![data logo] **data**

Data Descriptor

A Data Set of Portuguese Traditional Recipes Based on Published Cookery Books

Alexandra Soveral Dias and Luís Silva Dias *

Department of Biology, University of Évora, Ap. 94, 7000-554 Évora, Portugal; alxandra@uevora.pt
* Correspondence: lsdias@uevora.pt; Tel.: +351-266-760-881

Abstract: This paper presents a data set resulting from the abstraction of books of traditional recipes for Portuguese cuisine. Only starters, main courses, side dishes, and soups were considered. Desserts, cakes, sweets, puddings, and pastries were not included. Recipes were characterized by the province and ingredients regardless of quantities or preparation. An exploratory characterization of recipes and ingredients is presented. Results show that Portuguese traditional recipes organize differently among the eleven provinces considered, setting up the basis for more detailed analyses of the 1382 recipes and 421 ingredients inventoried.

Dataset: available as www.mdpi.com/2306-5729/3/1/9/s1.

---

1. Summary

Eating is one of the obligatory activities of human beings, but unlike many other activities, eating usually requires active and deliberate effort. In humans, this occurs hopefully once or twice a day, every day during our lifetime, requiring a more or less complex, extended, and permanent web of activities starting with production, through storage, transportation, delivery, and transformation, before foods are consumed.

In addition to being an obligatory activity, eating is also a pleasure, especially as we distance from survival alone. As stated almost two centuries ago by Brillat-Savarin in the seventh introductory aphorism to his *magnus opus*, "*the pleasure of the table is enjoyed at all ages, conditions, countries and days, can be associated with all other pleasures and is the last we enjoy, giving us solace for the loss of all others*" [1] (p. ix).

However, cultural habits and preferences create food tastes and choices as well as inhibitions and aversions, complicating the introduction of new foods and composite dishes while simultaneously, reinforcing the use of those already established [2] (p. 481). A good example of the former are the words of caution and simultaneously of seduction that an enthusiastic convert to a new cuisine offer to less knowledgeable and less fortunate [3] (pp. 16–18). An example of the latter is the persistence of recipes transmitted by word-of-mouth that are, by example and imitation, at the core of the so-called traditional cuisine. In this data set, we report the less-well-known traditional cuisine of Portugal.

Portugal is a relatively small country, located in the western-most part of the Iberian Peninsula which comprises also two Atlantic archipelagoes, Madeira and the Azores islands. In all, it has slightly more than 92,000 km^2 (89,000 km^2 in mainland with maximum N–S length of 577 km and W–E length of 286 km, plus about 801 km^2 in Madeira and 2322 km^2 in the Azores). According to the 2011 census, the resident population is about 10,000,000 inhabitants in the mainland plus about 500,000 in the two islands, slightly more in Madeira than in Azores [4]. However, in a number of aspects, Portugal has little internal coherence and homogeneity for its size; there is a possible division of mainland

Portugal (the islands being separate cases) between the more mountainous North and the flatter South. Alternately, between the Atlantic Northwest and the more continental Northeast, both contrasting with the Mediterranean-type South, with Atlantic and Mediterranean inroads in the South and North, respectively. Nevertheless, these climatic and orographic contrasts have been counteracted by long, intense, and homogenizing human activities [5] (pp. 144–167).

This heterogeneity was noted with admiration nearly two centuries ago, especially in relation to the climate and vegetation [6] (pp. 89–99, 140–143). Later, it was stated without exaggeration that "*few countries of its size present such high level of natural differences, so that someone instantly transported from the middle of Minho* (the northern half of Entre Douro e Minho; Figure 1) *to the middle of Alentejo would think to have travelled not about 80 leagues* (around 325 km) *that separates them but thousands of them*" [7] (p. xi). For a short but insightful presentation of the strong territorial heterogeneity, see also [8] (pp. 1–4).

Figure 1. Approximate limits of the Portuguese division in provinces adopted. Insets are of the Atlantic islands of Madeira and Azores.

Altogether, this patchwork of climates and landscapes explains the various attempts to define and delimit provinces as administrative surrogates of the so-called natural regions of Portugal. The last such attempt was in the 1930s. In 1933's Constitution, provinces were considered the upper level in the territorial and administrative organization of Portugal, their exact delimitation being done only in 1936 [9], largely based upon Girão's proposals [10]. However, only eleven provinces were considered in the mainland, instead of the proposed thirteen. In 1959, provinces were lawfully extinguished. Nevertheless, they have been recognized and used in everyday life until today and are—some more than others—a very important part of the territorial identity of Portuguese people.

We hypothesized that provinces—being climatic and environmentally natural regions that became identity-generating entities—would somehow also correspond to separate and recognizable sub-cuisines with identifiable and segregating traits. For example, in the diachronic use of herbs, spices, and other condiments; in the way alien ingredients disseminated throughout Portuguese cuisine or, in a more encompassing framework, the reproducibility at the small but highly variable scale of the so-called "Darwinian gastronomy" hypothesis put forward to explain the worldwide pattern of the

use of spices [11,12]. As a prerequisite to investigate these hypotheses, we set out to create a data set of recipes described by their ingredients and the province where they were traditionally cooked and eaten. For strictly practical reasons, derived from the provincial organization adopted in the cookery books, we further reduced the number of provinces, merging Minho and Douro Litoral under the more ancient designation of Entre Douro e Minho. Alto and Baixo Alentejo were also merged under the umbrella designation of Alentejo (Figure 1).

2. Data Description

Sources for the data set were recipe books explicitly dealing with or aimed at presenting Portuguese traditional cooking. To be included in the data set, the origin of recipes had to be known. Therefore, to be selected and abstracted, cookery books had to allow (1) all or at least most of the recipes to be assigned to any one of the provinces shown in Figure 1, and (2) no province was omitted, deliberately or otherwise. As a consequence, some published materials could not be used, namely the classic Olleboma's book first published in 1936 [13] or the near encyclopedic series of Portuguese cookery books authored or co-authored by A. Saramago, left incomplete because of his death in 2008 when only six provinces had been published [14–19].

Thus, six cookery books were used [20–26]. Dates of first publication ranged from 1981 to 2006, but in only two [25,26] was the first edition after 2000. Data were derived from the cookery books by one researcher and independently double-checked by another, thus ensuring that all decisions were reached by consensus. The greatest care was taken in building this inventory, but we cannot reject that errors might have gone unnoticed. If errors are found or suspected, please let us know so they can be corrected.

Data were organized in tabular form in an MS Excel® 2010 spreadsheet, which is available in the Supplementary Materials (Supplementary S1) as a non-proprietary comma-separated value (CSV) format file. Each line in the data set presents information for a single recipe in columnar form. The first column, headed "CODE", displays alphanumeric codes for recipes and is composed of a four-letter acronym of the province name and a three-digit identifier for the recipe within the province (AÇOR stands for Azores, ALEN for Alentejo, ALGA for Algarve, BALT for Beira Alta, BBAI for Beira Baixa, BLIT for Beira Litoral, EDMI for Entre Douro e Minho, ESTR for Estremadura, MADE for Madeira, RIBA for Ribatejo, and TMAD for Trás-os-Montes e Alto Douro). Recipes are displayed in alphabetical order by province and alphabetic order by recipe name in Portuguese, and if necessary by ascending order of number of ingredients. Precise location of the recipes is not shown because such information is mostly absent. Undetermined origin ranged from 32% in Ribatejo (second minimum of 38% in Beira Baixa) to 93% in Azores and Madeira (second maximum of 68% in Alentejo), with an average of 63% when the recipes were pooled together.

The second column, "RECIPE (Portuguese)", presents the name of the recipe in Portuguese, and the third, "RECIPE (English translation)", the translation of the name in English. Translation essentially followed the English edition of Modesto's cookery book published in 1989 [20,27]. The fourth and fifth columns ("REF" and "PAGE") provide reference numbers and pages where the full recipe can be found. Whenever the recipe was abstracted from Modesto's cookery book [20], two page numbers are displayed in "PAGE" separated by a slash. The first number refers to the Portuguese edition and the second to the English translation [27], unless the recipe occurs on the same page in the two editions. The full list of references used is presented at the end.

The remaining 421 columns display individual ingredients in the recipes in alphabetical order by Portuguese names (1 if present in the recipe, blank if absent). Each column is headed by an alphanumeric code composed of the letter "I" (for ingredient) and a three-digit identifier for the ingredient, followed by its name in Portuguese and its translation to English in parentheses, and whenever it existed by the EFSA FoodEx2 food code [28].

Translations of ingredient names to English relied heavily on the English edition of Modesto's cookery book [27], but other useful sources were also used [3,29,30]. Finally, some of the English

translations were reviewed by colleagues with in-depth expertise in specific terminologies, namely alcoholic beverages [31]; seafood and fresh water foods, including fish, shellfish, and mollusks [32]; and game, livestock, and poultry, including their parts [33]. For ease of use, we also provide an English–Portuguese glossary of ingredient names in the Supplementary Materials (Supplementary S2). It goes without saying that responsibility for any errors in the data set and the glossary are ours.

Despite our efforts, translation into English was not possible for a number of recipes, and especially certain ingredients, largely because such ingredients are not used in English-speaking culinary areas. In all, seventeen ingredients could not be translated, most being pork-based dry-cured sausages (53%) and wines (24%).

3. Methods

3.1. Ingredients

Salt was not included in the data set as an ingredient because it was always included in the recipe or, in relatively few cases where no mention of it was made, its presence is implicit through one or more ingredients, usually *bacalhau* (Atlantic cod, salted and dried; I035). In Portuguese cuisine, *bacalhau* is always the dry, highly-salted form that almost always has to be soaked in water for one or more days, the water being changed a couple of times so that most of the salt is removed. It is also worth remarking that *coentro* (coriander; I121) is widely consumed throughout the world in a variety of forms (mostly seeds) [34] but in Portuguese cuisine, only the leaves or sprigs composed of stem and leaves (preferably fresh) are used.

Decisions had to be taken in relation to ingredients that are mixtures of ingredients, which it is assumed everyone knows how to prepare without further instructions, specifically side dishes. In these cases, we considered a minimum set of components and included them in the recipe. Thus, potato puree was described as butter, milk, and potato (e.g., in *Cavalas recheadas*, Stuffed mackerel, AÇOR024 or *Coelho à caçadora*, Jugged rabbit, ESTR071), regardless of other ingredients that might be used, depending on the cook or preferences (e.g., white pepper or nutmeg). *Esparregado*, etymologically related to asparagus but usually done using spinach or other vegetables, traditionally contains bayleaf, flour, garlic, and olive oil (e.g., in *Galinholas à alentejana*, Alentejo style woodcocks, ALEN104). Also, *arroz branco* (white rice) is assumed to require only butter and rice (e.g., in *Lampreia*, Lamprey, BBAI045 or *Tripas à moda do Porto*, Oporto style tripes, EDMI175); ingredients adopted for white rice agreed with the description in one recipe from Madeira (*Atum assado*, Grilled tuna, MADE005). Finally, fried potatoes (chips, French fries) were listed as potatoes and oil for frying.

Conversely, dishes not prepared "in the moment" were taken as a single ingredient, regardless of their composition (e.g., *enchidos*, dry-cured sausages, or *molho inglês*, Worcester sauce, I252).

In a few cases, linguistic differences in Portuguese had to be accounted for, the most important involving *segurelha* (savory, I384), which is an ingredient in *Feijão verde à alentejana* (Alentejo style string beans, ALEN097). The name *segurelha* is also used in the Madeira Islands, except that in Madeira it refers to thyme [20] (p. 296), [27] (p. 314), [35] (p. 283). Another case was *pimenta da Jamaica* (allspice, I018), which is only used in recipes from the Azores and Madeira. In the latter, it is almost never referred to as *pimenta da Jamaica* but as *alcepás* (more rarely as *acepás*), clearly a corruption of its name in English.

In Portuguese, true saffron (*Crocus sativus* L.) is *açafrão*, but this term is also applied to the much less expensive turmeric (*Curcuma longa* L.) and safflower (*Carthamus tinctorius* L.). In addition to *açafrão*, the names *açaflor* and *açafroa* are also mentioned, the latter representing—in Portuguese—the grammatical feminine of *açafrão*. Throughout the data set, *açaflor* or *açafroa* are referred to as ingredients twice, in AÇOR019 and AÇOR038, respectively, while *açafrão* is referred to 26 times. In two of them, it is explicitly stated that true saffron should be used (EDMI009 and EDMI064). In the other 24, no clarification is made and we assumed it meant turmeric, less frequently safflower, but never saffron.

Throughout the cookery books, there is reference to several peppers, namely *pimenta preta* (black pepper, I314), *pimenta rosa* (rose pepper, I315), *pimenta verde* (green pepper, I316), *pimenta vermelha* (red pepper, I317), and *pimenta da Jamaica*, already mentioned. Singly or jointly, they are present in 54 recipes. Additionally, in about 780 recipes, there is mention of *pimenta* without any other quality attached; we assumed this always meant white pepper (I313).

3.2. Recipes

Recipes in the data set include starters, main courses, side dishes, and soups. Desserts, cakes, sweets, puddings, and pastries were not considered. Recipes are described by the ingredients they require in terms of presence or absence, completely disregarding quantities because presence might be expected to be a relatively stable characteristic—at least more stable than quantities, which strongly depend either on the availability of ingredients or on the taste and likings of those for whom meals are prepared. On the importance of availability and taste in this framework, see for example [36] (pp. 48–49).

Traditional cuisine is supposed, at the very least, to involve a large number of recipes eaten in individual households where if someone will dislike a given ingredient, vinegar for example, it is reduced or eliminated; the opposite also being true where an ingredient is favored. Obviously, one ingredient may be replaced with another (e.g., vinegar by lemon), which is taken to be more or less equivalent. We have seen this happen with, for example, turnips being replaced by potatoes in *Coelho à Capitão-Mor* (Rabbit "à Capitão-Mor", BALT036). This may help explain the frequency of instructions like "*use this or that ingredient*" or "*such and such ingredient can also be used*". An example of the former can be found in *Caldo-verde à Minhota* (Minho Style shredded cabbage soup, EDMI077) which indicates the use of *salpicão* (I359) or *chouriço* (I116). Of the latter, in *Cachola de porco* (Pork "cachola", ESTR046), it is stated that *água-pé* (I102) can be replaced by ordinary white wine although *Cachola* apparently "*tastes infinitely better with água-pé*" [20] (p. 209), [27] (p. 222). Whenever such instructions appeared, we retained only the first ingredient (i.e., we only listed *salpicão*) and disregarded ingredients that "*could also be used*" (i.e., we only listed *água-pé*), except when these were explicitly considered as a separate variant of the recipe.

When we finished abstracting the six cookery books [20–26], the data set totaled 1644 recipes and a precautionary check on repeated recipes was performed. Pairwise comparisons between each recipe and all others was done, and a likeness-value (*LI*) calculated as:

$$LI = N_{A,B}/\max(N_A, N_B), \tag{1}$$

where $N_{A,B}$ is the number of ingredients present simultaneously in recipes A and B, and $\max(N_A, N_B)$ is the number of ingredients in the recipe with the greater number (or the number of ingredients of recipe A or B if $N_A = N_B$), with *LI* ranging from 0 to 1. When *LI* = 0, there were no common ingredients between recipes. When *LI* = 1, recipes have the same number of ingredients and all ingredients of one recipe occur in the other. *LI* = 0.5 was obtained whenever the number of ingredients common to two recipes was half the maximum number of ingredients of the richest recipe. However, recipes could have *LI* = 1 and still be different, because of the way the dish is composed and cooked. Therefore, we checked ways of cooking for every pair of recipes with *LI* = 1; the result being that 262 where true repetitions, for the most part recipes from Alentejo (40), Algarve (34), Trás-os-Montes e Alto Douro (30), and Entre Douro e Minho (28). In addition, 19 recipes were found to have *LI* = 1 when compared with another, but were cooked in different ways, most involving recipes from Algarve (8) and Trás-os-Montes e Alto Douro (4). True repetitions were eliminated from the data set, which was reduced to 1382 recipes. Entries that were eliminated belonged to later publications.

3.3. Evaluating Bias in Cookery Books Examined

As explained above, the data set was based on six cookery books by different authors, which had different aims and methods of collecting and selecting recipes. Therefore, it is conceivable that authors—and thus cookery books and recipes—might be biased. The impact of putative biases is naturally dependent on the share each cookery book made to the total number of recipes.

Three books [20,24,25] out of six provided 85% of recipes, and always more than 80% of recipes in each province, with the exception of the Azores, and were selected to investigate for bias in their portrait of Portuguese traditional cuisine. The underlying rationale for the method used was that if bias exists, then from the occurrence of one or a combination of ingredients in a recipe it would be possible to predict the book from where that recipe came. We adopted a non-parametric tree-structured classification method [37,38], which is preferable to other methods such as discriminant analysis or logistic regression because of its robustness with respect to outliers and misclassified points, its applicability to any data structure, and its automatic stepwise selection and complexity reduction. In addition, it easily accommodated interactions between variables without prior selection of variables and gave estimates for incorrect identifications. Finally, it results in decision rules which are easy to understand and apply [37] (pp. 56–58).

Binary trees were generated using SPAD Data Mining & Text Mining, v. 6.5.0 (SPAD, Paris, France). Only recipes from the three books [20,24,25] were used. The relative cost of misidentification was constant and unitary, and 25 independent runs were done, randomly assigning 66% of recipes to the learning group and 33% to the validation group. Overall, misidentification of cookery books in the validation group was the major criterion for selection.

Optimal partitions were always obtained, and 21 out of 25 independent runs resulted in the same binary tree with only two ingredients: white pepper and *manteiga* (butter, I231). The same ingredients plus *safio/congro* (conger, I355) and *vinho branco* (white wine, I412) were present in two additional trees. The same four ingredients plus *azeite* (olive oil, I032) were present in two additional trees. Misidentification by the shorter tree ranged from 43.7% to 44.7% in the validation group, with recipes from Modesto [20] being generally better identified (between 93.0% and 94.1% of correct identification). Recipes from Guedes [24] were generally misidentified (31.2–36.0% of correct identification), while recipes from Valente [25] were always misidentified. Because longer trees did not noticeably reduce the percentage of misidentification, we kept the shorter and most frequent binary tree. Recipes were identified as belonging to Modesto [20] either when they lacked white pepper or when they had white pepper but lacked butter. Recipes were identified as belonging to Guedes [24] only when they had white pepper and butter. Recipes were never identified as belonging to Valente [25]. Therefore, some minor bias around white pepper and butter might be present, and it may be wise to exert some caution as to whether or not the conclusions of analyses appear to be dependent on these ingredients. Possible courses of action might be to perform analyses with and without white pepper and butter, or the two ingredients included as supplementary to evaluate the stability of conclusions.

3.4. Exploratory Characterization of Recipes and Ingredients

All statistics in this section were done with Statgraphics 4.2 (STSC, Inc., Rockville, MD, USA) or with MS Excel® 2010. We adopted a type I error rate $\alpha = 0.001$ as a threshold reference for strong evidence against the null hypothesis [39]. In general, data are presented as mean $\pm$ standard error (SE) and sample size (*n*).

3.4.1. Recipes

Alentejo clearly topped the rank for numbers of recipes with 209, followed by Trás-os-Montes e Alto Douro, Estremadura, and Entre Douro e Minho with 186, 182, and 180 recipes, respectively. These provinces were well above the average number of recipes, which was approximately 126. Additionally, with more than the average number of recipes was Algarve, with 147 recipes. Conversely, the other six

provinces (Azores, the three Beiras, Madeira, and Ribatejo) had totals well below the average, ranging from 70 to 89 (Figure 2).
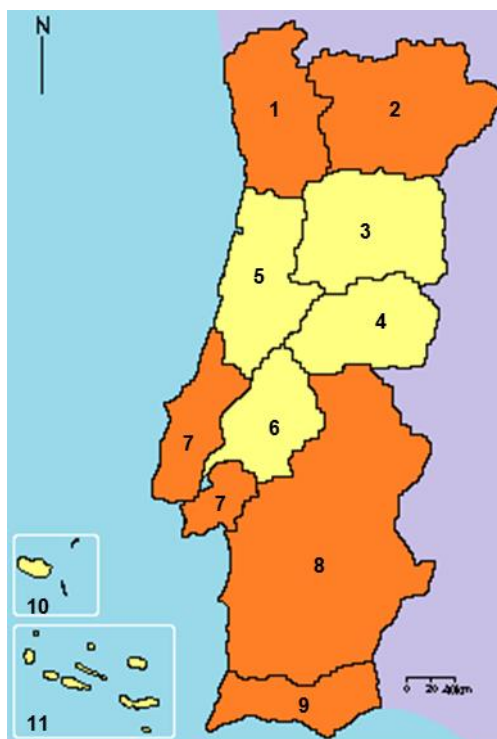

Figure 2. Over-representation (orange) and under-representation (light yellow) of number of recipes in provinces of Portugal. 1 Entre Douro e Minho, 2 Trás-os-Montes e Alto Douro, 3 Beira Alta, 4 Beira Baixa, 5 Beira Litoral, 6 Ribatejo, 7 Estremadura, 8 Alentejo, 9 Algarve, 10 Madeira, 11 Azores.

Considering all 1382 recipes, the number of ingredients per recipe ranged from one in two recipes (*Sável fumado*, Smoked shad, EDMI158; *Cracas*, Barnacles, MADE047), the ingredient being allis shad or barnacles, to 23 also in two recipes (*Cabrito com arroz à moda de Monção*, Monção style goatling with rice, BBAI019, and *Feijoada à transmontana*, Trás-os-Montes Style bean stew, TMAD108). The mean number of ingredients was (mean $\pm$ SE and sample size n) 9.5 $\pm$ 0.1, n = 1382 (median of 9), which is larger than the mean number of ingredients per recipe (nine ingredients including salt) found in combined worldwide inventories of recipes [40]. The coefficient of skewness g_1 = 0.647 was highly significant (p = 7.9 $\times$ 10^{-23}, two-tailed test), meaning the distribution of numbers of ingredients was skewed to the right. When skewness was tested separately for each province, there was strong evidence of highly skewed distributions for the numbers of ingredients to the right only in the two northernmost provinces of Entre Douro e Minho and Trás-os-Montes e Alto Douro (Figure 1) which had g_1-values of 0.721 and 0.826, and p-values 6.9 $\times$ 10^{-5} and 3.5 $\times$ 10^{-6}, respectively.

The mean numbers of ingredients in each province was compared with the mean numbers of ingredients of all provinces to assess over-, average-, and under-representation. For this, we determined the probability of a *t*-Student value necessary for inclusion of the means for each province in the confidence interval of all 1382 Portuguese recipes. Significantly over-represented provinces (mean number of recipes from 10.0 $\pm$ 0.4, n = 70 to 10.4 $\pm$ 0.3, n = 182, one-tailed p-values less than 3.2 $\times$ 10^{-8}), by increasing order of means, were Madeira, Entre Douro e Minho, Azores, and Estremadura. Significantly under-represented provinces (mean number of recipes from 9.0 $\pm$ 0.4, n = 82 to 7.8 $\pm$ 0.4, n = 79, one-tailed p-values less than 2.1 $\times$ 10^{-8}), by decreasing order of means, were Ribatejo, Algarve, Beira Baixa, and Beira Alta. Average representation (mean numbers of recipes from 9.2 $\pm$ 0.2, n = 209 to

9.6 ± 0.3, *n* = 86, one-tailed *p*-values from 0.006 to 0.424) was found for the remaining three provinces, Alentejo, Trás-os-Montes e Alto Douro, and Beira Litoral (Figure 3).

Figure 3. Over-representation (orange), average representation (white), and under-representation (light yellow) of mean number of ingredients of recipes in provinces of Portugal. 1 Entre Douro e Minho, 2 Trás-os-Montes e Alto Douro, 3 Beira Alta, 4 Beira Baixa, 5 Beira Litoral, 6 Ribatejo, 7 Estremadura, 8 Alentejo, 9 Algarve, 10 Madeira, 11 Azores.

Comparisons like those described where parts (provinces) are compared with the whole to which they belong (country) are not completely independent, which might help explain the lack of clarity in patterns of differences among provinces. The only exception was the group of over-representation that included provinces that were either islands (Azores and Madeira) or included the largest and wealthiest cities of Portugal, Lisbon in Estremadura and Oporto in Entre Douro e Minho. An alternative approach might be to test the mean numbers of ingredients for all provinces. Simultaneous comparisons were done using a least squares linear regression approach with dummy variables to prevent the ambiguity resulting from lack of "transitivity", which frequently arises in simultaneous test procedures. For example, mean A is not significantly different from mean B, mean B is not significantly different from mean C either, but means A and C are significantly different [41,42].

Forward stepwise selection with replication was used, and the complete candidate models included only qualitative independent variables (the provinces), binary coded as 0, 1. An experiment-wise type I error rate was adopted for the coefficients of regression and calculated using the Dunn–Šidák method [43]. Coefficients of determination (R^2) are presented as proportions of the maximum R^2 possible [44] (p. 246). The significant heteroscedasticity detected in untransformed data using the two-tailed *F* distribution ($p = 5.1 \times 10^{-6}$) was strongly reduced ($p = 0.002$), which was deemed acceptable, when the numbers of ingredients were transformed using natural logarithms.

A significant three-term polynomial could be fitted to logarithmically-transformed data. Significance levels of the coefficients were $p \leq 0.020$; $R^2 = 0.673$; lack of fit $F_{8,1371} = 1.787$, corresponding to $p = 0.075$, and thereby, significant differences among provinces were found. After solving the equation for the values 0, 1 of the binary variables, provinces could be separated as three groups.

Group 1, was composed of recipes from Beira Alta only, comparatively poor in ingredients (7.8 ± 0.4, *n* = 79); group 2 was composed of recipes from Estremadura only, comparatively rich in ingredients (10.4 ± 0.3, *n* = 182); and group 3 was composed of recipes from the remaining nine provinces pooled together, with an intermediate number of ingredients per recipe (9.5 ± 0.1, *n* = 1121).

Complementary to the previous approach was the analysis of likeness-values (Equation (1)). Considering all 954,271 *LI*-values, the most notable feature was the low likeness of recipes. *LI*-values range from 0 to 1, with a mean of 0.218 ± 0.0001, *n* = 954,271, and a median of 0.2. Skewness $g_1 = 0.424$ was highly significant ($p \approx 0$, two-tailed test). A detailed analysis of *LI*-values is beyond the scope of this paper, but a few points are worthy of examination. Firstly, comparing mean *LI*-values within provinces (main diagonal elements in Table 1 below) with the mean *LI*-value of Portugal allowed the identification of three groups. The first, with under-represented *LI*-values, was composed of Beira Alta, Beira Baixa, and Trás-os-Montes e Alto Douro. The second, of near-average representation, was composed of the Azores. The third, over-representation, was composed of the remaining seven provinces (Figure 4a). *LI*-values within provinces might mirror the intensity of interactions and common use of recipes among inhabitants of each province. Alternatively, they might mirror the heterogeneity of environmental conditions, and thus of food supply (or both). Either way, increased interactions and common use, and decreased heterogeneity of the food supply, would likely increase *LI*-values. Therefore, frequency and distribution within and between provinces of complete unlikeness (*LI* = 0) is worth analyzing.
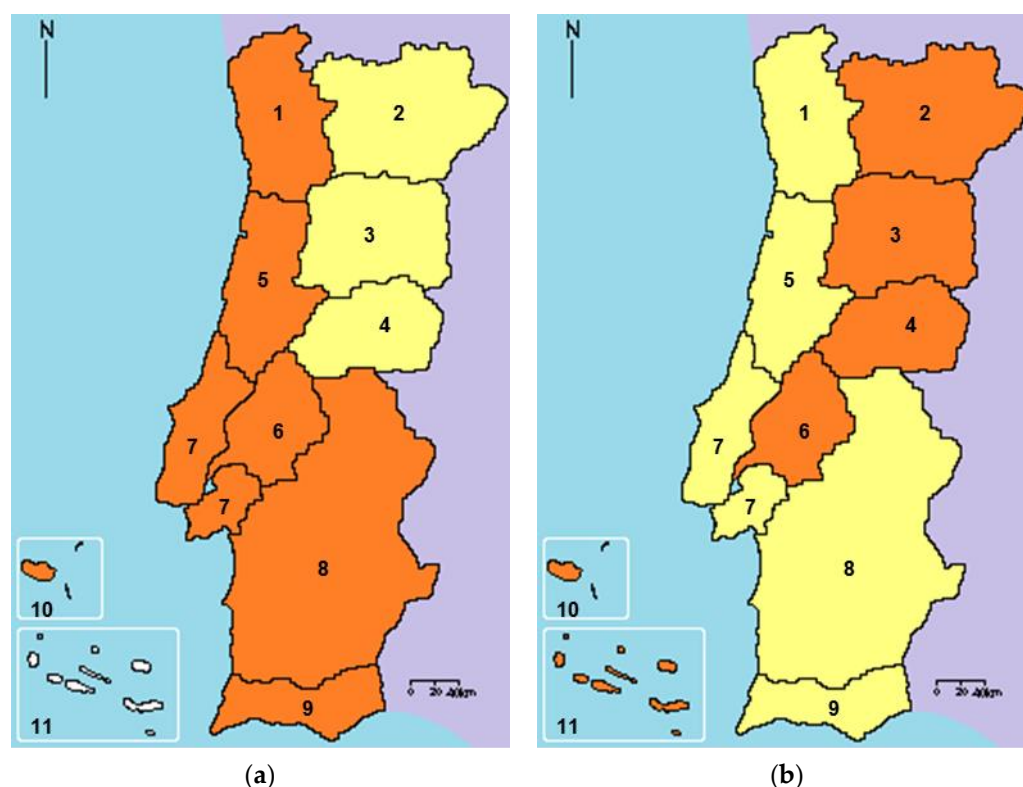

Figure 4. Over-representation (orange), average representation (white), and under-representation (light yellow) of (**a**) mean likeness-values within provinces of Portugal, (**b**) frequency of recipes of each province involved in comparisons with likeness-values *LI* = 0 weighted by the number of recipes of the province. 1 Entre Douro e Minho, 2 Trás-os-Montes e Alto Douro, 3 Beira Alta, 4 Beira Baixa, 5 Beira Litoral, 6 Ribatejo, 7 Estremadura, 8 Alentejo, 9 Algarve, 10 Madeira, 11 Azores.

Table 1. Means of likeness-values (*LI*) between recipes of a province and recipes of the same (main diagonal) or other province. Numbers in bold indicate that *LI*-values within a province were larger than mean *LI*-values between that province and all others.

	Alentejo	Algarve	Azores	Beira Alta	Beira Baixa	Beira Litoral	Entre Douro e Minho	Estremadura	Madeira	Ribatejo	Trás-os-Montes e Alto Douro
Alentejo	**0.258**										
Algarve	0.242	**0.264**									
Azores	0.212	0.218	**0.220**								
Beira Alta	0.200	0.211	0.183	0.194							
Beira Baixa	0.208	0.210	0.180	0.184	0.197						
Beira Litoral	0.245	0.253	0.217	0.220	0.212	**0.271**					
Entre Douro e Minho	0.226	0.242	0.208	0.213	0.208	0.250	0.250				
Estremadura	0.223	0.244	0.207	0.200	0.200	0.242	0.234	0.243			
Madeira	0.207	0.226	0.206	0.187	0.188	0.220	0.213	0.219	**0.248**		
Ribatejo	0.229	0.234	0.200	0.197	0.198	0.239	0.227	0.221	0.204	0.229	
Trás-os-Montes e Alto Douro	0.200	0.210	0.183	0.191	0.186	0.212	0.216	0.200	0.186	0.195	0.200

About 12% of *LI*-values were null. However, their impact on the mean *LI*-value of Portugal was negligible, because when null *LI*-values were removed, the mean increased slightly from 0.218 ± 0.0001, $n = 954{,}271$ to 0.248 ± 0.0001, $n = 841{,}726$, the median increased from 0.200 to 0.231, and skewness $g_1 = 0.703$ remained highly significant ($p \approx 0$, two-tailed test). Because it is likely that provinces with more recipes also had more recipes involved in the comparisons, all frequencies were weighted against numbers of recipes. Considering all provinces pooled together, the frequency of comparisons with $LI = 0$ weighted by numbers of recipes involved was 162.9. Using this value as the reference to compare each province, two groups were identified. The group of over-representation involved in comparisons having $LI = 0$ comprised Azores, Beira Alta, Beira Baixa, Madeira, Ribatejo, and Trás-os-Montes e Alto Douro (weighted frequencies from 173.5 to 217.8). The group of under-representation comprised Alentejo, Algarve, Beira Litoral, Entre Douro e Minho, and Estremadura (weighted frequencies from 129.9 to 155.0), which is almost the opposite of the distribution examined above (Figure 4b) with Madeira and Ribatejo moving between groups.

Despite the need for further investigation, our analyses suggests that *LI*-values might relate to characteristics of provinces, regardless of what these may be, but such characteristics support a hypothesis of individuality. This would imply that the more "unique" a province is (cookery-wise), the larger the likeness will be among its recipes, and the smaller these values will be when compared with recipes from other provinces. Data summarized in Table 1 provide the first insight into this hypothesis.

Alentejo, Algarve, Azores, Beira Litoral, and Madeira seemed to fit in a group of "unique" and highly distinctive provinces. Thus, besides Beira Litoral (which ranks second in relation to Alentejo and Algarve), this group includes the Atlantic islands and Southern Portugal. Entre Douro e Minho and Estremadura seem to be as alike as they are alike to Beira Litoral and Algarve, respectively. Conversely, the remaining provinces are more alike other provinces than among themselves. Beira Alta, Beira Baixa, and Ribatejo are more alike with Beira Litoral (an apparent attractor); Trás-os-Montes e Alto Douro is more alike to neighboring Entre Douro e Minho followed by Beira Litoral.

3.4.2. Use of Ingredients

As seen in Section 2, 421 different ingredients were found in 1382 recipes. The top three ingredients (in frequency but not necessarily in amounts consumed) were olive oil, *cebola* (onion, I105) and *alho* (garlic, I022), which were present in 66%, 64%, and 59% of recipes, respectively. The top five also included white pepper and *salsa* (parsley, I360), present in 56% and 41% of recipes. When provinces were considered, the five top ingredients in Algarve, Beira Alta, Beira Litoral, Entre Douro e Minho, and Estremadura were exactly the same, even if their rank might vary. Conversely, in Azores, Beira Alta, Madeira, Ribatejo, and Trás-os-Montes e Alto Douro, parsley was always replaced by *batata* (potato, I046) in the top five ingredients, while in Alentejo, parsley was replaced by *pão* (bread, I277), which in Portugal is made almost exclusively of wheat flour. *Broa/pão de milho* (corn bread, I060) and *pão de centeio* (rye bread, I278) were recorded as ingredients in far fewer recipes, localized to Beira Litoral and Entre Douro e Minho (corn bread) and Trás-os-Montes e Alto Douro (rye bread). However, the comparison of individual ingredients can be misleading because the same basic sources of food can be used in different ways. For example, pork appears in the data set in 66 different ways, including as *enchidos*, suckling pig and fats. Various fish, including shellfish and mollusks but excluding Atlantic cod, salted and dried, appeared in 85 different ways; beef appeared in 36 ways; lamb in 24; chicken and goat in 14 ways; hare in four; turkey in three; and duck and rabbit in two different ways each. In addition, wines appear in nine different ways and peppers appear in six. Therefore, we combined data from these ingredients and recalculated the frequencies.

As before, olive oil and onion were the most frequently used, but garlic was replaced by peppers, present in 61% of recipes. In fourth place was garlic, and the top five was completed by pork, present in one way or another in more than half (53%) the recipes, replacing parsley. However, when provinces were examined, a more fragmented pattern emerged; now, it was Beira Litoral, Entre Douro e Minho,

Ribatejo, and Trás-os-Montes e Alto Douro where the five top ingredients were the same, even if the rank of the ingredients might vary. Conversely, in Alentejo and Beira Baixa, peppers were replaced by bread and parsley, respectively; in Algarve and Estremadura, pork by fish and parsley; in the Azores, olive oil by wines; in Beira Alta, onion by potato; and in Madeira, garlic and pork by potato and fish.

Incidentally, Atlantic cod, salted and dried, appeared in only 129 recipes, ranking 23rd (in provinces 13th to 38th, Ribatejo and Azores, respectively)—well under the 1001 different ways to cook Atlantic cod claimed by the Portuguese [3] (p. 24) or the more modest 365 ways, one for each day of the year [29] (p. 37), [45].

3.5. Final Remarks

Clearly, Portuguese traditional recipes organize differently across provinces. In the exploratory characterization, Entre Douro e Minho and Estremadura always occurred together, as did Beira Alta and Beira Baixa. Alentejo and Algarve almost always appeared together with the first group, while Ribatejo and Azores occurred together with the second of hinterland Beiras. The remaining two, Trás-os-Montes e Alto Douro and Beira Litoral, appeared with Madeira. It is also worth noting that Beira Litoral had an "attractor" role in terms of likeness-values for ingredients. Detailed analyses could confirm or inform the tentative grouping presented, providing a clearer picture of Portuguese traditional cuisine and hopefully pointing to possible explanations.

The data set presented here represents a comprehensive review that brings together an otherwise dispersed body of published knowledge on Portuguese traditional recipes. As such, it may constitute a reference inventory for investigating and testing a number of hypotheses on the distribution and pattern of use of ingredients within and among natural regions of Portugal. For example, how do provinces group in relation to the presence of different types of ingredients, namely herbs and spices, or vegetables, meats, fishes, and shellfish? Further, what might be the relation, if any, of the groups formed and factors like climatic, demographic, economic, and historic characteristics or events?

The data set might also be useful to investigate the extent and reality of tradition and traditional use attributed to the recipes. When building the data set, and in this paper, we accepted as good the classification of recipes as traditional without attempting to define or characterize this term. For the moment, we have to assume that the recipes fall somewhere between two opposing poles, where one is truly traditional (i.e., part of an aggregate of customs and practices that give continuity to a culture and to a social group [46] (p. 84)), and the other comprised of "invented traditions", more or less loosely based on past customs that once had the capacity of change and evolution, but are now frozen in various ideals or ritualized ways [47].

References

1.   Brillat-Savarin, J.-A. *Physiologie du Gout, ou Méditations de Gastronomie Transcendante; Ouvrage Théorique, Historique et a l'Ordre du Jour, Dédié aux Gastronomes Parisiens, par un Professeur, Membre de Plusieurs Sociétés*

Littéraires et Savantes; A. Sautelet et Cie Libraires: Paris, France, 1826; Volume 1, p. 390. Available online: http://catalogue.bnf.fr/ark:/12148/cb30161657n.public (accessed on 23 October 2017).

2. Helen Hemingway. Macropedia. In *The New Enciclopædia Britannica*, 15th ed.; Helen Hemingway: Chicago, IL, USA, 1974; Volume 7, p. 1135.
3. Sarvis, S. *A Taste of Portugal*; Charles Scribner's Sons: New York, NY, USA, 1967; p. 192.
4. Instituto Nacional de Estatística. Statistics Portugal. Available online: https://www.ine.pt/xportal/xmain?xpgid=ine_main&xpid=INE (accessed on 15 December 2017).
5. Ribeiro, O. *Portugal. O Mediterrâneo e o Atlântico*, 5th ed.; Revised and Augmented; Sá da Costa: Lisboa, Portugal, 1987; p. 189.
6. Balbi, A. *Essai Statistique sur le Royaume de Portugal et d'Algarve, Comparé aux Autres États de l'Europe, et Suivi d'un Coup d'Oeil sur l'État Actuel des Sciences, des Lettres et des Beaux-Arts Parmi les Portugais des Deux Hémisphères*; Rey et Gravier, Libraires: Paris, France, 1822; Volume 1, p. 481.
7. Costa, B.C.C.; Castro, L. *Le Portugal du Point de Vue Agricole*; Imprimerie Nationale: Lisbonne, Portugal, 1900; p. 965. Available online: https://ia601700.us.archive.org/20/items/leportugalaupoin00cost/leportugalaupoin00cost.pdf (accessed on 4 December 2017).
8. Disney, A.R. *A History of Portugal and the Portuguese Empire*; Cambridge University Press: Cambridge, UK, 2009; Volume 1, p. 386.
9. *Código Administrativo (Aprovado pelo Decreto-Lei n.° 27-424 de 31 de Dezembro de 1936) Contendo as Leis n.os 1:940 de 3 de Abril de 1936, e 1:945 e 1:1946 de 21 de Dezembro de 1936 que Promulgam as Bases da Organização Administrativa e Dão Nova Redacção a Alguns Artigos da Constituição, e Inserindo Todas as Tabelas e Mapas Anexos ao Código*; Emprêsa Jurídica Editora: Coimbra, Portugal, 1937; p. 248.
10. Girão, A.A. *Esbôço Duma Carta Regional de Portugal*, 2nd ed.; Rewritten and Augmented; Imprensa da Universidade: Coimbra, Portugal, 1933; p. 224.
11. Billing, J.; Sherman, P.W. Antimicrobial functions of spices: Why some like it hot. *Q. Rev. Biol.* **1998**, *73*, 3–49. [CrossRef] [PubMed]
12. Sherman, P.W.; Billing, J. Darwinian gastronomy: Why we use spices. *BioScience* **1999**, *49*, 453–463. [CrossRef]
13. Bello, A.M.O. *Culinária Portuguesa*; Assírio & Alvim: Lisboa, Portugal, 1994; p. 454. ISBN 972-37-0370-X.
14. Saramago, A.; Fialho, M. *Cozinha Alentejana*; Assírio & Alvim: Lisboa, Portugal, 1998; p. 271. ISBN 972-37-0494-3.
15. Saramago, A. *Cozinha Transmontana: Enquadramento Histórico e Receitas*; Assírio & Alvim: Lisboa, Portugal, 1999; p. 275.
16. Saramago, A. *Cozinha do Minho: Enquadramento Histórico e Receitas*; Assírio & Alvim: Lisboa, Portugal, 2000; p. 254.
17. Saramago, A. *Cozinha Algarvia: Enquadramento Histórico e Receitas*; Assírio & Alvim: Lisboa, Portugal, 2001; p. 228.
18. Saramago, A. *Cozinha da Beira Interior: Enquadramento Histórico e Receitas*; Assírio & Alvim: Lisboa, Portugal, 2002; p. 254.
19. Saramago, A. *Cozinha da Beira Litoral: Enquadramento Histórico e Receitas*; Assírio & Alvim: Lisboa, Portugal, 2003; p. 238.
20. Modesto, M.L. *Cozinha Tradicional Portuguesa*, 2nd ed.; Editorial Verbo: Lisboa, Portugal, 1982; p. 335.
21. Modesto, M.L.; Praça, A. *Festas e Comeres do Povo Português*, 2nd ed.; Editorial Verbo: Lisboa, Portugal, 2000; Volume 1, p. 214.
22. Modesto, M.L.; Praça, A. *Festas e Comeres do Povo Português*; Editorial Verbo: Lisboa, Portugal, 1999; Volume 2, p. 218.
23. Bourguette, V.; Mourato, D. *Gastronomia Regional Portuguesa*, 5th ed.; Impala Editores: Ranholas, Portugal, 2005; p. 240.
24. Guedes, F. *Receitas Portuguesas. Os Pratos Típicos das Regiões*, 3rd ed.; D. Quixote: Lisboa, Portugal, 2005; p. 301.
25. Valente, M.O.C. *O Melhor da Cozinha Regional Portuguesa*; Círculo de Leitores: Rio de Mouro, 2004; p. 248.
26. Ribeiro, M. *Do Minho ao Algarve. Açores e Madeira*; Planeta Editora: Lisboa, Portugal, 2006; p. 370.
27. Modesto, M.L. *Traditional Portuguese Cooking*, 6th ed.; Verbo: Lisboa, Portugal, 2014; p. 359.
28. European Food Safety Authority. Foodex2 Browsing Tool (Updated: 18 August 2017). Available online: https://www.efsa.europa.eu/en/data/data-standardisation (accessed on 21 January 2018).

29.　Anderson, J. *The Food of Portugal*; William Morrow and Company: New York, NY, USA, 1986; p. 304.

30.　Carter-Tripp, M. Guide to Food in Portugal. Available online: http://www.xacobeo.fr/ZF1.07.lex.en-pt_food-term.pdf (accessed on 24 June 2016).

31.　Cabrita, M.J.; University of Évora, Évora, Portugal. Personal communication, 2017.

32.　Almeida, P.R.; University of Évora, Évora, Portugal. Personal communication, 2017.

33.　Elias, M.; University of Évora, Évora, Portugal. Personal communication, 2017.

34.　Diederichsen, A. *Coriander (Coriandrum sativum L.)*; IPK; IPGRI: Rome, Italy; Gatersleben, Germany, 1996; p. 83.

35.　Silva, F.A.; Menezes, C.A. *Elucidário Madeirense. Volume Terceiro O–Z*, 3rd ed.; Regional Secretariat for Tourism and Culture, Regional Directorate for Cultural Affairs: Funchal, Portugal, 1966; p. 413.

36.　Turmo, I.G. *200 Años de Cocina. Historia y Antropología de la Alimentación*; Cultiva Libros: Madrid, Spain, 2013; p. 375.

37.　Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; CRC Press: Boca Raton, FL, USA, 1984; p. 358.

38.　Guegen, A.; Nakache, J.P. Méthode de discrimination basée sur la construction d'un arbre de decision binaire. *Rev. Stat. Appl.* **1998**, *36*, 19–38. Available online: http://www.numdam.org/item?id=RSA_1988__36_1_19_0 (accessed on 17 December 2017).

39.　Sterne, J.A.C.; Smith, J.D. Sifting the evidence—What's wrong with significance tests? *Br. Med. J.* **2001**, *322*, 226–231. [CrossRef]

40.　Chatterjee, U.; Kumar, V.; Madalli, D.P. Formalizing food ingredients for data analysis and knowledge organization. *Collnet. J. Scientometr. Inf. Manag.* **2016**, *10*, 289–309. [CrossRef]

41.　Chew, V. Comparing treatment means: A compendium. *HortScience* **1976**, *11*, 348–357.

42.　Penas, A.C.; Dias, L.S.; Mota, M.M. Precision and selection of extraction methods of aphelenchid nematodes from maritime pinewood, *Pinus pinaster* L. *J. Nematol.* **2002**, *34*, 62–65. Available online: http://journals.fcla.edu/jon/article/view/69396 (accessed on 12 December 2017). [PubMed]

43.　Ury, H.K. A comparison of four procedures for multiple comparisons among means (pairwise contrasts) for arbitrary sample sizes. *Technometrics* **1976**, *18*, 89–97. Available online: http://amstat.tandfonline.com/doi/abs/10.1080/00401706.1976.10489405#.WjBXJKNLHiy (accessed on 12 December 2017). [CrossRef]

44.　Draper, N.R.; Smith, H. *Applied Regression Analysis*, 3rd ed.; John Wiley: New York, NY, USA, 1998; p. 706.

45.　Costa, A.P. Portugal. A dialogue of cultures. In *Culinary Cultures of Europe. Identity, Diversity and Dialogue*; Goldstein, D., Merkle, K., Eds.; Council of Europe Publishing: Strasbourg, France, 2005; pp. 347–356.

46.　Helen Hemingway. Micropedia. In *The New Enciclopædia Britannica*, 15th ed.; Helen Hemingway: Chicago, IL, USA, 1974; Volume X, p. 1031.

47.　Hobsbawm, E. Introduction: Inventing traditions. In *The Invention of Tradition*; Hobsbawm, E., Ranger, T., Eds.; Cambridge University Press: Cambridge, UK, 1983; pp. 1–14.