SI Appendix: Modeling the role of voyaging in the coastal spread of the Early Neolithic in the West Mediterranean

Neus Isern\textsuperscript{a,1}, João Zilhão\textsuperscript{b,c}, Joaquim Fort\textsuperscript{a,c}, and Albert J. Ammerman\textsuperscript{d}

\textsuperscript{a} Complex Systems Lab and Physics Department, Universitat de Girona, 17003 Girona, Spain.
\textsuperscript{b} Departament d’Història i Arqueologia (SERP; Grup de Recerca SGR2014-00108), University of Barcelona, 08001 Barcelona, Spain.
\textsuperscript{c} Catalan Institute for Research and Advanced Studies (ICREA), 08010 Barcelona, Spain.
\textsuperscript{d} Department of Classics, Colgate University, Hamilton, NY 13346, United States of America.
\textsuperscript{1} Corresponding author. Email: neus.isern@udg.edu

Database notes

Selection criteria

When preparing the database, only the earliest date for each site was retained, following the criteria in Ref. (1), i.e., excluding any and all rank 7 results, relying principally on results ranked 1-4, and resorting to results in ranks 5-6 only when nothing better was available. Despite the problems pointed out in Ref. (2), the database accepts the Nerja/Vestíbulo date (rank 5) on the basis of consistency with results for the adjacent geography (2). Two dates on shell, for the Portuguese sites of Cabranosa and Padrão, are included; their calibrated ages were calculated with the terrestrial curve after correction for the reservoir effect following Ref. (3).

The stringent criteria used in the selection of dating results find justification in the nature of the research question. The use of statistical techniques can be of help in sorting out signal from noise when dealing with large datasets, and such an approach makes sense when the questions being addressed concern such things as the duration of an archeological culture or the timing of its floruit. When dealing with “first arrival” issues, however, the “chronometric hygiene” approach used here is mandatory, and indeed the standard practice in comparable situations, namely the spread of the Lapita complex across the Pacific, or the time of settlement of New Zealand or the Hawaiian archipelago, not to mention, in the Mediterranean basin itself, the time of first human colonization of the Balearic Islands (4-6).

Two examples suffice to illustrate why, for the study of the spread of farming across the West Mediterranean, a “chronometric hygiene” approach is unavoidable. When comparing the two extremes of the distribution of the Impressa-Cardial complex, southern Italy and Portugal, a time lag of some 400 years is readily apparent; yet, as shown by the dating of the lakeside settlement of La Draga, that much can be the difference between the real age of the settlement as derived from short-lived, bone or cereal samples, and the apparent age obtained when dating the oak trees therein used as timber (7). The other example concerns the dating of Cueva de Nerja, and shows how the pattern of spread can be affected not only by old wood effects but also by issues of sample identification; here, a bone initially classified
SI Appendix. Modeling the West Mediterranean Neolithic spread

Zoo-archeologically as domesticated oviscaprid returned a result some 700 years too old, falling squarely in the time range of the regional Mesolithic and indicating that the sample was in fact of the extant, wild species *Capra pyrenaica* (2). Needless to say, uncritical use of such problematic dates would result in a “first arrival” scenario quite distinct from that in Fig. 1.

The downside of the “chronometric hygiene” approach consists in the risk of losing, or missing information concerning areas were quality dates are too few, resulting in underestimated “first arrival” dates for those particular areas. In our case, however, this is not a real problem, for several reasons, namely: (a) we are dealing with a process of spatio-temporal expansion in which the key factor in the calculation of such parameters as the rate of spread is a good control of the dating at the two ends—here, western Liguria and central Portugal—and the number of quality results in our dataset is, for these areas, sufficient for the chronological boundaries of the process to be considered robust; and (b) a rather large number of quality results is available for two key intermediate areas, Catalonia and Valencia, and they are consistent with the gradient to be expected if the process indeed was one of westward expansion with an origin in northern Italy.

The database was compiled in 2014 using the then available data and the models were developed on that basis; here (Dataset S1), we append to it results that could not be used in the models because they were obtained or published since (or only since came to our attention). For the sake of completeness, this new set includes one result for the open-air site of Cabecicos Negros (Almería) obtained on a sample of *Cerastoderma edule*, which we calibrated with the marine curve, as the local reservoir effect is unknown. Note that the site is located 2 km inland, meaning that the sample, if derived from a live mollusk collected for consumption, most probably comes from the immediate brackish water environment described by Camalich-Massieu et al. (8); therefore, the difference between its true and apparent ages may well be even larger than if the sample had been collected in a marine environment. Except for a clearly anomalous 7280±40 BP date (Beta-347630) that most likely reflects the use of dead, beach-collected shells of *C. edule* for the manufacture of beads, an activity well documented at the site (9), the result we retained is the oldest date obtained for Cabecicos Negros (where all dates are on samples of that same taxon).

The appended results remain entirely consistent with the general picture in Fig. 1, supporting the notion that we can be confident that, within statistical error, we have indeed picked-up the time of first appearance of farming in the areas concerned. We have compared a few test runs against such new results (10-12), and indeed have found that at least in the case of model 4 (with no interaction and jumps of 450 km), the simulations remain consistent with them. Even though their inclusion might affect the goodness of the prediction for particular sets of parameters, we expect these additional results to modify neither the global picture nor our conclusions.

The only major geographical gap in our dataset concerning an area with a well-known and rather dense network of Early Neolithic sites is the French Languedoc, as is well apparent in Fig. 1. The lack of quality dates for these sites means that there is at present little that can be done to remedy the situation. On the other hand, this example serves to highlight another advantage of our “chronometric hygiene” approach, namely the possibility to use the general pattern in Fig. 3 to predict a time of first arrival of
the Neolithic in that area around 5650 cal BC. Indeed, such a time frame is suggested by some of the charcoal dates obtained for the sites of Peiro Signado and Pont de Roque-Haute, and would be consistent with the associated Ligurian ceramics (13).

Comparison with model results
As shown in Fig. 1, the archaeological data are clustered in nine areas, which we denote as: Italy/France, Catalonia, Valencia, Andalusia, Northern Africa, Algarve, Central Portugal, Cantabria and Interior. Even though only the first seven areas, those directly related to the coastal spread along the Mediterranean shores, are of interest here, we also compare the results with the Interior area as a reference for the inland expansion.

Since we are interested in the patterning of the spread, testing whether the models correctly match the archaeological evidence means assessing a given simulation run's goodness of fit for each area. The eight dates selected to test the simulation outcomes (represented by diamonds in Fig. 1b) are listed in Table S1. We consider that the model predicts the observed arrival times in the respective areas correctly if the simulated arrival times fall within the 2σ calibrated range of these eight earliest dates. Otherwise, the simulations may be too slow (if some areas are reached after the 2σ range) or too fast (if some areas are reached before the 2σ range).

<table>
<thead>
<tr>
<th>Area</th>
<th>Site</th>
<th>2σ cal BC range</th>
<th>Median of the cal BC range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italy/France</td>
<td>Arene Candide</td>
<td>5,638 – 5,791</td>
<td>5,711</td>
</tr>
<tr>
<td>Catalonia</td>
<td>Guixeres de Vilobí</td>
<td>5,490 – 5,643</td>
<td>5,582</td>
</tr>
<tr>
<td>Valencia</td>
<td>Mas d’Is</td>
<td>5,481 – 5,621</td>
<td>5,547</td>
</tr>
<tr>
<td>Andalusia</td>
<td>Cueva de Nerja</td>
<td>5,481 – 5,616</td>
<td>5,538</td>
</tr>
<tr>
<td>Northern Africa</td>
<td>Kef Taht el Ghar</td>
<td>5,076 – 5,483</td>
<td>5,333</td>
</tr>
<tr>
<td>Algarve</td>
<td>Cabranosa</td>
<td>5,374 – 5,623</td>
<td>5,513</td>
</tr>
<tr>
<td>Central Portugal</td>
<td>Galeria da Cisterna</td>
<td>5,326 – 5,481</td>
<td>5,414</td>
</tr>
<tr>
<td>Interior</td>
<td>Cueva de Chaves</td>
<td>5,478 – 5,615</td>
<td>5,528</td>
</tr>
</tbody>
</table>

Computational model: extended description
We have developed four computational models with the aim of simulating and reproducing the Neolithic coastal spread in the West Mediterranean. Here we present an extended description of our models, which have been outlined in the Methods section. We have written the models in Fortran 90/95, and the code is available at http://copernic.udg.es/QuimFort/WestMedNeo_Program_Files.zip.

Our model follows a dispersion-interaction-reproduction scheme, which allows us to study different possibilities for each of the three steps. The two main effects we focus on are: the importance of coastal voyaging (which we explore using different travelling kernels) and the effect of cultural transmission (which allows us to explore a range of processes, from purely demic to mixed demic-cultural with varying degrees of interbreeding and acculturation).

We run our simulations on a grid of square cells of 50 km x 50 km, prepared using an Albers Equal Area Conic Projection for Europe, which we have adapted from Ref. (14). The 50 km value is the characteristic
distance moved by preindustrial farmers per generation, as estimated from observed mobility and persistence data (see (15), note [32]). In addition, the land-based model studied by Bernabeu Aubán et al. (16) for Iberia suggested that settlements were most likely spaced about 25–100 km.

Each cell in the grid is identified as being either a sea cell, a coastal cell or an inland cell. As shown in Fig. S1, coastal cells are defined so that no inland cell has a sea cell on any of its four nearest neighbors.

In agreement with archeological evidence, we assume that Neolithic populations expanding along the West Mediterranean coast originate from northwest Italy. We take as the starting point of our simulations a cell located in northern Italy, at the northeastern extreme of our region of study, and with central coordinates (44.474N, 8.407E). The site of Arene Candide (site 1 in Table S1), which is the easternmost site in all figures and has the oldest calibrated date median, 7660 cal BP, falls within this cell. Using its date as representative, the simulations begin at 7700 BP (5751 BC). In the simulations we also explore the consequences of shifting this starting point one or two cells westward as well as eastward (keeping the same starting date).

We initially set the Neolithic population of this source cell at its maximum value, $N_{\text{max}} = 3,200$ individuals (initially, all other cells have no farming population). This value for the maximum population per cell is obtained from a realistic value of the carrying capacity of Neolithic populations, 1.28 individuals/km$^2$ (17)—knowing that each cell covers an area of 50 km × 50 km.

When we consider Neolithic-Mesolithic interactions, we assume that Mesolithic populations are present only in the coastal cells, following the currently available archaeological data (18). We estimate the Mesolithic population number and distribution from data on the aboriginal population of Southwest Tasmania (19), a hunter-gatherer population living in coastal and riverside contexts. From these data, we assume that the Mesolithic bands were distributed approximately every 50 km (one band per cell) with a population of 50–80 individuals per band.

**Fig. S1:** Model grid. The diagram shows the three kinds of cells used in the model: inland cells in brown, coastal cells in yellow and sea cells in light blue (coastal cells are defined so that no inland cell has a sea cell as any of its four nearest neighbors). The red lines joining coastal cells indicate how the coastal distance between cells is computed.
The simulations run for several iterations, each one corresponding to a generation, until the Neolithic has spread through all of the West Mediterranean. We use a realistic value of the generation time $T = 32\,\text{yr} \,(20)$; this corresponds to the average age difference between the parents and one of their children measured for pre-industrial populations. At the end of each iteration, we identify the position of the Neolithic front as the cells that have reached for the first time a population of 300 Neolithic individuals during that iteration. We use this value, about 10% of the maximum population per cell in the simulations (see above), as a practical threshold slightly above the minimum size required for a human reproductive network to be viable (21) and bearing in mind that the earliest archeologically detected Neolithic in a given cell may have had as-yet undetected local settlement antecedents. Using a lower or higher threshold would modify neither the minimum sea-travel range nor the ranges that yield the best results. The arrival times obtained for each cell can then be plotted on a map and/or be compared with the archeological data in order to quantify the goodness of fit with the simulated results.

The fact that our models use a dispersion-interaction-reproduction scheme means that at every iteration—namely, a generation—the model performs three sequential steps, which involve population dispersal, population interaction and population growth. We describe the three steps below. Changing the order of the three steps does not affect the results.

**Step 1: Population dispersal**

Regarding the dispersal process, we have designed four different approaches to describe how to include voyaging when modelling the West Mediterranean expansion. In all cases, voyaging takes place as a cabotage process (i.e. traveling along the coast); the differences between the four approaches depend then on how probable it is for the voyaging first farmers to settle at the possible destinations.

In all of our models, at each iteration and for each inhabited cell, a fraction $p_c = 0.38$ (persistence) of the Neolithic individuals remains in the same cell. This value for $p_c$ has been estimated previously from ethnographic data (15). The rest of the population moves to other cells according to the following travel rules, which depend on the type of the cell of origin: inland or coastal.

If the origin cell is an **inland cell**, the remaining individuals are equally distributed among the four nearest neighboring cells—each receiving a fraction $(1 - p_c)/4$ of the initial population (Fig. S2a). When a fractional value is obtained, the number of individuals who move is rounded to the nearest integer (with 0.5 rounded up to 1). The program checks to see that the final number of individuals—those who jump plus those who stay—is equal to the initial number; if necessary, the possible offset produced by rounding is applied to the number of individuals who stay at the cell of origin (the same rounding approach is applied when dealing with coastal cells).

When the cell of origin is a coastal cell, some of its nearest neighbors are sea cells, and thus ineligible destinations for settlement (Figs S2b, c). Below we describe how this situation is dealt with according to four different voyaging models.
Fig. S2: Land travel scheme. Black arrows stand for land travel when the origin is an inland cell (a), or a coastal cell (b,c). Red arrows stand for the fraction of the population that will travel by sea.

**Model 1. No voyaging.** In this first approximation, we do not allow sea travel. Thus, here, the individuals that would move to sea cells are redistributed in equal parts among the other neighboring cells that are eligible destinations (i.e., those classified as inland or coastal cells; black arrows in Figs. 2b, c). In Fig. S2b, for example, the three land neighbors would each receive \((1 - p_c) / 3\) of the original population and no first farmer would travel by sea.

**Model 2. Voyaging with decreasing probability to settle with distance.** In this model we assume that Neolithic individuals can travel by sea in the form of cabotage, establishing new settlements in each coastal cell within a certain range (e.g., 300 km), though with diminishing probability to settle with increasing distance (i.e., there is a preference to relocate close to home). Therefore, in this case, the individuals who would move from the cell of origin to sea cells are redistributed among the coastal cells within the travel range and according to a given probability distribution. We consider a Gaussian distribution

\[
p(x) = K_1 \exp \left( -\frac{1}{2} \frac{x^2}{\left(x_{\text{max}}/2\right)^2} \right),
\]

where \(x_{\text{max}}\) is the sea travel range, \(x\) is the distance to each coastal node between 50 km and \(x_{\text{max}}\), and \(K_1\) is a normalization constant. Fig. S3a shows how the voyaging population is distributed according to this model. Since we assume that voyaging takes place in the form of cabotage, the sea-travel distances are computed along the coast as the cumulative distance between coastal cells (as indicated by the red line in Fig. S1).

**Model 3. Voyaging with uniform probability to settle.** In this model we assume that Neolithic individuals can travel by sea and relocate along the coast up to a certain range (e.g., 300 km), with the probability of settling not depending on distance (i.e., all cells within the prescribed range receive the same fraction of voyagers; see Fig. S3b). Again, the distance between origin and destinations is computed along the coast, as shown in Fig. S1.
Fig. S3: Maritime voyaging scheme. Each box shows the relative distribution of the population after traveling by sea from the leftmost cell according to each of the three voyaging models, considering a range of 250 km (i.e., five aligned cells).

For models allowing sea travel (models 2–4), voyaging takes place in an onward direction. This means that voyaging at the front takes place toward unknown territories, rather than toward areas already occupied by first farmers. Increasing the probability that the individuals who relocate travel by sea rather than by land does not decrease the minimum sea-travel range necessary to fit the archaeological pattern, although slightly lower growth rates might be necessary.

We assume that the Mesolithic populations are well established and stationary in terms of population numbers, so any population movement between hunter-gatherer bands would be balanced.

**Step 2: Population interaction**

When assuming a purely demic process, the Neolithic population advances following a dispersal-reproduction scheme, so we would not apply this interaction step. However, if we consider that the expansion was a mixed cultural-demic process, we have to include cultural transmission in the model. Here we consider two processes: cross-mating (vertical cultural transmission) and acculturation (horizontal/oblique cultural transmission). At each iteration, we compute the new population incorporated into the farmer community per cell as a result of the interaction applying the following equations.

**Vertical cultural transmission.** This relates to cross-mating between populations (23). It is characterized by a parameter \( \eta \) (with values between 0 and 1) that indicates the degree of interbreeding. If \( N \) is the number of farmers and \( M \) the number of hunter-gatherers, then the number of hunter-gatherers that mate with farmers is (24)

\[
I_{VT} = \eta \frac{N \cdot M}{N + M},
\]

Note that this equation is not exactly the same as in Ref. (24). Here we are computing the number of mixed couples (or individuals of each population that will mate with individuals of the other population), while the interaction term in Ref. (24) computed the new population contributed by the mixed couples after reproduction; here we compute the population growth in the next step, which leads to the same final output.
In principle the case $\eta > 1$ is also possible, but it would lead to more cross-matings than under random mating (25), which is not realistic for the case of farmers and hunter-gatherers according to ethnographic parallels (25, 26).

**Horizontal/oblique cultural transmission.** This relates to acculturation of hunter-gatherers (23). Accordingly, it accounts for the number of hunter-gatherers that are incorporated in the farming community, though not for mating reasons (in other words, they can be couples). The number of Mesolithic people incorporated in the farming population due to this process is (27)

$$I_{HT} = f \frac{N \cdot M}{N + \gamma M}, \quad (S3)$$

where $f$ indicates the intensity of acculturation and $\gamma$ the preference by hunter-gatherers to copy either farmers ($\gamma < 1$) or other hunter-gatherers ($\gamma > 1$). However, as shown analytically in (27, 28), the speed of the Neolithic wave of advance depends only on a single acculturation parameter, namely $C = f / \gamma$. This parameter involves the number of hunter-gatherers incorporated in the farming community per pioneer farmer (i.e., when $N \approx 0$) and per generation (or it also could indicate the number of hunter-gatherer couples incorporated per pioneer farming couple and per generation, since they are not incorporated for mating reasons).

Since the value of $C$ is highly uncertain, we compute the results of applying horizontal/oblique cultural transmission for three illustrative values of the acculturation rate: $C = 0.2$ (five farmers are needed for a hunter-gatherer to adopt farming per generation), $C = 1$ (a hunter-gatherer is converted per every farmer and generation) and $C = 10$ (every farmer converts 10 hunter-gatherers into farming per generation). In the model, however, we have to input the values of the parameters $f$ and $\gamma$ when applying Eq. (S3). We have checked that, as expected, for a single value of $C = f / \gamma$, any other pair of values for the parameters $f$ and $\gamma$ yields roughly the same results (in agreement with Refs. (27, 28)), and thus we fix $\gamma = 10$ for the purpose of modelling the effect of acculturation.

**Step 3: Population growth**

The final population in each cell after a generation (iteration) is computed by applying a reproduction step. The population growth process it computed for each cell of the grid using a reproduction coefficient $R_0 = \exp(aT)$, where $a$ is the intrinsic growth rate and $T$ the generation time, and limiting the growth to the maximum value of individuals per cell $N_{max}$ (recall that $N_{max} = 3200$ individuals). Therefore, the number of farmers in each coastal or inland cell after the population growth process, $N(t + T)$, is computed as follows

$$\begin{cases} N(t + T) = R_0 N(t) & \text{if } N < N_{max} \\ N(t + T) = N_{max} & \text{if } N \geq N_{max} \end{cases}, \quad (S4)$$
where $N$ is the number of individuals per cell after the dispersion and interaction processes. Since the final number of farmers in a cell must be an integer value, the results from Eq. (S4) are rounded to the nearest integer.

Other approaches are also valid, e.g., a logistic (i.e., quadratic rather than linear) function in the first line in Eq. (S4); however Eq. (S4) does not yield negative population numbers when using finite-difference equations (as in the present study), in contrast with the logistic function (15, 29). Alternatively, we could use the complete solution of the logistic equation; however, including a higher-order approximation does not affect significantly the speed of the front.

Since we assume, as a first approximation, that the Mesolithic populations are well established and in a stationary state in terms of population numbers, we do not consider net population growth for Mesolithic individuals.

**Interaction parameters**

**Acculturation parameter**

In the main paper (Fig. 4) we show the results obtained when vertical and horizontal cultural transmission are applied with $\eta = 0.5$ and $C = 0.2$. Increasing the values of $C$ does not modify, however, the results significantly, as shown in Fig. S4 for $C = 1$ (Fig. S4a) and $C = 10$ (Fig. S4b). Therefore, the exact value of $C$ does not seem to be very important.

**Size of Mesolithic populations**

All of the results that include interaction discussed so far (both here and in the main paper) have been computed for bands of 50 hunter-gatherers. Increasing the number of people per band to 80 (also consistent with the evidence for Tasmanian populations (19)) allows for a faster spread, but very slightly, so that it does not change the conclusions reached. This is shown by the following figures (S5–S6), where the diamonds indicating the goodness of fit with the prediction remain unchanged when compared with the results for bands of 50 people (Figs. 4 and S4).
Fig. S4: Effect of combined vertical (cross-mating) and horizontal/oblique (acculturation) cultural transmission on the expansion process. Results for model 4, with $\sigma=1.8\%$, $\eta=0.5$ with bands of 50 people, and for different acculturation levels: (a) $C=1$, and (b) $C=10$. We used a sea-travel range of 350 km. White diamonds represent areas reached within the calibrated range; black diamonds represent areas reached by the model later than the calibrated range.
Fig. S5: This figure is the same as Figs. 4b and 4c in the main paper, but it is based on bands of 80 individuals (instead of 50 people). It shows the effect of Neolithic-Mesolithic interaction on the rate of spread. The results are those of model 4, with a sea-travel range of 350 km and intrinsic growth rate $a=1.8\%$ for two cases: (a) vertical cultural transmission only ($\eta=0.5$, $C=0$, as in the case of Fig. 4b), and (b) combined vertical and horizontal cultural transmission ($\eta=0.5$, $C=0.2$, as in the case of Fig. 4c). White diamonds represent areas reached within the calibrated range and black diamonds those areas reached later than the calibrated range (the pattern of the colored diamonds is the same as in Figs. 4b,c, so the conclusions do not change).
**Fig. S6**: This figure is the same as Fig. S4 above, but taking a band to have 80 individuals (instead of 50). It shows the combined effect of vertical and horizontal/oblique cultural transmission on the rate of spread. The results are for model 4, with $\sigma=1.8\%$, $\eta=0.5$ with bands of 80 people, and for different acculturation levels: (a) $C=1$, and (b) $C=10$. In all cases, the sea travel range is 350 km. White diamonds represent areas reached within the calibrated range and black diamonds areas reached later than the calibrated range (the pattern of diamonds is again the same as in Fig. S4, so the conclusions do not change).
References


