live in the Anatolian cell and there are initially no farmers elsewhere in Europe.

The simulation runs as follows. Each cell containing farmers “looks at” its eight neighboring cells to find out whether these cells are available for occupation. A cell is available for occupation unless it is already occupied by other farmers or the cell is entirely unfit for farming (e.g., high mountains or desert). For each cell the number of neighboring cells that are available for occupation is determined and if this number is 4 or more, the population of the cell undergoes a demographic growth of 3.5%. If the number is 2 or 3, the growth is 2.5%; if it is 1 or 0, there is no growth. Updating the population of farmers living in all cells completes a cycle, and then a new cycle starts. A cycle represents one year since it reflects the demographic growth rate.

When the number of farmers living in a cell reaches 45% of the cell’s Carrying Capacity, a migration occurs: a portion of the cell’s inhabitants (25%) moves to one of the empty neighboring cells selecting the cell with the highest Carrying Capacity. Cells that represent sea cannot be occupied but they can be penetrated by the migration movement. Three conditions govern sea crossing: (1) a randomly selected neighboring sea cell is penetrated if none of the neighboring land cells is available for occupation; (2) once a population has penetrated a sea cell it automatically continues to move in each cycle from sea cell to sea cell in a straight line until it reaches a land cell; (3) if a land cell is not reached within a distance of 20 sea cells (around 150 km) from the beginning of the crossing, the population dies out. Condition (3) reflects the level of navigational technology and skills reached at the time.

Figures 1-8 show the expansion process at different times expressed in years since the beginning of the process in the IX millennium. Grey shades represent density of occupation by farmers (white = no occupation; black = maximum density). The Middle East area should be disregarded since it is probable that a different expansion process also tied to the Neolithic farming revolution had taken place in that area prior to the European expansion process originating in Anatolia which is the object of the present model.

INSERT FIGURE from 1 to 8 ABOUT HERE

Figure captions indicate the time taken by farmers to reach selected locations in Europe and the distance of these locations from point of origin (this distance may include sea crossings not exceeding 20 cells). An interesting result emerging from these data is that the expansion process turns out to be based on an average rate of advancement of around 1 km per year (somewhat more in the advancement along the Adriatic coasts and somewhat less towards the Baltic coast and the Easters territories). This is in agreement with the model of Ammerman and Cavalli-Sforza [5] and Renfrew [6] (see also [7]) and with arachaeological evidence of Neolithic settlements on the terrain. However, it is important that while in Ammerman and Cavalli-Sforza’s and Renfrew’s model the rate of advancement of 1 km/year is assumed as such and fitted into the model, in our simulation it results spontaneously from the demographic growth rate, the nature of the terrain, and the expansion rules postulated by the simulation. Ammerman and Cavalli-Sforza derive the expansion velocity from the archaeological evidence and in order to include this value in their model they assume that at each 25-year generation interval the population moves 18 km on. In contrast, the simulation simply assumes the demographic growth rates, which at a micro-scale level may be different in different cells, feeds them into a representation of the real terrain encountered at the micro-scale level, and generates as a result (in a by no means obvious way since it is a result of literally millions of local interactions) a macro-scale expansion velocity averaging 1 km/year.

This result seems to give independent support to Ammerman and Cavalli-Sforza's and Renfrew's model. By including in the simulation only the demographic growth rates and by making the mechanisms and effects of population growth and expansion process solely dependent on the properties of the occupied and immediately surrounding terrain at the very fine scale of 70 sq. km units, the simulation succeeds rather well in reproducing the real behavior of Neolithic farmers in the specific environment constituted by the terrain mapped in the simulation. The fact that this behavior reconstructed at such a microscale produces through millions of local interactions macro-patterns that are in agreement with the observable evidence militates in favor of these local interactions having actually taken place.

A methodologically interesting aspect of the model concerns the level of navigational technology and skill as reflected in the maximum distance that could be navigated without touching land. In the model that we have described a distance of 20 sea cells, that is, around 150 km, is assumed. If we vary this value and assume a somewhat higher level of navigational technology/skill, say, 30 sea cells, i.e., around 225 km, the same overall results are obtained but with some quantitative and qualitative changes. For example, with only 20 sea cells of navigational autonomy the Black Sea is circumnavigated but not traversed from South to North. With a higher level of navigational autonomy of 30 sea cells the Black Sea is both circumnavigated and traversed. Another difference is that with a higher level of navigational technology/skill the West Mediterranean regions are reached earlier and, in Italy, the expansion along the Thyrrenian coast plays a more important role even if the expansion along the Adriatic coast remains prominent. It may not be clear if these differences militate in favour of the higher level of navigational technology/skill in the period considered, or if the lower level assumed previously gives more historically accurate results. What is important is the light these different results shed on simulations as a research method. First, one can clearly see that simulations tend to generate a very rich and differentiated set of results that can be compared with different types of empirical and historical evidence, in our case archaeological evidence on expansion of farming but also on level of navigational technology/skill (which, incidentally, may have varied during the period considered, with the possibility to incorporate these variations in the simulation). Second, simulations clearly show how different factors, in our case level of navigational technology/skill and geographical factors, may interact with each other in complex ways that may be difficult to capture by simple reasoning. An assumed level of navigational technology/skill may give good results for how Italy was colonized by farmers but bad results for how the regions around the Black Sea were colonized. This represents a challenge to formulate explicit and self-consistent models for interpreting the past.

4. Origins and differentiation of European languages

Because the simulation reproduces the expansion process cell-by-cell and year-by year, for each cell it is possible to know when it was first occupied (number of years since the beginning of the expansion process) and to trace the entire path followed by farmers from the point of origin of the process to the particular cell. We can use this information to construct trees of expansion paths with a common “root” in Anatolia and branches in the various regions of Europe. If we assume that the expanding people had a culture or a language that they carried with themselves and that became progressively differentiated because of internal changes and the lack of interactions with people living in distant cells, we can interpret these branching trees as reflecting different degrees of cultural/linguistic similarity among people living in different regions in Europe. Cultural/linguistic differentiation sets in as a function of the temporal length of separation of different groups from a single community originally having a shared culture and a common language. By choosing any two cells in Europe, tracking their paths of provenance, and computing the time elapsed since the separation of these paths, we can determine the degree of cultural/linguistic differentiation between the two cells.

We have done this for a number of “critical” cells (geographical locations in Europe). Figure 9 shows the various paths of provenance of these cells, which obviously collectively take the form of a branching tree rooted in the Anatolian point of origin of the expansion process.

INSERT FIGURE 9 ABOUT HERE

What is the meaning of this tree? We can interpret the tree in linguistic terms and use it to assess the validity of one of the two main hypotheses that have been advanced to explain the similarities among most of the languages currently (and in the past) spoken in Europe. According to this hypothesis [6], almost all the European languages spoken in Europe derive from a common proto-language which was spoken by farmers living in Southwest Anatolia nine thousand years ago. The demic expansion of these farmers produced the different languages spoken in Europe, with the similarity tree of these languages derivable from the tree of expansion paths of Figure 9. (Another direction of expansion of these farmers was towards West Asia and this direction of expansion produced the Indo-Iranian branch of the Indo-European family, but this is outside the scope of our simulation.)

Let us try to read in these terms the branching pattern of Figure 9 in relation to its time depth. This can be done by comparing the branching pattern with the different states of advancement shown in Figures 1-8.

There are four initial branches that start directly from the root. They go off approximately towards the four directions North-West, North-East, South-East, and South-West. After 700 years (see Figure 1) - a period of time long enough to determine significant linguistic differences - the South-West branch has touched Greece having gone through Rhodes, the South-East branch is pointing towards the Caucasus after having gone through Cyprus, the North-East branch traverses the central part of Anatolia, the North-West branch has reached the Bosphorus strait, crossed it, and touched what today is Bulgaria.

For these first four main population branchings we can posit four major linguistic groups. However, in order to gain some insight into their correspondence with the Indo-European language family we have to consider their further development.

The first branch, the South-West one, divides into two: one goes north to settle Greece and the internal part of southern Balkans, the second tours around the Greek peninsula to land in Italy (Otranto). Here it divides up into two again, one group going up along the Adriatic Sea and the other one along the Tyrrhenian Sea. This last will eventually settle France, Spain, and England. It is not difficult to identify a Greco-Italo-Celtic language group, with Greek becoming separated from Italo-Celtic at an early stage (well before the end of the first millennium; cf. Figure 2), and Celtic branching off towards the end of the second millennium (date of arrival to Provence is at 2200; cf. Figure 4).

The second branch, the South-East branch, crosses the whole of Anatolia in the direction of Armenia, which is reached in around 1750 years (cf. Figure 7). Then it goes on towards the southern shore of the Caspian Sea, into Iran (reached after another 1000 years; cf. Figure 8), further proceeding towards Afghanistan. Although the simulation geographically stops there, one can easily foresee a further advance towards Pakistan and Northern India. In Indo-European terms, this branch corresponds to the Armeno-Aryan group, with Armenian branching off and leaving an Indo-Iranian group, and a subsequent branching of the Iranian languages spoken in Iran and Afghanistan from the languages of Northern India.

The third, North-East, branch occupies the whole of Central Anatolia and the southern shore of the Black sea but it does not proceed anywhere else: it essentially dies within Anatolia. We can identify it with what is known to be a separate family of (all extinct) languages within Indo-European: the Anatolian languages.

The fourth, North-West, branch, after crossing the Bosphorus departs north both from Bulgaria and from Odessa. It should be noticed that those two departure points are practically simultaneous (700 and 750 years), so that no linguistic distinction can be implied, and that the colonization proceeds through a thick ray-like pattern of contiguous branches of which we have traced only a few. Therefore, linguistic differentiation is here better seen as a function of time elapsed rather than geographic separation. This branch expands through the whole Slavic area and reaches the Baltic area after approximately 2200 years (cf. Figure 6). It, therefore, corresponds to the Balto-Slavic language group.

In sum, the simulation correctly predicts the existence of four major linguistic groups within the Indo-European family and, furthermore, it correctly predicts their internal constituency: Anatolian, Aryan-Armenic (with the Aryan later articulated into Indo-Iranian), Greco-Italo-Celtic (with Greek splitting off long before Italo-Celtic), Balto-Slavic. This degree of fitness between groupings of expanding farmers generated by the simulation and internal comparative linguistic analysis, two completely unrelated sets of data, seems to us, again, interesting evidence in favor of the hypothesized expansion process.

Two pieces of evidence stand instead in apparent contrast. They both concern Germanic. In traditional comparative analysis Germanic is classed with the Balto-Slavic group. Our simulation tells a different story. In Figure 9, we can clearly see an autonomous branch that occupies present-day Germany filling the entire space between the Celtic group to the west and the Slavic to the east. This branch originates, however, at approximately 1500 year, from the eastern Italic branch north of the Gulf of Trieste, corresponding to today Southern Austria. The Germanic group would thus be more closely related to the Italic. We have no explanation for this. However, it is worth noting that Cavalli-Sforza et al. [8] obtained the same close relation between Germanic and Italic by using both the UPGMA and the NJ reconstruction methods (see Cavalli-Sforza [9], fig. 15).

The second anomaly is related to the first in that it also involves Germanic. As it can be seen in Figure 9, the whole Scandinavian area is colonized by people coming from the Balto-Slavic group. Instead, Scandinavian languages are traditionally classed as a branch of Germanic.

On the other hand, one nonobvious phenomenon receives a clearcut explanation from the simulation. The great North European band of territory extending from the Atlantic coast of France to the Ural Mountains does not present any significant orographic barrier and hence no natural obstacle to the regular and continuous spreading of population from east to west or viceversa [10]. Yet both today and in early times it has always appeared as fractionated into three linguistic groups (albeit with variable borders): Celtic (French), Germanic, Slavic. The early Neolithic settlement, as determined by the purely demographic expansion and the nature of the terrain, if one accepts its origin in Anatolia, produces exactly this result. The three groups are sharply separated by the very dynamics of the expansion process. The central branching, the Germanic, starts its northern ascension from southern Austria at 1500 years from origin and proceeds for a long time completely separated to the east from the Slavic group coming up from Bulgaria and the Black Sea: the two will touch each other, determining a reciprocal border to their expansion, only 800-900 year later, at around 2300-2400 years from origin. The same thing happens for the separation between Germanic and Celtic. The Celtic line of expansion begins its ascension around 2200 and will reach and touch the

Germanic line of expansion to the east only more than a thousand years later, thus determining the second fracture line in the Northern European plain.

Worth of note is also the situation that the simulation describes for Italy. From Figure 9 we can clearly see that the Neolithic settling of Italy is entirely operated (with the exception of a very thin Tyrrhenian coastal strip) by people coming from the Adriatic and crossing the Appennine mountains in successive parallel waves from south to north. This type of movement is widely attested, at least for later epochs, although perhaps the Tyrrhenian expansion may have been more important if we assume a higher level of navigational technology/skill (see above). These same populations also cross the Adriatic sea and expand into the Balkans. Today this territory is of late Slavic occupation. Not so in ancient times when it was occupied by so-called Illyrian populations, whose languages are all extinct: the simulation predicts their close relation to the Italic languages.

Finally, the whole process of Italian peninsular colonization, taking place from south (Apulia) to north (Po Plain) and from east to west, takes about 1000 year. If we place the origin of expansion in Anatolia at 6500 year BC, Italian Neolithic settlement happened between 5500 and 4500 BC along the specified path: available archaeological evidence is in substantial agreement with this prediction.

The remarkable correspondence between, on one side, the branching tree of expansion paths generated by our simulation, with the associated time depths at different geographical locations, and on the other side, the similarity tree of European languages gives some support to one of the two main hypotheses concerning the origins of European languages: the Proto-Indo-European language was spoken in Anatolia nine thousand years ago and it expanded and differentiated in the subsequent millennia in Europe along with the diffusion of farming. This hypothesis appears to be rather well supported by the evidence [11] [12] and is consistent with other evidence which links language differentiation with diffusion of farming in other parts of the Earth [13]. The alternative hypothesis, more popular among historical linguists and based on glottochronological analyses and time depth considerations, is that Proto-Indo-European was instead spoken by nomadic people living north of the Black and Caspian Seas 5-6 thousand years ago [14] [15]. This alternative hypothesis assumes that these nomadic peoples or, more probably, their culture/language expanded in Europe in an east-bound direction and that the similarity structure of European languages is the result of this expansion. We cannot test this alternative hypothesis with the current model because the cell parameters and the expansion rules used in our model are appropriate for earlier times and for another kind of expansion (farming). However, it should in principle be possibile to modify our model by selecting different cell properties and different expansion rules for the cellular automata that are more appropriate for a later expansion of a different nature. In this way one could determine whether the resulting tree of expansion lines matches the similarity tree of European languages equally well as the current version of the model which is inspired by the hypothesis of a more ancient Anatolian origin associated with farming.

5. Demic or cultural?

The model that we have described presupposes a demic interpretation of the arrival of farming in Europe. Farmers actually moved from Anatolia and their descendants (i.e., their genes) progressively occupied the entire Europe. This interpretation may explain the genetic gradient which has been found with point of origin in the Near East and gradual decrements as one moves North and West in Europe [16]. However, the model can be modified to incorporate a cultural rather than demic interpretation of the expansion of farming in Europe. The model as we have described it ignores the Mesolithic people already living in Europe and adopting a hunting and gathering subsistence strategy. However, more recent analyses of mithochondrial and Y chromosome DNA suggest that the demic component of farming diffusion in Europe, i.e., the replacement of Mesolithic hunters/gatherers by Neolithic farmers, may have been less strong than assumed by the demic expansion model [17] [18] [19]. Our model can be interpreted as assuming not that Mesolithic people were physically (genetically) canceled by the expanding farmers but that they adopted farming by cultural/technological assimilation to the successful farmers living in neighboring cells. In this cultural interpretation of the model farming, not farmers, expanded in previously nonfarming cells, with a greater probability for farming to be adopted by cells which were more suitable for farming.

We have run this cultural version of the model with results that are generally similar to those of the demic version. In particular, the tree of expansion paths matches the similarity structure of European languages equally well as the tree obtained with the demic model. This seems to indicate that geographical factors linked to suitability for farming of different regions of Europe and the point of origin of the process in Anatolia may be the critical factors that determine both the branching tree of expansions paths and the similarity structure of European languages. These factors play a role in both the demic and the cultural versions of the model.

One difference between the two models is that with the cultural model farming invades the whole of Europe in hundreds, not thousands, of cycles of the cellar automaton. But the cellular automaton should be temporally re-calibrated if the expansion is interpreted as cultural. In the demic model the expansion of farming is linked to demographic growth which can be very fast and can lead to migration of farmers in one year. Cultural adoption of farming may be much slower and one cycle of the cellular automata may correspond not to one year but to, say, 10 years. With these adjustments, there is correspondence between these two models even from this point of view.

Of course, one might also assume, and test, a mixed demic and cultural model, with the early expansion of farming in the geographical regions near to Anatolia mainly due to demic factors and later diffusion in the rest of Europe mainly due to cultural factors. In any case an important consequence of considering the cultural interpretation of the model is that additional historical data, not included in the present version of the model, can be integrated in the model such as demographic and geographical distribution data on Mesolithic people already living in Europe, including the changes in these data during the period considered [17] [18] [19].

6. Further developments of the model

The current version of the model assumes that languages or cultures with a common origin become more and more different with the increasing length of time elapsed since separation. But the phenomenon of progressive divergence which is simply “assumed” in the current version of the model can be actually “observed” by modifying the model.

The current model assumes that progressive differentiation is a function of time since separation in that internal changes accumulate in a culture or in a language and, provided there are no interactions among cells that may lead to unidirectional or reciprocal assimilation, these changes lead to progressive divergence with time. In the case of language this may result in the appearence first of different dialects and then of different languages. But imagine we explicitly model culture or language as a string of bits, with each bit indicating the presence (1) or absence (0) of a cultural or linguistic trait. The initial Anatolian cell is assigned a randomly generated bitstring that represents the culture or the language spoken by the originary farmers in Anatolia nine thousand years ago and expansion is modeled as copying the bitstring of an expanding cell in the cell which is being occupied. Intrinsic changes in culture/language are modeled as random changes in bitstrings that can occur with a given probability. This allows us to actually observe (in the simulation) the progressive differentiation that occurs as a function of the time elapsed from separation from a common cell, i.e., from a shared culture/language. Random changes in bitstrings accumulate and lead to progressive divergence.

The new model, which explicitly, althoughy very schematically, represents cultures/languages, makes it possible to observe two important phenomena which it was impossible to simulate with the preceding version. First, the new model can simulate culture/language contact and cultural/linguistic influence or assimilation, which is another mechanism for cultural/linguistic change in addition to internal changes. The model can implement a variety of cultural or linguistic assimilation rules that can all be tested to identify the rule that gives the best approximation to historical data. One assimilation rule is that in each cycle the bitstring of every cell is modified by making, with a given probability, the value of one or more of the bits of the cell’s bitstring identical to the value of the corresponding bit of the majority of neighboring cells that contain a bitstring (i.e., of already occupied neighboring cells). This implements the notion of “frequency bias” proposed by Boyd and Richerson in their model of cultural change [20]: cultural/linguistic change is towards the most frequent value of a given trait. Another assimilation rule assumes that the value of each bit in a cell’s bitstring is made identical, again with a given probability, to the value of the corresponding bit of one of the neighboring cells, either randomly selected or selected on the basis of the already existing similarity between the two cells - with existing similarity favouring further assimilation. (This last rule was proposed by Axelrod [21]).

It has already been shown, using abstract cellular automata in which there is no geography, i.e., no natural obstacles to interaction among cells, that starting from randomly assigned bitstrings and using a variety of assimilation rules, assimilation between neighboring cells does not lead to complete homogeneity, with all the cells eventually containing the same bitstring [21] [22]. On the contrary, stable boundaries tend to emerge between internally homogeneous neighboring regions comprising many cells (cultural or linguistic communities) and these boundaries keep the bitstrings of different regions different. If one uses Axelrod’s assimilation rule which postulates assimilation with a single neighboring cell based on already existing similarity, this results from the fact that strong existing dissimilarity prevents assimilation. If one uses the majority assimilation rule, stable boundaries emerge because greater internal pressures not to change on a cell currently belonging to a given community tend to win over external pressures to change emanating from other neghboring communities. This is true even if the bitstrings of all the cells of the cellular automaton originate from a single cell possessing a single bitstring, as in our model, and are the result of a process of expansion accompanied by the random internal changes in each cell’s bitstring that we have already discussed [22]. Furthermore, if one adds to the model an abstract geography of “mountains” separating different groups of cells, the eventual number of different cultural/linguistic regions turns out to be even greater [22].

A second important phenomenon that can be observed with the new version of the model is the actual geographical extent of internally homogeneous cultural/linguistic regions/communities. The preceding version allowed us to identify the date of arrival to isolated cells in Europe and the path followed to reach each cell, and this made it possible to measure similarity between cells. However, the preceding model tells us nothing about the geographical limits of multi-cell areas possessing the same culture or language. Explicitly representing cultures/languages as bitstrings allows us to determine which neighboring cells have the same culture/language, and therefore the geographical boundaries of internally homogeneous regions, and to observe how the number of different cultures/languages increases in Europe starting from the single culture/language of the originary Anatolian cell.

Preliminary experiments show that the number of different cultures/languages in Europe does not increase very much in the first 1000 cycles (first 1000 years after the beginning of the process, in the demic expansion model), increases moderately up until around cycle 3000, and has a more rapid increase after cycle 3000. However, if we allow the simulation to continue well beyond the completion of the farming diffusion process in the whole of Europe until cycle 8000, the number of different cultural/linguistic regions becomes stable. Predictably, the results tend to be sensitive not only to the type of assimilation rule which is adopted, but also to the length of the bitstring representing language/culture, with longer bitstrings producing more different cultural/linguistic regions, and to the rate of random mutations representing internal changes in cultures/languages, with higher rates again resulting in larger number of different regions. Furthermore, all these results are based on a notion of cultural/linguistic homogeneity which requires that all the cells in a cultural/linguistic region possess exactly the same bitstring. Weaker notions of cultural/linguistic homogeneity may require only a sufficient degree on similarity among cells belonging to the same region. Notice that a precise, quantitative notion of similarity/distance between two bitstrings can be used: the number of bits which have a different value in the two bitstrings (Hamming distance). This may allow us to identify subregions (dialects or sub-cultures) inside a region and gradients which can make boundaries between neighboring regions fuzzy rather than clearcut.

A further development of the model is to define in terms of the model a notion of "ethnicity" and to make the probability of assimilation between the bitstrings of neighboring cells dependent on the degree of co-ethnicity of the cells. Two or more cells are considered as belonging to the same "ethnos" if the they have a recent common ancestor cell, i.e., if the time since separation from a common ancestor cell is not too great. Two neighboring cells can belong to the same "ethnos" but they can also belong to different "ethnos" if they result from two spatially diverging expansion paths which for some reason have re-converged. Cultural/linguistic assimilation between two cells can reflect their co-ethnicity if the probability of assimilation between the two cells is a negative function of time elapsed since the time the two cells had a common ancestor cell.

7. Conclusion

We have simulated using a cellular automaton a demic model of the expansion of farming in Europe which began nine thousand years ago in Anatolia and the correlated process of expansion and differentiation of European languages. The model is able to generate the yearly expansion rate of 1 km which appears to be supported by the archaeological evidence on the basis of local geographical data and the assumption of reasonable demographic expansion rates due to adoption of farming. Furthermore, by modeling the actual geographical paths followed by farmers in their expansion in Europe and their time of arrival to various geographical locations, the simulation generates a tree of progressive linguistic differentiation which matches rather well the similarity tree of European languages posited by linguists.

It is quite probable that the demic model of agricultural/linguistic expansion in Europe will have to take into consideration various expansion and migration processes which took place in Europe both before and after the Neolithic expansion and that it will have to consider in more detail the actual processes of demic substitution vs. technological/cultural expansion and of language change. However, the simulation reported in this paper appears to give some support to the hypothesis that links the expansion of farming to the differentiation of European languages and in any case it seems to capture important aspects of what took place in Europe during the Neolithic.

However, the goal of the present chapter was not to support a particular hypothesis about the origin and expansion of farming in Europe or about the origin and differentiation of European languages. Our goal was to illustrate with some concrete examples the potential of computer simulations as a new way to express scientific theories and models in the social and historical sciences.

Simulations make it inevitable to express theories and models in explicit, detailed, quantitative terms, which is rarely the case for theories and models in the social and historical sciences. For example, in our simulation model we had to specify, in explicit, precise, quantitative terms, the critical factors that can account for the historical phenomena of farming expansion and linguistic differentiation: geographical maps, level of navigational technology and skills, demographical growth rates, types of cultural/linguistic assimilation rules, etc.

Furthermore, simulations generate a rich variety of results that represent the empirical predictions derived from the theory or model expressed in the simulation, and this gives a rich and detailed empirical content to theories which normally lack such a content and guarantees that specific empirical predictions actually derive from a theory/model. Our simulation generates results on time taken to reach particular regions in Europe, on paths followed to reach these regions, etc., and these predictions are guaranteed to actually derive from the assumptions of the model incorporated in the simulation.

Finally, simulations are virtual experimental laboratories in which the researcher can manipulate the conditions, variables, and value of parameters which influence the results that are obtained. In our simulation we can manipulate such conditions and parameter values as the geographical map underlying the simulation, the level of navigational technology/skill, the assimilation rules, the probability of internal changes in culture/languages, etc. This allows the researcher in the social and historical sciences to do experiments - a very powerful tool in the natural sciences - with phenomena that are too big, too complex, or are simply no more existent to be brought into the physical laboratory.

References

[1] Von Neumann, J. *Theory of Self-Reproducing Automata*. Urbana, Ill., University of Illinois, Press, 1966.

[2] Wolfram, S. *A New Kind of Science*. Wolfram Media, 2002.

[3] Parisi, D. A cellular automata model of the expansion of the Assyrian empire. In S. Bandini, R. Serra, & F.S. Liverani (eds.) *Cellular Automata: Research towards Industry*. London, Springer, 1998.

[4] Antinucci, F., Cecconi, F., Natale, F., & Parisi, D. *Simulating the Indo-European Expansion through a Cellular Automata*. Institute of Cognitive Sciences and Technologies, National Research Council Rome, 2002.

[5] Ammerman, A.J., & Cavalli-Sforza, L.L. *Neolithic Transition and the Genetics of Population in Europe*. Princeton, Princeton University Press, 1984.

[6] Renfrew, C. *Archaeology and Language. The Puzzle of Indo-European Origins*. London, Jonathan Cape,1987.

[7] Semino, O., Passarino, G., Brega, A., Fellous, M. & Santachiara-Benerecetti, A. S. A view of the Neolithic demic diffusion in Europe through two Y chromosome-specific markers. *American Journal of Human Genetics*, 59, 1996, 964-968.

[8] Cavalli-Sforza, L.L., Minch, E. & Mountain, J. Coevolution of genes and language revisited. *Proceedings of the National Academy of Science*, 1992, 89, 5620-5624.

[9] Cavalli-Sforza, L.L. *Genes, Peoples, and Languages*. Los Angeles, University of California Press, 2001.

[10] Diamond, J. *Guns, Germs, and Steel: the Fate of Human Societies*. New York. Norton, 1997.

[11] Barbujani, G. & Bertorelle, G. Genetics and the population history of Europe. *Proceedings of the National Association of Sciences*, 2001, 98, 22-25.

[12] Bellwood, R. Early agriculturalist population diasporas? Farming, languages, and genes. *Annual Review of Anthropology*, 2001, 30, 181-207.

[13] Diamond & Bellwood

[14] Gimbutas, L. Proto-Indo-European culture: the Kurgan culture during the fifth, fourth and third millennia B.C. In G. Cardona, H.M. Hoenigswald & A. Senn (eds.) *Indo-European and Indo-Europeans*, Philadephia, University of Pennsylvania Press,1970, 155-197.

[15] Gimbutas, M. The Kurgan wave migration (c. 3400-3200 B.C.) into Europe and the following transformation of culture. *Journal of Near Eastern Studies*,1980, 8, 273-315.

[16] Cavalli-Sforza, L.L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes*. Princeton, Princeton University Press, 1994.

[17] Semino, O., Passarono, G., Oefner, P.J., Lin, A.A., Arbuzova, S., Beckman, L.E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiae, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, A.S., Cavalli-Sforza, L.L. & Underhill, P.A. The genetic legacy of Paleolithic Homo Sapiens in extant Europeans: A Y chromosome perspective, *Science*, 2000, 290, 1155-1159.

[18] Torroni, A., Bandelt, H.J., D'Urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savontaus, M.L., Bonne-Tamir, B. & Scozzari, R. mtDNA analysis reveals a major late Paleolithic population expansion from Southwestern to Northeastern Europe, *American Journal of Human Genetics*, 1998, 62, 1137-1152.

[19] Torroni, A., Bandelt, H.J., Macaulay, V., Richards, M., Cruciani, F., Rengo, C., Martinez-Cabrera, V., Villems, R., Kivisild, T., Metspalu, E., Parik, J., Tolk, H.V., Tambets, K., Forster, P., Karger, B., Francalacci, P., Rudan, P., Janicijevic, B., Rickards, O., Savontaus, M.L., Huoponen, K., Laitinen, V., Koivumaki, S., Sykes, B., Hickey, E., Novelletto, A., Moral, P., Sellitto, D., Coppa, A., Al-Zaheri, N., Santachiara-Benerecetti, A.S., Semino, P. & Scozzari, R. A signal, from human mtDNA, of post-glacial recolonization in Europe. *American Journal of Human Genetics*, 2001, 69, 844-852.

[20] Boyd, R. & Richerson, P.J. *Culture and the Evolutionary Process*. Chicago, Chicago University Press, 1985.

[21] Axelrod, R. The dissemination of culture: a model with local convergence and global polarization. *Journal of Conflict Resolution*, 1997, 41, 203-226.

[22] Parisi, D., Cecconi, F. & Natale, F. Cultural change in spatial environments: the role of cultural assimilation and of internal changes in cultures. *Journal of Conflict Resolution*, 2003, 47, 163-179.

Figure 1. Arrival to Greece through Rhodes at around 670 years: distance from origin (DFO) 640 km. Arrival to southern coast of Bulgaria after crossing the Bosphorus at around 700 years: DFO 700 km.

Figure 2. Arrival to Odessa along the Black Sea coast at 770 years: DFO 750 km. Arrival to Italy (Otranto, Apulia) at about 1000 years: DFO 1100 km.

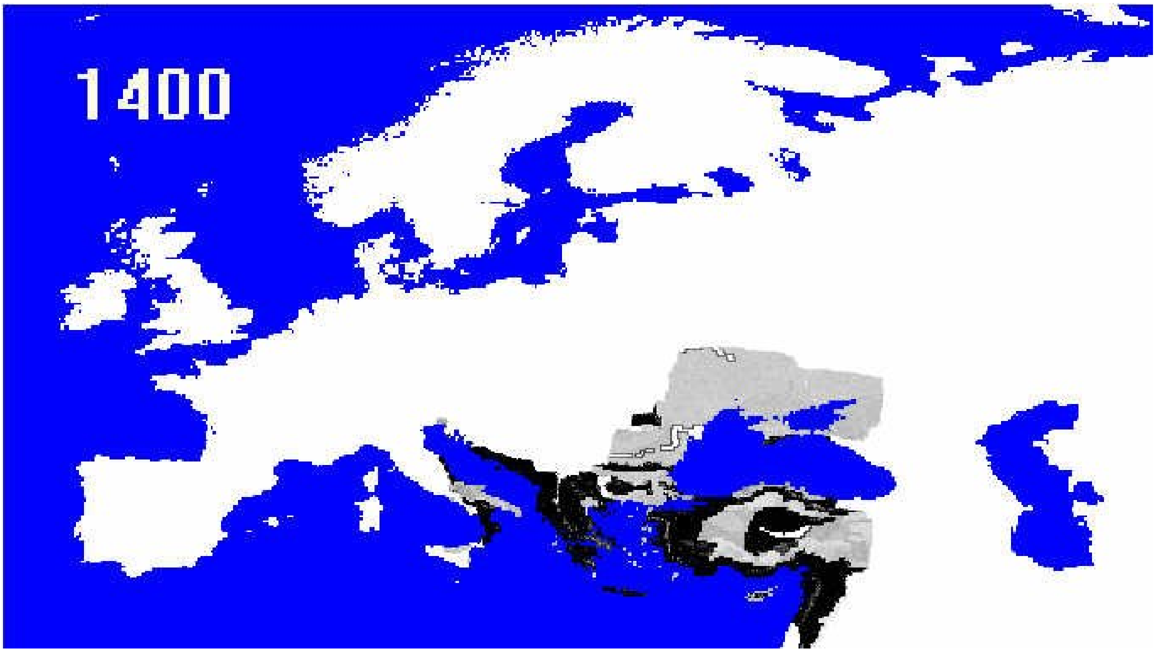

Figure 3. Arrival to Ancona (Adriatic coast) at 1400 years: DFO 1550 km.

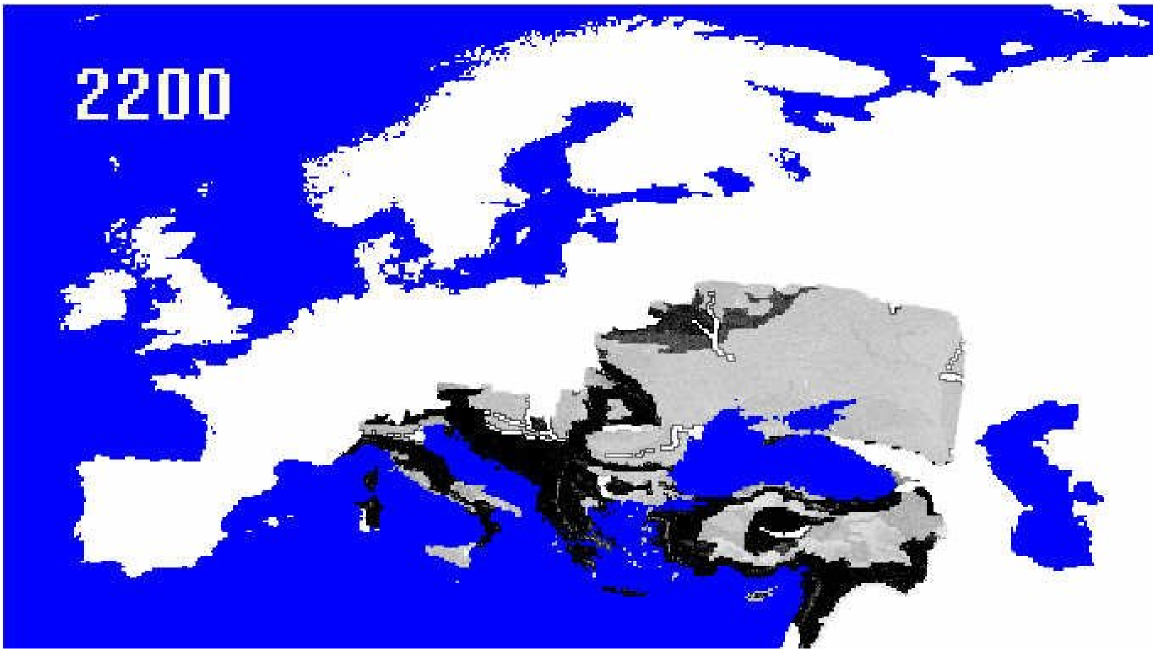

Figure 4. Arrival to Provence via Corsica and Sardinia at about 2200 years: DFO 2200 km.

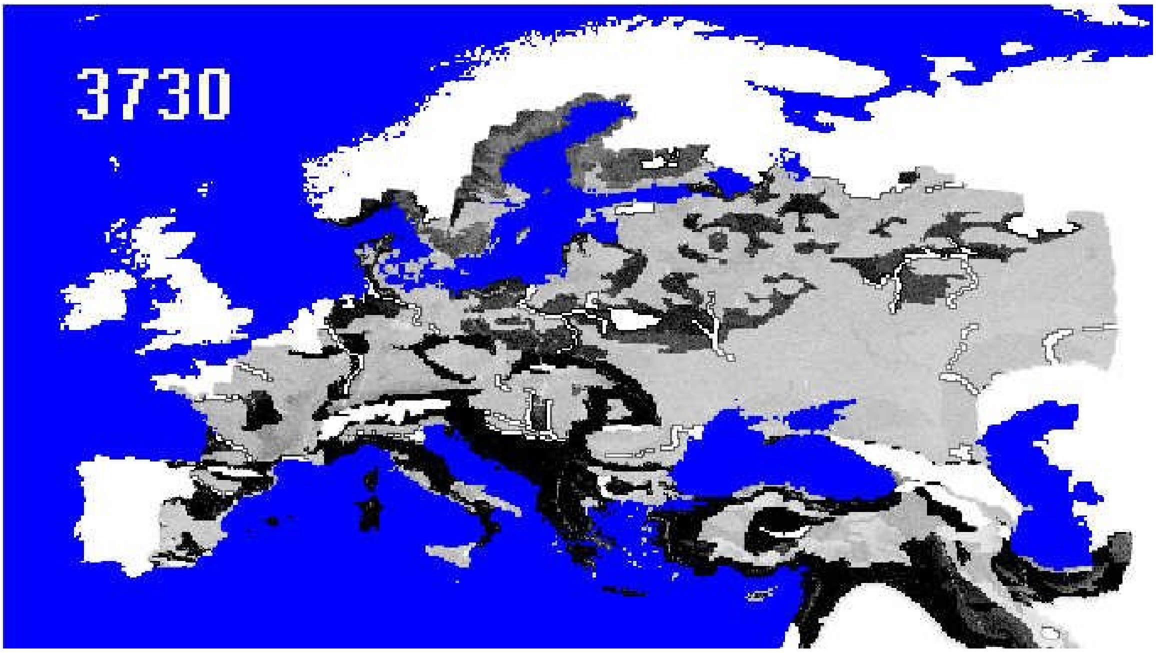

Figure 5. Arrival to England at 3730 years: DFO 3600 km.

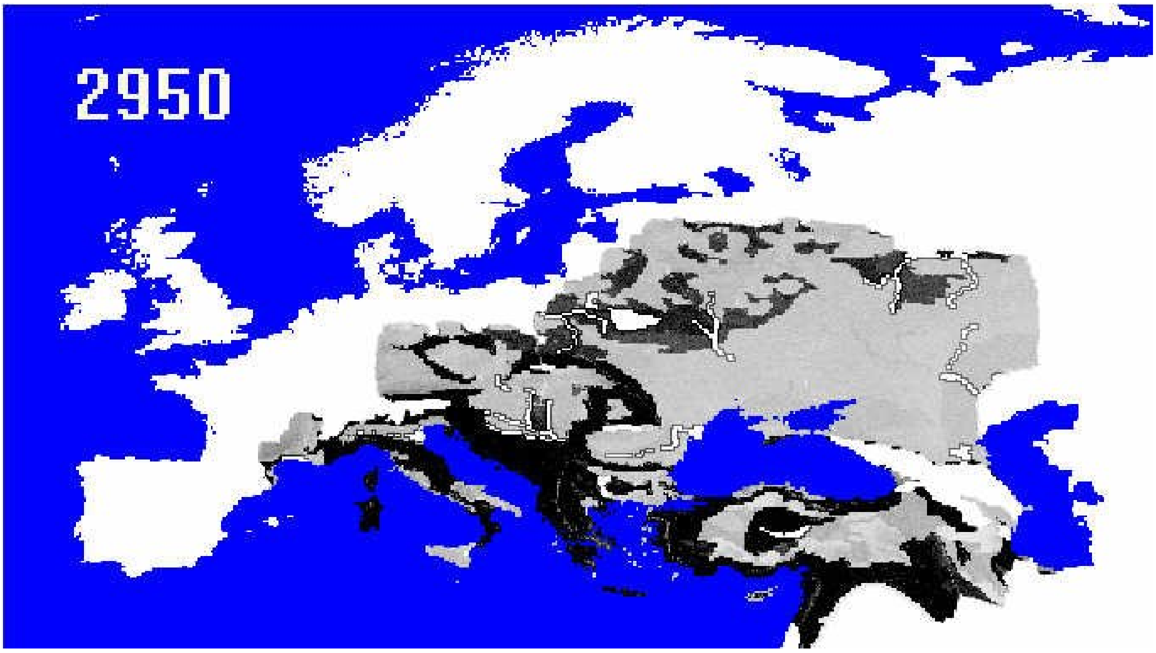

Figure 6. Arrival to Baltic at 2950 years: DFO 2650 km.

Figure 7. Arrival to Armenia at 1750 years: DFO 1650 km.

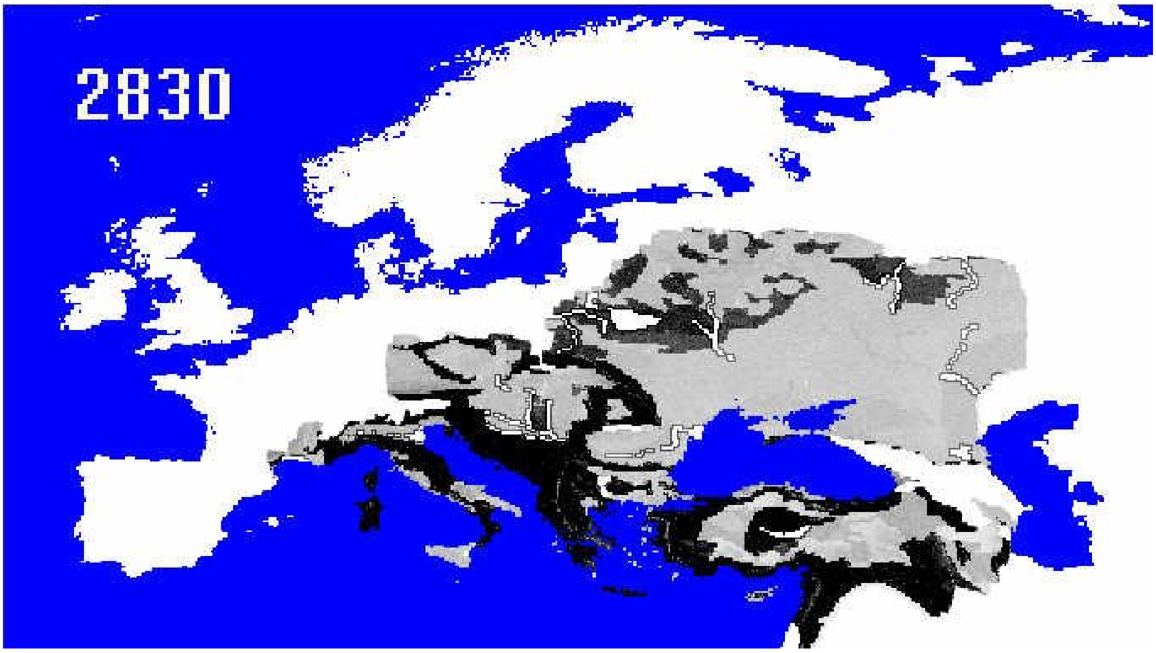

Figure 8. Arrival to Iran at 2830: DFO 2650 km.

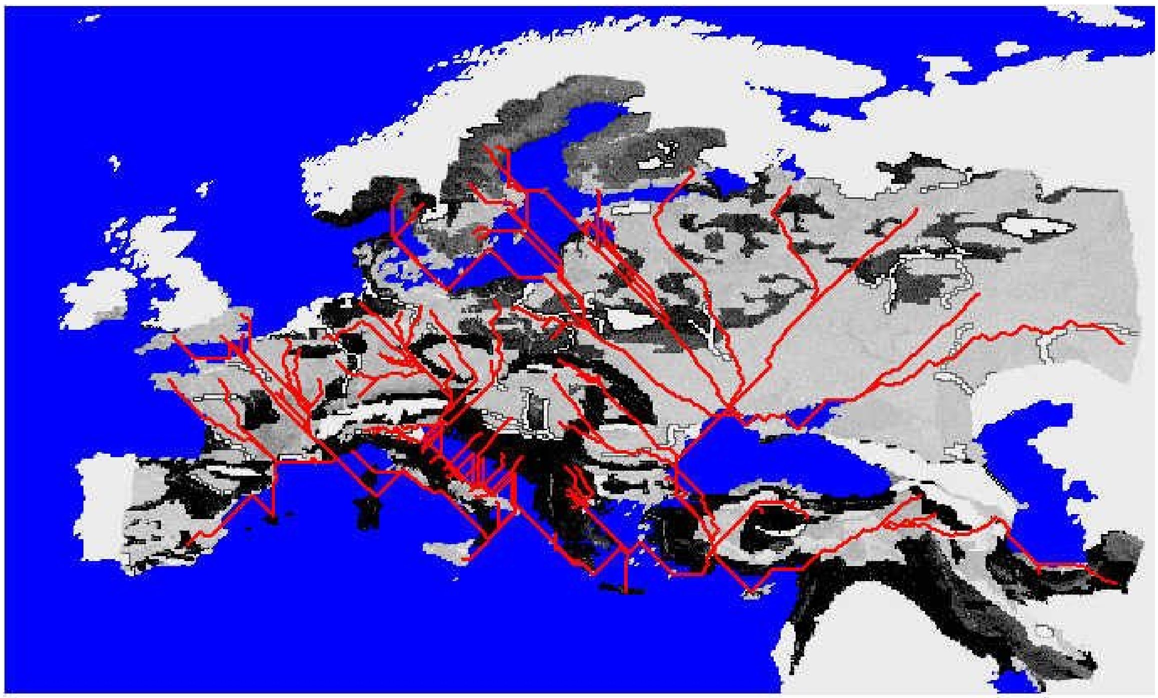

Figure 9. Branching tree of expansion paths with root in the point of origin of the expansion process.