5
Analysis of Food Pairing in Regional Cuisines of India

5.1 INDIAN REGIONAL CUISINES

Details of recipes, ingredients, and their corresponding flavor compounds constitute the primary data required for the study of food pairing in a cuisine. Much of this is documented in the form of books and recently through online recipe sources. We obtained the Indian cuisine recipes data from one of the popular cookery websites *TarlaDalal.com* [Dalal, 2014]. The flavor profiles of ingredients were compiled using previously published data [Ahn et al., 2011] and through an extensive literature survey. Table 5.1 lists details of recipes and ingredients in each of the regional cuisines.

Table 5.1: Statistics of regional cuisines

Cuisine	Recipe count*	Ingredient count
Bengali	156	102
Gujarati	392	112
Jain	447	138
Maharashtrian	130	93
Mughlai	179	105
Punjabi	1013	152
Rajasthani	126	78
South Indian	474	114

* Recipes of size ≥2 were considered for the purpose of flavor analysis.

The ingredients belonged to the following 15 categories: spice, vegetable, fruit, plant derivative, nut/seed, cereal/crop, dairy, plant, pulse, herb, meat, fish/seafood, beverage, animal product, and flower. Category-wise ingredient statistics of regional cuisines is provided in Annexure A.4, Table A.4.1.

5.2 STATISTICS OF RECIPE SIZE AND INGREDIENT FREQUENCY

We started with the investigation of preliminary statistics of regional cuisines. All the eight regional cuisines under consideration showed bounded recipe-size distribution (Figure 5.1). While most cuisines followed a uni-modal distribution, Mughlai cuisine showed a strong bimodal distribution and had recipes with large sizes when compared with the rest. This could be an indication of the fact that Mughlai is the derivative of royal cuisine. To understand the ingredient usage pattern, we ranked ingredients according to decreasing usage frequency within each cuisine. As shown in Figure 5.2, all cuisines showed strikingly similar ingredient usage profile reflecting the pattern of Indian cuisine (Figure 5.2, inset). While indicating a generic culinary growth mechanism, the distributions also show that certain ingredients are excessively used in cuisines depicting their inherent 'fitness' or popularity within the cuisine.

Table 5.2: List of ingredient categories and corresponding ingredient counts as found in all sub-cuisines.

Ingredient Category	Bengali	Gujarati	Jain	Maharashtrian	Mughlai	Punjabi	Rajasthani	South Indian
spice	25	23	26	25	24	33	21	25
vegetable	14	23	29	14	15	29	16	23
fruit	13	19	25	9	16	22	5	14
plant derivative	8	7	11	7	8	13	4	6
nut/seed	12	12	12	11	11	13	8	10
cereal/crop	6	10	11	6	9	12	7	9
dairy	7	6	8	6	7	10	5	7
plant	2	3	3	3	4	5	4	5
pulse	4	6	5	4	5	6	5	6
herb	2	2	5	3	3	4	2	3
meat	3	0	0	2	0	1	0	0
beverage	1	0	1	1	0	1	0	0
fish/seafood	2	0	0	0	0	0	0	2
animalproduct	2	0	1	1	2	2	0	2
flower	1	1	1	1	1	1	1	1
additive	0	0	0	0	0	0	0	1

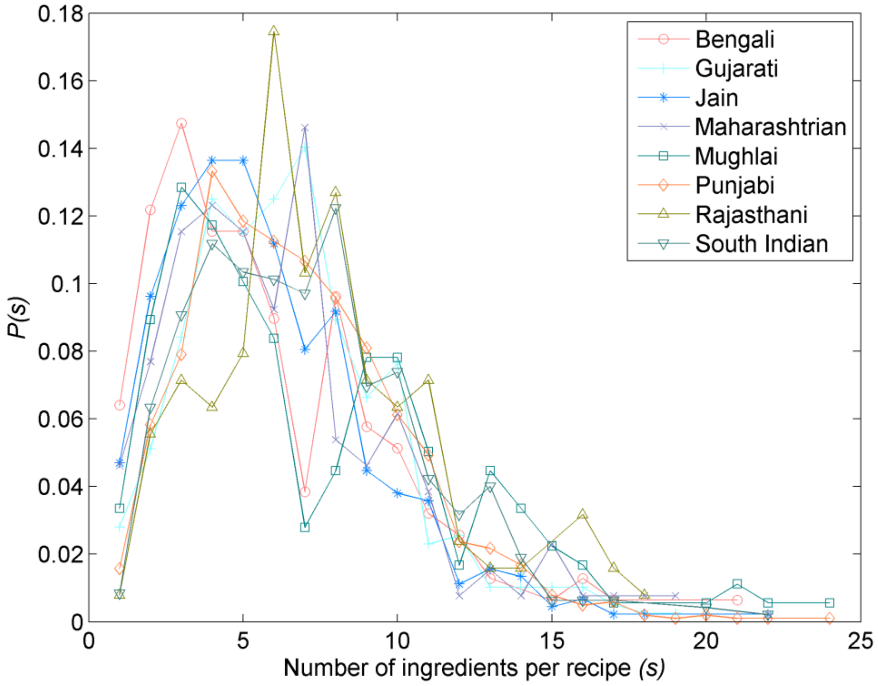

Figure 5.1: Recipe size distributions. The plot of probability of finding a recipe of size s in the cuisine. Consistent with other cuisines, the distributions are bounded. Mughlai and Punjabi cuisines have recipes of large sizes compared to other cuisines.

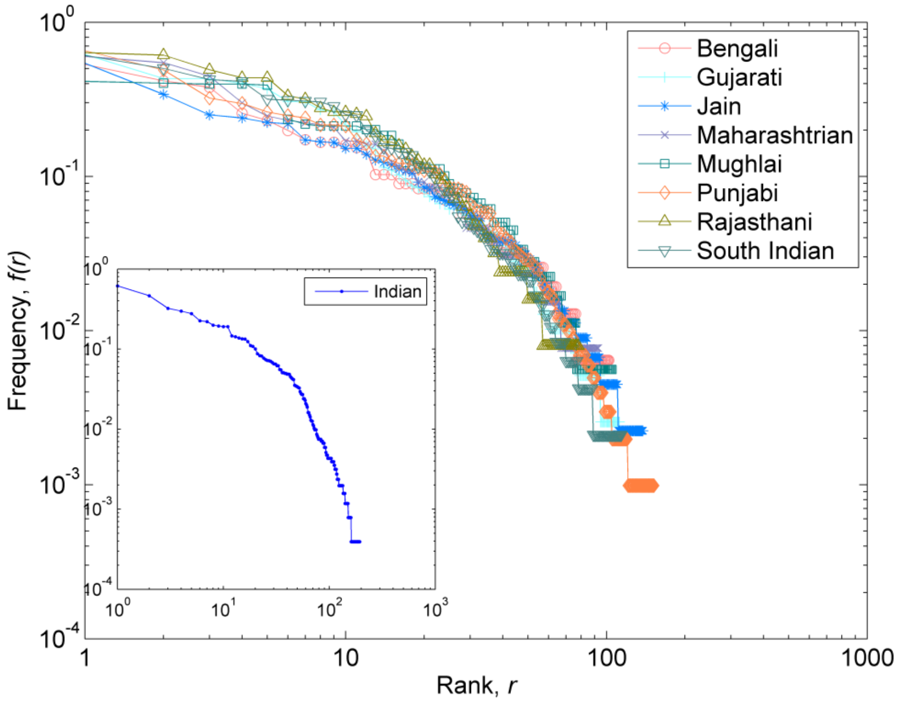

Figure 5.2: Frequency-Rank distributions. Ingredients ranked as per their frequency of use in the cuisine. Higher the occurrence, better the rank of the ingredient. All the cuisines have similar ingredient distribution profile indicating generic culinary growth mechanism. Inset shows the ingredient frequency-rank distribution for the whole Indian cuisine.

5.3 REGIONAL CUISINES OF INDIA EXHIBIT NEGATIVE FOOD PAIRING

We found that all regional cuisines are invariantly characterized by average food pairing lesser than expected by chance. This characteristic negative food pairing, however, varied in its extent across cuisines. Mughlai cuisine, for example, displayed the least inclination towards negative pairing ($\Delta N_s = \overline{N}_s^{Mughlai} - \overline{N}_s^{Rand} =$ -0.758 and Z-score of -10.232), whereas, Maharashtrian cuisine showed the most negative food pairing ($\Delta N_s = \overline{N}_s^{Maharashtrian} - \overline{N}_s^{Rand}$ =-4.523 and Z-score of -52.047). Figure 5.3 depicts the generic food pairing pattern observed across the regional cuisines of India. We found that the negative food pairing is independent of recipe size, as shown in Figure 5.4. This indicates that the bias in food pairing is not an artefact of averaging over recipes of all sizes and is a quintessential feature of all regional cuisines of India. Note that, across cuisines, the majority of recipes are in the size-range of around 3 to 12. Hence the significance of food pairing statistics is relevant below the recipe size cut-off of ~12.

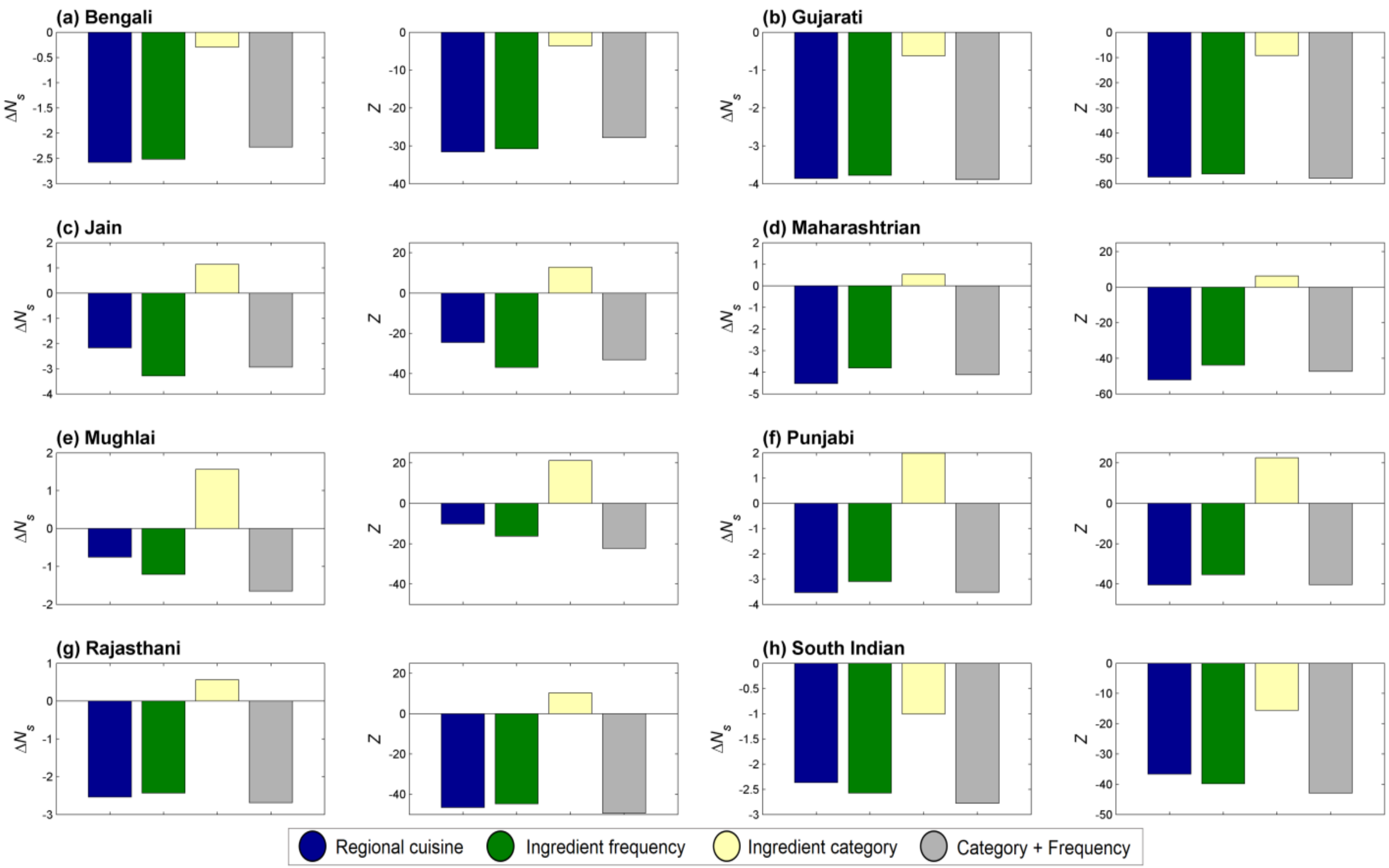

Figure 5.3: ΔN_s and its statistical significance. The variation in ΔN_s for regional cuisines and corresponding random controls signifying the extent of bias in food pairing. Statistical significance of ΔN_s is shown in terms of Z-score. 'Regional cuisine' refers to each of the eight cuisines analyzed; 'Ingredient frequency' refers to the frequency controlled random cuisine; 'Ingredient category' refers to ingredient category controlling random cuisine; and 'Category + Frequency' refers to random control preserving both ingredient frequency and category. Among all regional cuisines, Mughlai cuisine showed least negative food paring (ΔN_s = -0.758) while Maharashtrian cuisine had most negative food pairing (ΔN_s= -4.523).

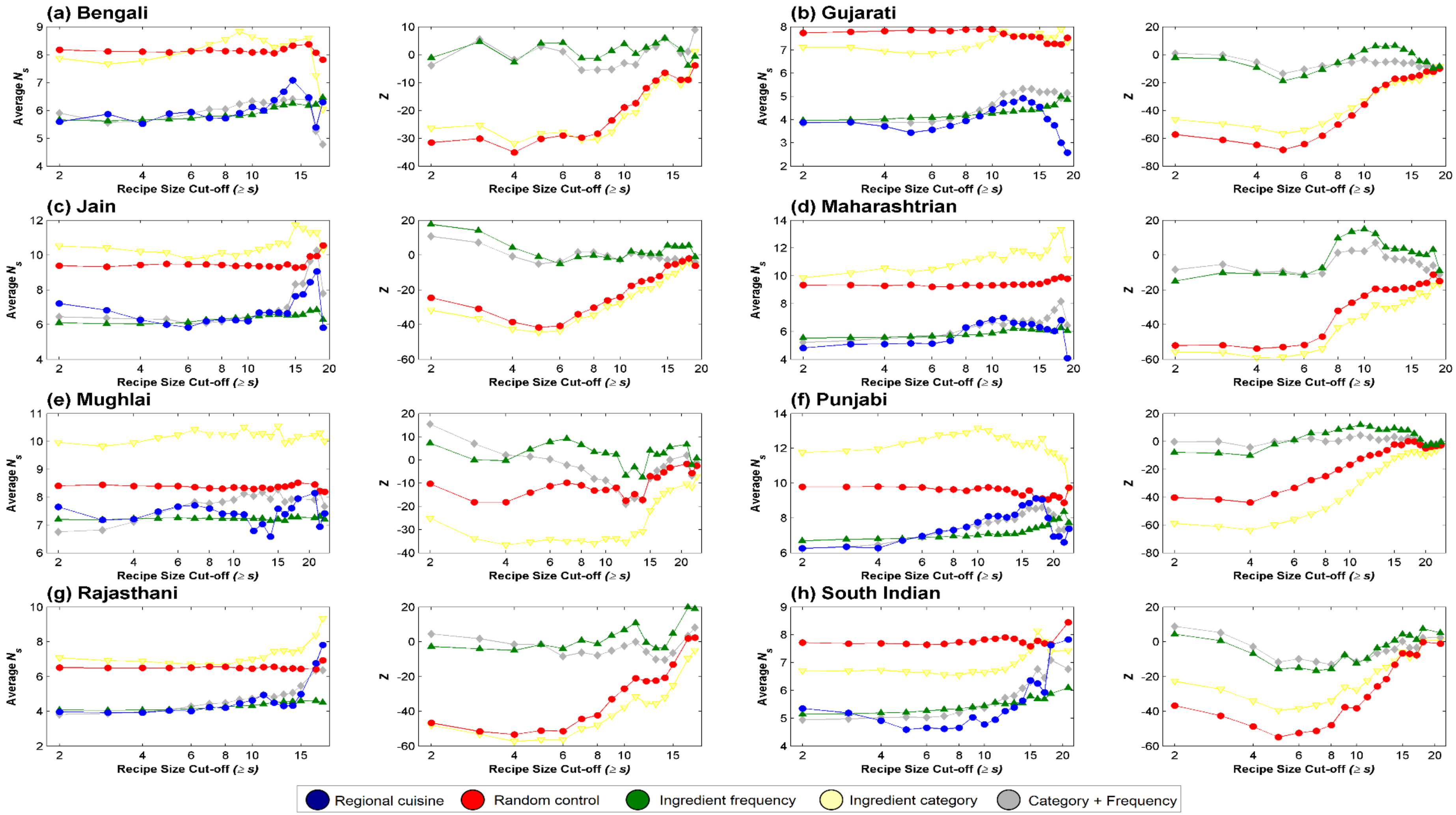

Figure 5.4: Variation in average N_s and its statistical significance. Change in $\overline{N_s}$ with varying recipe size cut-offs reveals the nature of food pairing across the spectrum of recipe sizes. The $\overline{N_s}$ values for regional cuisines were consistently on the lower side compared to their random counterparts. Category controlled random cuisine displayed average N_s variation close to that of the 'Random control'. Frequency controlled as well as 'Category + Frequency' controlled random cuisines, on the other hand, displayed average N_s variations close to that of the real-world cuisine.

We further investigated for possible factors that could explain the negative food pairing pattern observed in regional cuisines. We created randomized controls for each regional cuisine to explore different aspects that may contribute to the bias in food pairing. In the first control, frequency of occurrence of each ingredient was preserved at the cuisine level ('Ingredient frequency'). In the second control, the category composition of each recipe was preserved ('Ingredient category'). A third composite control was created by preserving both category composition of each recipe as well as the frequency of occurrence of ingredients ('Category + Frequency').

Interestingly, ingredient frequency came out to be a critical factor that could explain the observed bias in food pairing as reflected in $\overline{N_s}$ (Figure 5.3). The pattern of food pairing across different size-range of recipes is also consistent with this observation (Figure 5.4). On the contrary, category composition itself turned out to be irrelevant and led to food pairing that was similar to that of a randomized cuisine. Further, the control implementing a composite model featuring both the above aspects recreated food pairing observed in regional cuisines. Thus the frequency of occurrence of ingredients emerged as the most central aspect which is critical for rendering the characteristic food pairing.

5.4 FOOD PAIRING AT RECIPE LEVEL

Looking into the food pairing at recipe level, we analyzed the nature of distribution of food pairing among recipes (N_s^R). Our analysis showed that the negative ΔN_s observed for cuisines was not an averaging effect. The N_s^R values tend to follow an exponential distribution, indicating that the number of recipes exponentially decays with increasing N_s^R. To address the noise due to the small size of cuisines, we computed cumulative distribution ($P(\leq N_s^R)$) as depicted in Figure 5.5. The nature of cumulative distribution for an exponential probability distribution function ($P(\leq N_s^R) \propto e^{-\alpha N_s^R}$) would be of the following form:

$$P(\leq N_s^R) = a + \frac{(k-a)}{1+e^{-\alpha N_s^R}} \tag{5.1}$$

We found that all regional cuisines show a strong bias towards recipes of low N_s^R values as observed in Figure 5.5. For each regional cuisine, the bias was accentuated in comparison to corresponding random cuisines, as reflected in the exponents shown in Table A.3.2. Once again Mughlai cuisine emerged as an outlier, as the nature of its N_s^R distribution did not indicate a clear distinction from that of its random control. Consistent with the observation made with $\overline{N_s}$ and ΔN_s statistics (Figure 5.3 and Figure 5.4), we found that controlling for frequency of occurrence of ingredients reproduces the nature of N_s^R distribution across all regional cuisines (barring the Mughlai cuisine). This further highlights the role of ingredient frequency as a key factor in specifying food pairing at the level of recipes as well.

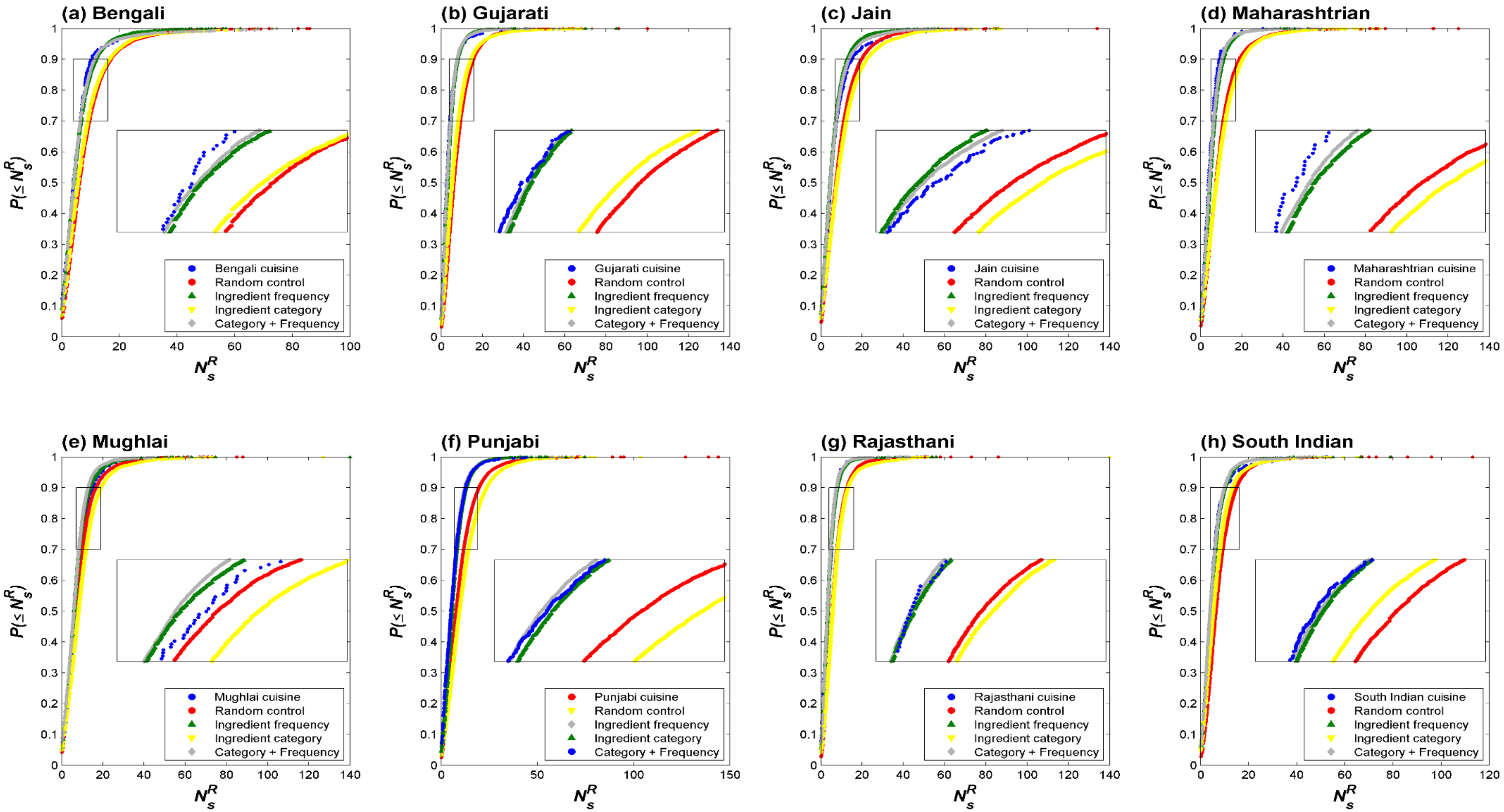

Figure 5.5: Cumulative probability distribution of N_s^R values for regional cuisines and their random controls. Cumulative distribution of N_s^R indicates the probability of finding a recipe having food pairing less than or equal to N_s^R. The data of regional cuisines as well as those of their controls were fitted with a sigmoid equation indicating that the $P(N_s^R)$ values fall exponentially. The exponent α (Equation 5.1) refers to the rate of decay; larger the α more prominent is the negative food pairing in recipes of a cuisine. As evident from Table A.3.2, N_s^R distribution of the controls based on 'Ingredient Frequency' as well as 'Category + Frequency' displayed recipe level food pairing similar to

5.5 FOOD PAIRING AT THE LEVEL OF INGREDIENT PAIRS

Beyond the level of cuisine and recipes, the bias in food pairing can be studied at the level of ingredient pairs. We computed the co-occurrence of ingredients in the cuisine for increasing value of flavor profile overlap (N). We found that the fraction of pairs of ingredients with a certain overlap of flavor profiles ($f(N)$) followed a power law distribution $f(N) \propto N^{\gamma}$ (Figure 5.6). This indicates that the higher the extent of flavor overlap between a pair of ingredients, the lesser is its usage in these cuisines. Table A.3.3 lists the γ values for each of the regional cuisines.

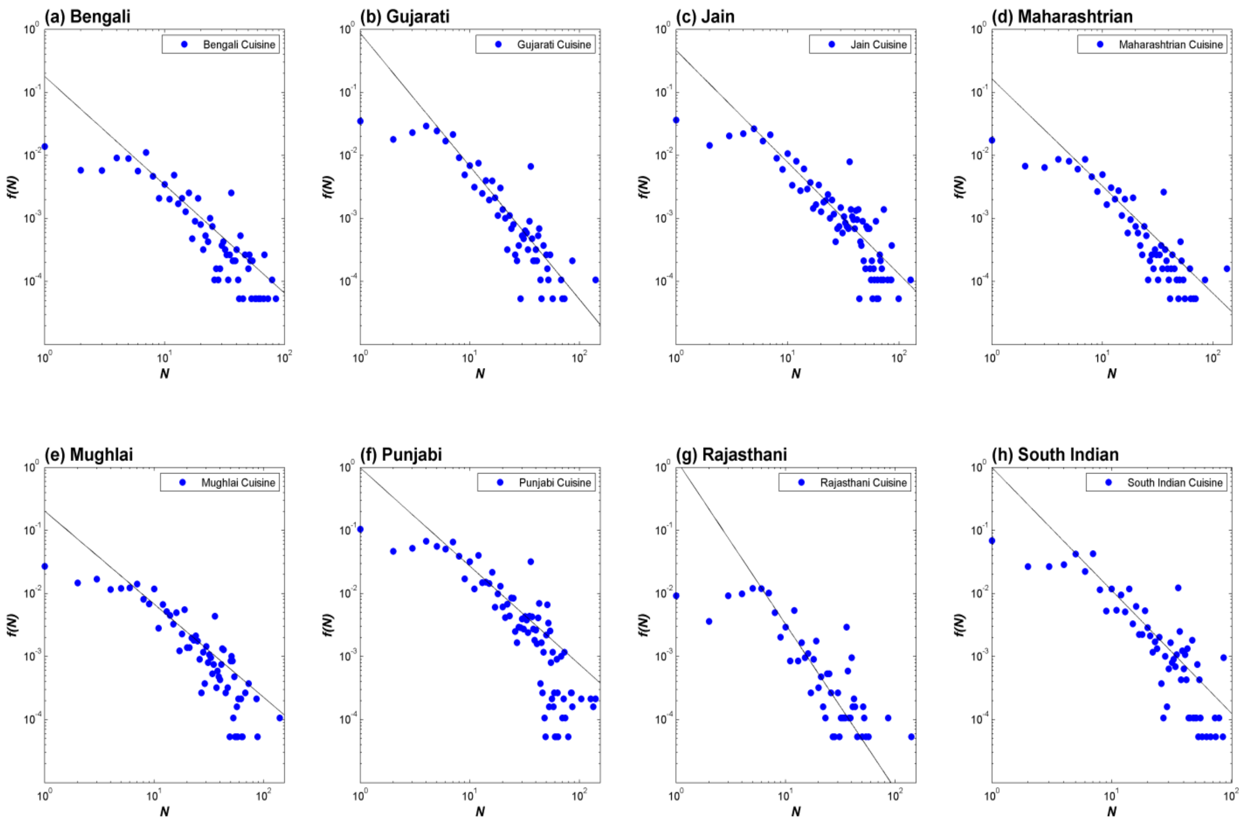

Figure 5.6: Co-occurrence of ingredients with increasing extent of flavor profile overlap. Fraction of ingredient pair occurrence (f(N)) with a certain extent of flavor profile overlap (N) was computed to assess the nature of food pairing at the level of ingredient pairs. Generically across the cuisines, it was observed that the occurrences of ingredient pairs dropped as a power law with increasing extent of flavor profile sharing. This further ascertained the negative food pairing pattern in regional cuisines, beyond the coarse-grained levels of cuisine and recipes.

5.6 CONTRIBUTION OF INDIVIDUAL INGREDIENTS TOWARDS FOOD PAIRING

For each of the regional cuisines, we calculated the contribution of ingredients (χ_i) towards the food pairing pattern. For an ingredient whose presence in the cuisine does not lead to any bias, the value of χ_i is expected to be around zero. With an increasing role in biasing food pairing towards the positive (negative) side, χ_i is expected to be proportionately higher (lower). Figure 5.7 shows the distribution of ingredient contribution (χ_i) and its frequency of occurrence, for each regional cuisine. Ingredients that make a significant contribution towards food pairing could be located, in either the positive or negative side, away from the neutral vertical axis around χ_i=0. Significantly, spices were consistently present towards the negative side, while milk and certain dairy products were present on the positive side across cuisines. Prominently among the spices, cayenne consistently contributed to the negative food pairing of all regional cuisines. Certain ingredients appeared to be ambivalent in their contribution to food pairing. While cardamom contributed to the positive food pairing in Gujarati, Mughlai, Rajasthani, and South Indian cuisines, it added to negative food pairing in Maharashtrian cuisine. Green bell pepper tends to contribute to negative food pairing across the cuisines except in the case of Rajasthani cuisine. Details of χ_i values of prominent ingredients for each regional cuisine are presented in Table A.3.4.

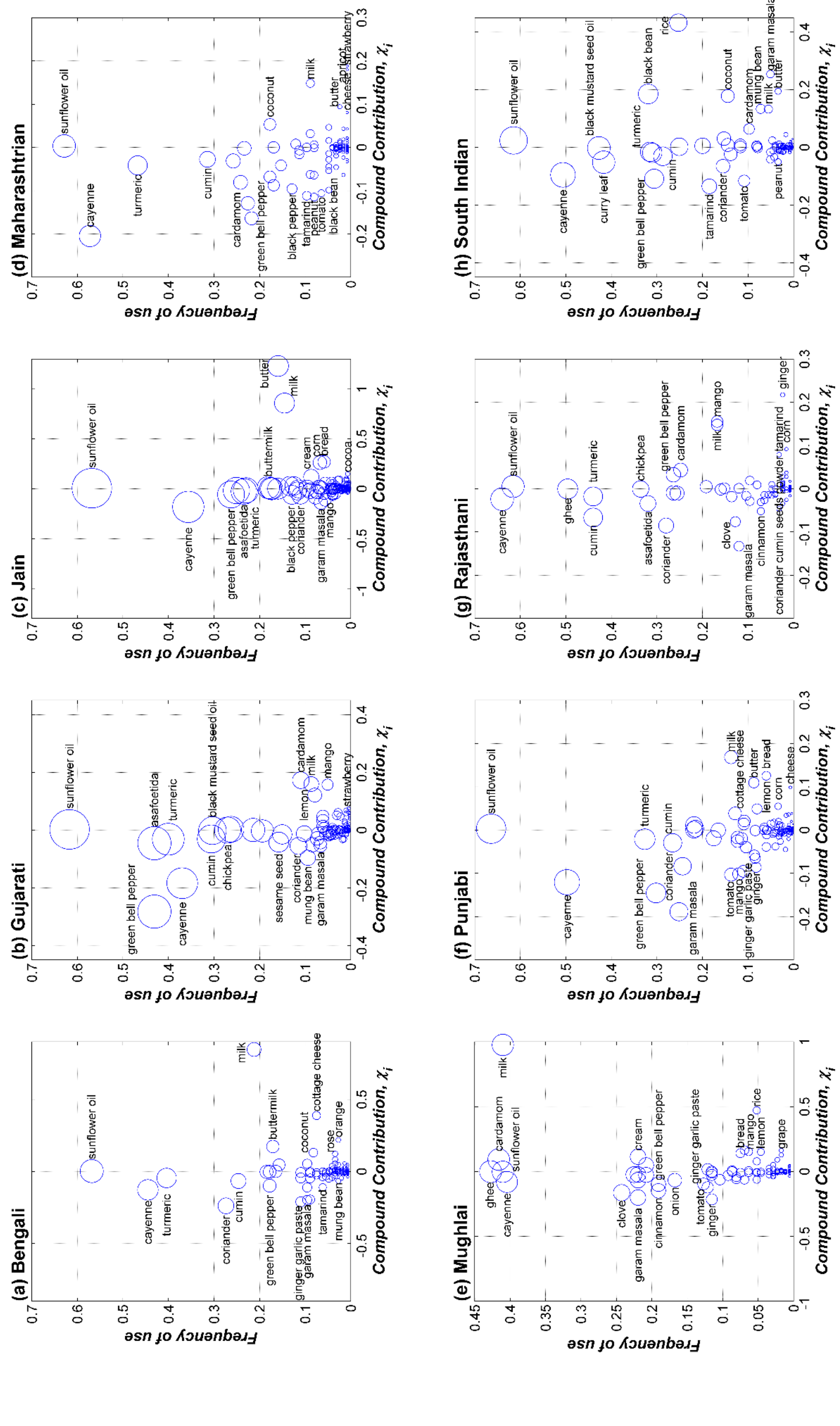

Figure 5.7: Contribution of ingredients (χ_i) towards flavor pairing. For all eight regional cuisines we calculated the χi value of ingredients that indicates their contribution to flavor pairing pattern of the cuisine and plotted them against their frequency of appearance. Sizes of circles are proportional to frequency of ingredients. Across cuisines, prominent negative contributors largely comprised of spices, whereas a few dairy products consistently appeared on the positive side.

5.7 ROLE OF INGREDIENT CATEGORIES IN FOOD PAIRING

As discussed earlier, the random cuisine where only the category composition of recipes was conserved tends to have food pairing similar to that of the 'Random control' (Figure 5.3 and Figure 5.4). This raises the question whether the ingredient category has any role in determining the food pairing pattern of the cuisine. Towards answering this question, we created random cuisines wherein we randomized ingredients within one category, while preserving the category and frequency distribution for the rest of the ingredients. The extent of contribution of an ingredient category towards the observed food pairing in the cuisine is represented by ΔN_s^{cat}. Figure 5.8 depicts the significance of ingredient categories towards food pairing of each regional cuisine. Interestingly, the pattern of category contributions presents itself as a 'culinary fingerprint' of the cuisine.

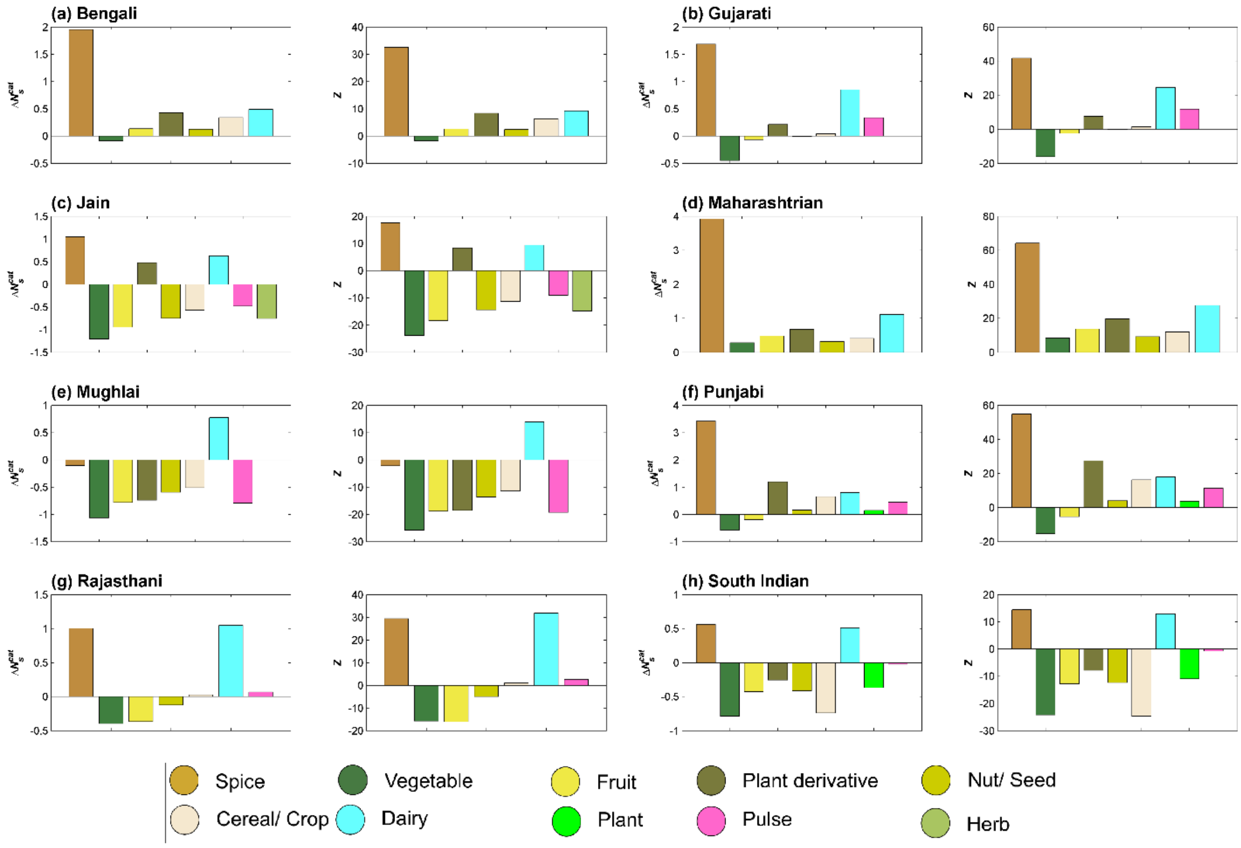

Figure 5.8: Contribution of individual categories (ΔN_s^{cat}) towards food pairing bias and its statistical significance. Randomizing ingredients within a certain category provides an insight into their contribution towards bias in food pairing. The spice and dairy category showed up as prominent categories contributing to the negative food pairing of regional cuisines.

The 'spice' category was the most significant contributor to negative food pairing across cuisines with the exception of Mughlai cuisine. Another category that consistently contributed to negative food pairing was 'dairy'. On the other hand, 'vegetable' and 'fruit' categories tend to bias most cuisines towards positive food pairing. Compared to the above-mentioned categories, 'nut/seed', 'cereal/crop', 'pulse' and 'plant derivative' did not show any consistent trend. 'Plant' and 'herb' categories, sparsely represented in cuisines, tend to tilt the food pairing towards the positive side. In Mughlai cuisine, all ingredient categories, except 'dairy', tend to contribute towards positive food pairing. This could be a reflection of the meager negative food pairing observed for the cuisine (Figure 5.3). The above observations were found to be consistent across the spectrum of recipe sizes (Figure 5.9).

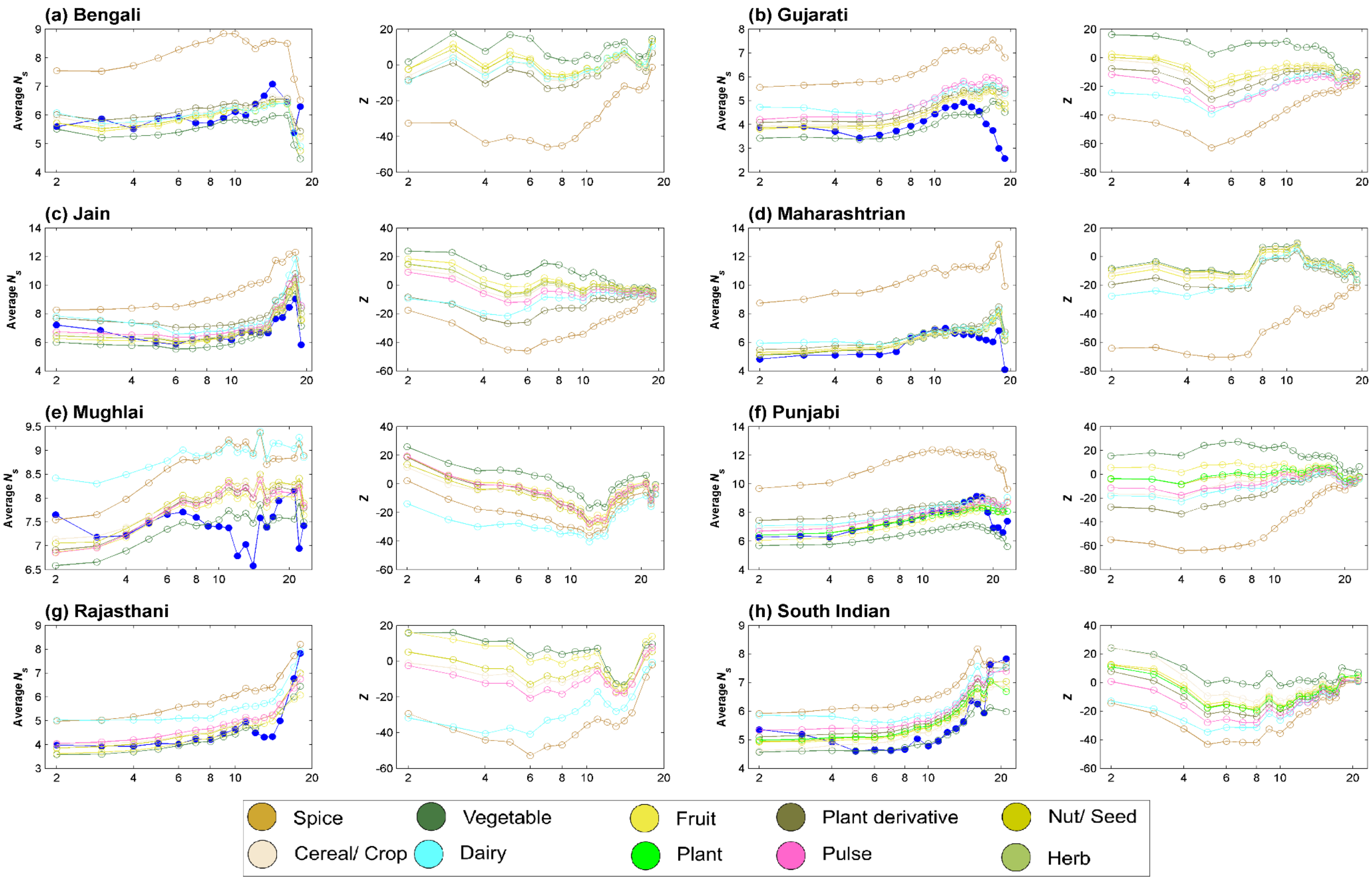

Figure 5.9: Variation in category contribution and its statistical significance. Across the spectrum of recipe sizes, we observed broadly consistent trend of contribution of individual categories towards food pairing bias.