Food Components as Markers Linking Health and Environment: Statistical Invariance Analysis of *in natura* Diet

Masatoshi Funabashi

Sony Computer Science Laboratories, Inc., Tokyo, Japan

Email address:
masa_funabashi@csl.sony.co.jp

Abstract: Diets are key factors that link environmental and human health. Global degradation of ecosystems and health state are firmly related to diet transition and production system. We propose a distinction of *in cultura* and *in natura* diet by the culture condition and consequent environmental load it imposes, which leads to the definition of *in natura* diet as a possible alternative for sustainable diet. By considering food components as markers linking health and environment, we investigate statistically invariant features that characterize the difference between *in cultura/natura* diets on 2 independent databases, INFOODS food composition database and Synecoculture products. Plural distinctive features between *in cultura/natura* diets were discovered in numerically sampled intake distribution. Taking the food diversity limit, *in natura* diet tended to be more consistent in relation to larger population with major components and minerals, and a significant difference with *in cultura* diet was encrypted in variance component. Possible interpretation of the results may relate recent health burden to historical transition from *in natura* to *in cultura* diet.

Keywords: Food Components, Statistical Analysis, Sustainable Diet

1. Introduction

1.1. Diet and Food Production: Health and Environmental Burden

Food, health and environment is a trilemma to achieve sustainable social-ecological system [1]. Recent increase of chronic diseases is raising medicare expenditures and has become one of the tight burdens of national budgets in many countries (e.g. [2] [3]). This problem presumably originates from the change of lifestyle along with the globalization, in which diet is an essential factor [4] [5].

Besides, modern production systems that support food supply is considered to exceed the acceptable limit of environmental burden. Agricultural land use is reported as the most devastating factor of biodiversity on the globe [6], and exacerbate the plant extinction rate more than climate change [7]. Among a wide range of stakeholders including governments, scientists, and civil society, conventional agriculture and distribution system is considered not a sustainable option for the next generation [8].

For sustainability purpose, we need to seek ways to measure the impact of food systems on both health and environment, and define the condition of sustainable diet that is both health promoting and ecologically sound.

1.2. Bottom-up Components Analysis

Modern nutriology has clarified the five major nutrients necessary to support the basic functioning of our metabolism: proteins, carbohydrates, lipids, vitamins and minerals [9]. While the major components are decisive to performances in short-time scale, a huge variety of complementary compounds such as phytochemicals and trace elements are considered to support long-term health protective effect [10] [11] [12]. These long-tail compounds are increasingly studied with the development of empirical measurement technology such as metabolomics [13]. Especially phytochemicals are the source of most pharmaceutical products [14], and are candidates for long-tail drug discovery that is expected to be more compatible to the whole functioning of cellular metabolism than target-specific drugs [15].

While studying the long-tail part of food components, complexity of interaction arises in determining the net effect

of bioactive compounds working in combination. Phytochemicals are known to express its diverse functionality as additive and synergetic interactions with other elements [16]. Theoretically, the complexity of such combination can extend to the whole system size [17]. This implies that the reconstruction of diet from element level such as synthetic supplement is not effective considering the complexity of the actual interaction in whole food, neither sustainable for the cost to reconstruct the whole biological process artificially.

1.3. Top-down Cohort Analysis

Traditional food studies have also focused from macroscopic perspective with cohort analysis based on whole food consumption. Functional effect of food containing bioactive chemicals has been increasingly identified (e.g. [18] [19]). Nutritional and health state comparison has been investigated in plural regions with changing dietary patterns [20]. Recently, particular examples have been investigated on rapid food system transition in China, in relation to health risks [21] [22] [23].

Among empirical studies that connect between food systems, human, health, and environmental factors, beneficial food systems are reported from indigenous and traditional food communities with a sustainable use of local ecosystems (e.g. [24] [25] [26]). The common condition is the wide introduction of various food items available in natural local environment, especially edible wild plants. If widely adopted, these alternative diets could both reduce health risk and improve ecological state [1].

Bioactive compounds with health-protective effect in wild edible plants are widely studied [27] [28] [29]. In some case, it is possible to substitute conventional diet with wild food without losing major nutrition profile [30]. The use of wild edible plants has common feature with traditional preventive medicine such as traditional Chinese medicine and Ayurveda, which have evolved with heuristic discovery of wild plants utility.

Culture conditions such as wild plants and cultivated crops also involve important health factors other than nutrition, such as gut microbe [31] [32]. It is known that soil microbiological flora affects gut microbe composition through food intake, which can only be evaluated with the analysis of net effect including the epiphenomena associated with food production systems.

1.4. Synthetic Perspective: Food Components as Markers Linking Health and Environment

As an integrated perspective to characterize sustainable diet, I propose to make use of food components databases to derive statistical measures as markers linking health and environment. Since food components are both influenced by culture condition and related to human health, it is possible to consider diet as an interface of health and environment [33].

The physiological property of bioactive compounds and traditional practice of ecologically sound food system both contain some necessary conditions for sustainable diet in terms of health and environment. Though, it has not been integrated yet to achieve a sufficient condition to define sustainable diet in a global context.

Sufficient condition for sustainable diet is an integrated attribute of food system that satisfies desired health and environmental quality, to which I expect to find consistent statistical measures with respect to food components distributions.

More precisely, health effect is biologically referring to missing inheritability of disease, in which epigenetic profile based on the life-course activity including diet habit accounts for the susceptibility to disease [34]. This is a common objective with current nutriology seeking for a complete list of bioactive compound and its effect through food intake [11] (p.283). Environmental quality is typically the degree of biodiversity that is under direct threat of agricultural activity [6] [7]. It also projects influence on compounds distribution, especially phytochemicals as it mediates the interaction between plants and other species [35].

2. *in natura* Diet: Definition

2.1. *in cultura* and *in natura* Diet

In order to yield a possible definition of sustainable diet, I proposed to divide the current food systems into 2 categories based on the culture condition [36]:

1) *in cultura* diet, that refers to the food grown by humans with the conventional agricultural practice, i.e. tillage, fertilizer, and chemicals. It principally consists of growing cultivars, aiming at physiologically optimum growth of a single crop. It can be represented as the progress toward the complete control of culture condition regardless of environmental load, and can be symbolized with a factory culture system in spaceship as a social-scientific ideal type [37].

2) *in natura* diet, based on the harvest from wild edible species or cultivars growing in natural condition based on the self-organization of ecosystem. The production is based on the ecological optimum that represents actual observed range of conditions in nature, where a species grows in association with others [38]. In case of human intervention such as introduction of edible species and harvest, only point disturbance that stimulates totally positive biodiversity response is accepted. As an ideal type, it can be symbolized with pre-anthropogenic ecosystem such as the era of the dinosaurs.

The terms *in cultura* and *in natura* [39] is an ecological extension of biological terms *in vitro* (cells in artificial culture medium) and *in vivo* (experiments on live organisms in a lab), in this case they divide culture conditions in a non-laboratory environment according to the degree of human intervention on the growing process of food.

The definition of *in cultura* and *in natura* diet is summarized in Table.1. In contrast to the conventional supply of *in cultura* diet, I propose a possible candidate for sustainable diet as *in natura* diet, if sufficiently productive

practice is possible without environmental degradation.

Table 1. *Definition of in cultura and in natura diet.*

	in cultura	*in natura*
Crop genotype	Cultivars	Wild species, Cultivars
Culture condition	Tillage, fertilizer, chemicals	Self-organization, point disturbance
Optimization	Physiological	Ecological
Typical fields	Farmland	Natural ecosystem, Synecoculture
Ideal type	Factory culture in a spaceship	Natural ecosystems before anthropogenesis

2.2. Synecoculture Experiment: Sustainable Production of *in natura* Diet

Synecoculture project has been testing the high-density mixed polyculture of edible plants to establish environmentally sound production system of *in natura* diet [33] [40]. So far, evidences on biodiversity and partial food composition were obtained: Synecoculture system promoted the field biodiversity and the products were proved to express health-beneficial secondary metabolites, particularly terpenoids and flavonoids, with the metabolome and taste analysis [41] [42] [43]. Minerals concentration in the Synecoculture products also complemented the soil dose deficiency with respect to the conventional standard, which presumably attributes to the effect of enhanced biodiversity [44]. Taken together, it is possible to consider the Synecoculture as an environmentally sound system for the production of *in natura* diet. Including the Synecoculture products, we will further study the statistical feature of *in natura* food components in comparison to *in cultura* counterpart.

3. Statistical Invariance Analysis

3.1. Theory

3.1.1. Sampling of Food Systems and Intake Distributions Based on the Food Diversity Limit

We consider a general framework of nutrition intake from a food system in order to establish statistical measures invariant to the particularity irrelevant to *in cultura/natura* distinction. Statistical invariance of *in cultura/natura* diets signifies statistical features that are conditional only to *in cultura/natura* distinction and remain invariant to other conditions.

As depicted in Figure 1, consider the database of all diets of all food components in this world. It represents the true distribution of all existing food components in this world. Next, consider a particular food system of l food items with *in cultura/natura* attribute. This is a possibly biased sampling depending on local availability and people's preferences. For those living on this food system, actual nutrition intake is the sampling from l food items, which generally converges to a normal distribution by central limit theorem regardless of the original distribution, if sufficient diversity of food and choice is assured. Let's define that our diet is the meal-wise sum (linear combination) of n food items out of l, that iterates m times to express the physiological effect. Then, usually $n > 30$ is sufficient to observe convergence to normal distribution with large m (results not shown). By finely choosing n and m, one can randomly assimilate any history of component intake that converges to the normal distribution.

We denote this normal distribution representing the food component oral ingestion with *in cultura* and *in natura* diets as the intake distributions $C(x|\mu_C, \sigma_C)$ and $N(x|\mu_N, \sigma_N)$, respectively, where x is the dose of a component, μ and σ represent its mean value and variance. For simplicity, we consider the component-wise intake distribution with one-dimensional variable x, but the generalization to multivariate normal distribution is possible. Since actual σ increases by a factor of n, we renormalize σ by dividing with n, to yield the intake distribution invariant to n.

The random sampling premise to justify the central limit theorem corresponds to the diversity of food. The convergence of the intake distribution to a normal distribution is proportional to the variety of recipes in the food system. We consider the highest limit of food diversity as random uniform sampling, and numerically yield the intake distribution for the analysis.

Based on the intake distributions, we consider statistically invariant properties that are sensitive to *in cultura/natura* difference and remain invariant to other bias that may arise in this model.
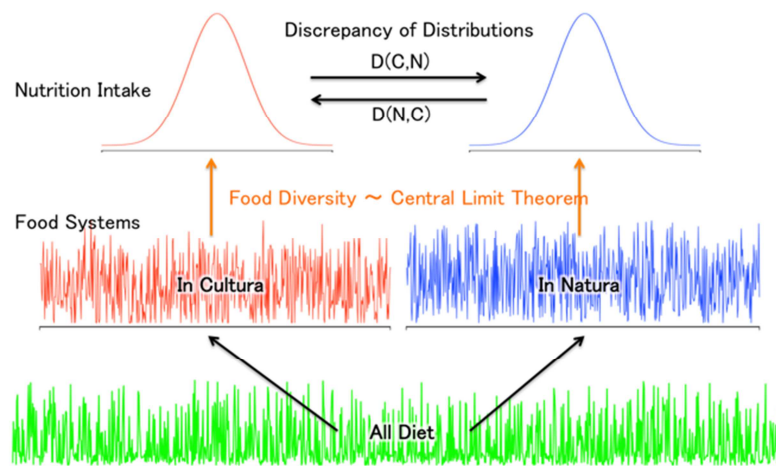
Figure 1. *Sampling of food systems and nutrition intake distributions.*

3.1.2. Actual Confidence Level Based on the Food Diversity Limit

Intake distributions are based on the random sampling from corresponding food systems, which are also based on unknown sampling from all possible foods. In order to discuss the general difference between *in cultura/natura* diets, one needs to infer their true distribution, or their population mean and variance. In reality, this can never be achieved as the complete listing of *in cultura/natura* diets is impossible. Still, it is possible to work on the convergence of the estimates with respect to the diversity of food systems: Which of in cultura/natura diets can better represent the larger population when the given food system is limited in elements number l ?

For that purpose, we can consider the actual confidence level of the estimates of mean value µ and variance σ of intake distributions conditioned by l. It can be formulated based on the standard theory of estimation of population mean and variance, which leads to examine their $t-$based and χ^2-based 95% confidence intervals with degrees of freedom l and $l-1$, respectively: How much intake distributions from food system with l_2 elements can infer the mean/variance values of those with $l_1 (> l_2)$?

To measure that, we can define the actual confidence level of mean and variance, $ACL(\mu)$ and $ACL(\sigma)$, with parameters l_1 and l_2. The actual confidence level is defined as the percentage that the actual mean/variance value of the intake distribution from food system with l_1 species is situated within the 95% confidence range of the estimation from the food system with l_2.

This analysis will evaluate the consistency of intake distribution, therefore the stability of components distribution among diverse food systems, when culture condition differed between *in cultura* and *in natura*.

3.1.3. Asymmetric KL-divergence Between in Cultura and in natura Diet Intake Distributions

The derivation of intake distribution may contain lots of bias that affect mean and variance parameters: Dose unit and measurement method may differ by database. Sampling from a food system may be biased by preferential recipe, which may not be uniform in a long run. Additionally, intake distribution does not necessary represent actual intake in the metabolism, as it goes through complex digestion process with diverse absorption rate, influenced by the genetic profile of eating person, gut microbiota. In-body combinatory interactions of bioactive compounds may also alter consequent physiological effect [16]. To get rid of these variabilities and obtain invariant statistical measures, we consider the use of Kullback-Leibler divergence $D(C, N)$ and $D(N, C)$ between *in cultura/natura* intake distributions, which are defined as follows [45]:

$$D(\mathrm{C}, N) := \int C \, (\log C - \log N) dx, \qquad (1)$$

$$D(\mathrm{N}, C) := \int N \, (\log N - \log C) dx. \qquad (2)$$

The KL-divergence between continuous distributions is

known to remain invariant under any continuous transformation of parameters, including non-linear deformation. Therefore, these measures are statistically invariant to dose unit change, measurement methods, food ratio bias, and possibly buffers intake efficiency, genetic background diversity, synergetic interactions of components, etc, as long as they are expressed as continuous transformation commonly applied to *in cultura/natura* intake distributions. This is a quite universal assumption that is adopted to study invariant structure of statistical manifold in information geometry [46]. Longer and combinatorial effects of biological variabilities that affect the eating person's background, however, does not necessary remain invariant because these become specific to *in cultura/natura* conditions.

The KL-divergence is asymmetric by definition and represents the degree of discrepancy seen from one distribution to the other. In information theoretical interpretation, $D(C, N)$ for example, represents the amount of information gain when the sample distribution changed from N to C through observation. More simply, it represents the amount of change in terms of distribution from N to C.

3.1.4. Decomposition of KL-divergence Between Mean and Variance Components

Based on the information geometry, it is possible to decompose the KL-divergence into the following [43]:

$$D(\mathrm{C}, N) = D(\mathrm{C}, CN) + D(CN, N), \qquad (3)$$

$$D(\mathrm{N}, C) = D(\mathrm{N}, NC) + D(NC, C). \qquad (4)$$

Where $CN = CN(\mathrm{x}|\mu_N, \sigma_C)$ and $NC = NC(\mathrm{x}|\mu_C, \sigma_N)$. This means that the term $D(C, CN)$ and $D(N, NC)$ represent the discrepancy with the difference of the mean value components, while $D(CN, N)$ and $D(NC, C)$ correspond to those of the variance components. For simplicity, we call $D(C, CN)$ and $D(N, NC)$ as the mean components, $D(CN, N)$ and $D(NC, C)$ as the variance components of $D(C, N)$ and $D(N, C)$, respectively.

This decomposition can orthogonally separate the effect of mean and variance parameters on the discrepancy of intake distributions. It allows us to study the high-order statistics such as variance that may be important to consider the intermittent effect on our metabolism, which has been mostly neglected.

3.2. Databases

3.2.1. INFOODS Database

As an empirical database of food components that contains both *in cultura/natura* diets, we analyzed the INFOODS database with the distinction of *in natura* foods with "Wild (W)" parameter in the Type code, and *in cultura* food without [47]. The species may differ between sampled *in cultura/natura* food systems. Due to available number of data, the analysis performed was limited to the databases of starchy roots & tubers, nuts & seeds, vegetables, fruits, and meat, on the food components that could satisfy sufficiently large sampling number as food systems ($l = 50$, except nuts & seeds with $l = 30$).

3.2.2. Synecoculture Database and Corresponding Japanese Food Composition Table

As another independent example of *in natura* diet, we analyzed the Synecoculture database that comprises dose data of 4 minerals (Na, K, Mg, Ca) in 140 samples of Synecoculture products from 37 vegetable species [43]. The corresponding *in cultura* diet of the same species was obtained from the standard food composition table in Japan [48].

4. Results

4.1. Numerical Simulations of Intake Distributions of in cultura/natura Diets

We simulated the intake distributions of *in cultura/natura* diets from INFOODS database with parameters $n = 300, m = 1000, l = 50$. We took 100 random sampling, which means we obtained 100 different intake distributions for each of *in cultura/natura* diets. This results in the comparison of $100 \times 100 = 10000$ pairs of *in cultura/natura* intake distributions for each food categories (starchy roots & tubers, nuts & seeds, vegetables, fruits, and meat), in varying species profile.

As for Synecoculture database, we took the parameters $n = 100, m = 10000, l = \{5, 10, 15, 20, 25, 30, 35, 37\}$, regarding the convergence to normal distribution and to investigate different orders of food system as the number of species are limited to 37. We estimated the population mean and variance of each mineral component with 95% intervals for both *in cultura/natura* diets with $l = 37$, from which we obtained randomly 1000 pairs of *in cultura/natura* intake distributions, which are expected to represent the random sampling from a larger population of Synecoculture and conventional products in 4 minerals. This assumption is partially supported by the analysis of the actual confidence level that only decreases linearly in smaller sample (Figure 4).

4.2. Means and Variances of Food Components Intake Distributions

We first compare the magnitude relation between *in cultura/natura* mean values of the intake distributions. We calculated the following 2 indices, IE_μ and R_μ, to judge the magnitude relation with respect to inequality and ratio between *in cultura/natura* mean values:

$$IE_\mu := log\{\#(\mu_C > \mu_N)/\#(\mu_N > \mu_C)\}, \quad (5)$$

$$R_\mu := log\{mean(\mu_C/\mu_N)\}, \quad (6)$$

where $\#(\ )$ describes a function that returns the number of pairs μ_C and μ_N satisfying the inequality condition. The $mean(\ )$ function calculates the mean value for all given pair of μ_C and μ_N. The logarithm was taken to compare the ratio between μ_C and μ_N in an equivalent scale.

Table 2 summarizes the results. For both IE_μ and R_μ, the score more than 0 signifies $\mu_C > \mu_N$, while less than 0 corresponds to $\mu_C < \mu_N$. Results of Synecoculture were generated with species-wise means of data. General tendency between μ_C and μ_N could not be strongly supported.

Table 2. *Magnitude relation between mean values of in cultura/natura diets intake distributions.*

Database	IE_μ	R_μ
INFOODS	0.0501271756	0.2243004077
Starchy roots & tubers	$\mu_C > \mu_N$	$\mu_C > \mu_N$
INFOODS	-0.1401787032	0.0575588684
Nuts & seeds	$\mu_C < \mu_N$	$\mu_C > \mu_N$
INFOODS	-0.7365250108	-0.2611843511
Vegetables	$\mu_C < \mu_N$	$\mu_C < \mu_N$
INFOODS	-0.2557792665	0.0734863848
Fruits	$\mu_C < \mu_N$	$\mu_C > \mu_N$
INFOODS	-0.0110553656	0.0396328071
Meat	$\mu_C < \mu_N$	$\mu_C > \mu_N$
Synecoculture	1.0110568367	-0.04227311
Na	$\mu_C > \mu_N$	$\mu_C < \mu_N$
Synecoculture	Inf	0.2858255342
K	$\mu_C > \mu_N$	$\mu_C > \mu_N$
Synecoculture	0.9024795941	0.110054178
Mg	$\mu_C > \mu_N$	$\mu_C > \mu_N$
Synecoculture	-1.5693181629	-0.1531869362
Ca	$\mu_C < \mu_N$	$\mu_C < \mu_N$
Synecoculture	0.0521780117	0.0818531682
Na, K, Mg, Ca	$\mu_C > \mu_N$	$\mu_C > \mu_N$

The results for the variance with the following definition of IE_σ and R_σ are summarized in Table 3.

$$IE_\sigma := log\{\#(\sigma_C > \sigma_N)/\#(\sigma_N > \sigma_C)\}, \quad (7)$$

$$R_\sigma := log\{mean(\sigma_C/\sigma_N)\}. \quad (8)$$

More apparent tendency of $\sigma_C > \sigma_N$ than mean values can be observed, especially in R_σ. This implies that *in cultura* diet may contain larger variance ratio in the intake distribution. Since this tendency is common to Synecoculture products with the comparison of the same species in conventional data, this property might necessarily attributes to the phenotypic plasticity of crops, in response to the artificial intervention on culture condition.

Table 3. *Magnitude relation between variances of in cultura/natura diets intake distributions.*

Database	IE_σ	R_σ
INFOODS	0.0747806822	2.1680785106
Starchy roots & tubers	$\sigma_C > \sigma_N$	$\sigma_C > \sigma_N$
INFOODS	-0.1797134192	1.2607719073
Nuts & seeds	$\sigma_C < \sigma_N$	$\sigma_C > \sigma_N$
INFOODS	-0.8809364652	-0.5393708599
Vegetables	$\sigma_C < \sigma_N$	$\sigma_C < \sigma_N$
INFOODS	0.0054287517	0.7219661015
Fruits	$\sigma_C > \sigma_N$	$\sigma_C > \sigma_N$
INFOODS	0.1320868295	1.0591251652
Meat	$\sigma_C > \sigma_N$	$\sigma_C > \sigma_N$
Synecoculture	0.8500828234	0.4185646996
Na	$\sigma_C > \sigma_N$	$\sigma_C > \sigma_N$
Synecoculture	1.4913616938	0.6949693934
K	$\sigma_C > \sigma_N$	$\sigma_C > \sigma_N$
Synecoculture	1.3267908578	0.9372598959
Mg	$\sigma_C > \sigma_N$	$\sigma_C > \sigma_N$
Synecoculture	-1.2046971183	-0.2437281525
Ca	$\sigma_C < \sigma_N$	$\sigma_C < \sigma_N$
Synecoculture	0.3987953871	0.6232782575
Na, K, Mg, Ca	$\sigma_C > \sigma_N$	$\sigma_C > \sigma_N$

4.3. Actual Confidence Level of Food Components Intake Distributions

The results of the actual confidence level with respect to the mean value $ACL(\mu)$ and variance $ACL(\sigma)$ were classified into 4 qualitative relations: > and < in terms of the overall magnitude relation, $\sim$ as qualitatively difficult to distinguish, and $\subset$ as inclusion when the right distribution covers the left distribution.

The qualitative relations were judged based on the plots of $ACL(\mu)$ and $ACL(\sigma)$ with respect to the available range within $l_1 = \{20, 30, 40, 50, 60, 70, 80, 90, 100\}$, $l_2 = \{10, 20, \cdots, l_1\}$, for INFOODS database. The examples of results with starchy roots & tubers are depicted in Figures 2 and 3.

The result of Synecoculture database was obtained with a quantitative analysis with linear regression of $ACL(\mu)$ and $ACL(\sigma)$ with respect to $l = \{5, 10, 15, 20, 25, 30, 35, 37\}$. The difference of slope Ds of the liner regression was evaluated by the following formula, with the slope of linear regression s_C and s_N of *in cultura/natura* diets:

$$Ds := s_C - s_N, \tag{9}$$

which translates to $ACL(\mu_N) > ACL(\mu_C)$ and $ACL(\sigma_N) > ACL(\sigma_C)$ when $Ds > 0$, inversely $ACL(\mu_N) < ACL(\mu_C)$ and $ACL(\sigma_N) < ACL(\sigma_C)$ when $Ds < 0$. The results are depicted in Figure 4. All linear regressions were significant (p<0.01), which implies that the mean complexity of the 4 mineral components in these species can reduce only linearly with respect to the food system size.
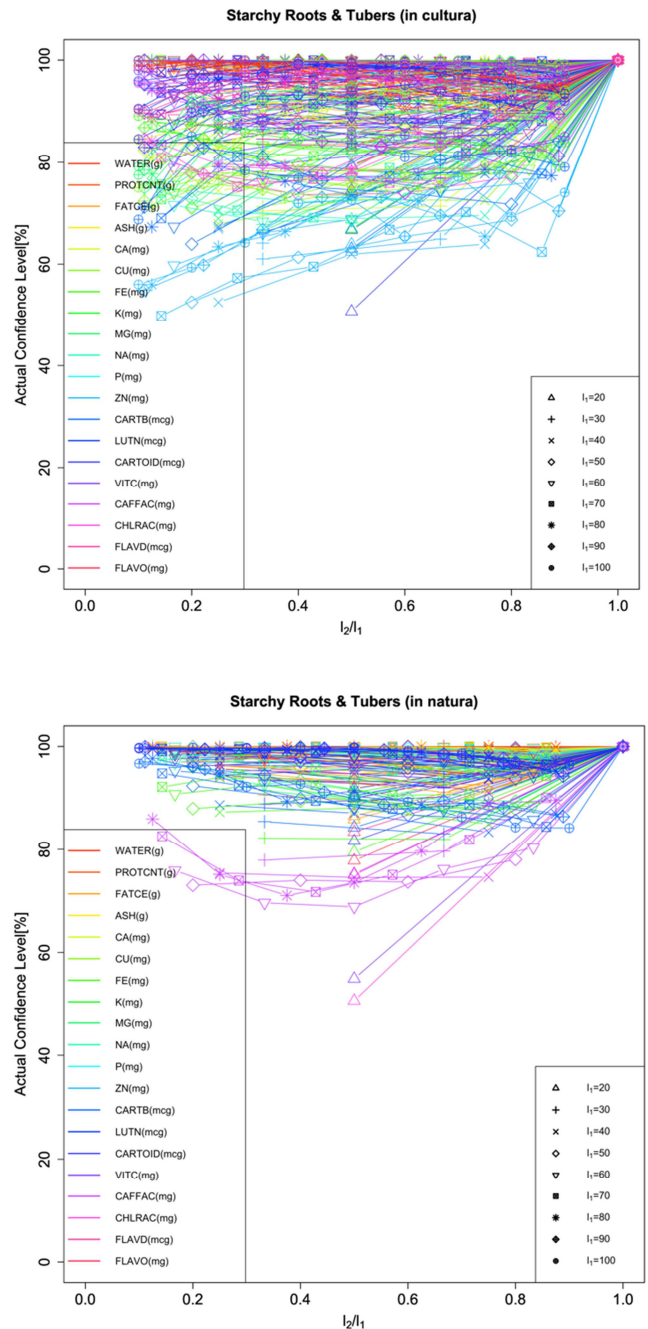
The summary is listed in Table 4. $\#c$ is the number of food components used for the analysis, with which phytochemical number is associated with parentheses. There exists overall tendency of $ACL(\mu_N) > ACL(\mu_C)$, when compared in larger ensemble with respect to the number of food components $\#c$.

Among INFOODS databases, only the starchy roots & tubers category was accessible to phytochemicals such as beta-caroten, lutein, carotenoids, caffeic acid, chlorogenic acid, flavonoids and flavonols. Other categories of food did not contain sufficient dataset of phytochemicals for the analysis. The results may therefore be limited to investigated range mainly with major components (proteins, carbohydrates, fats) and minerals.

Table 4. *Qualitative comparison of actual confidence level of mean and variance with in cultura/natura diets intake distributions.*

Database	$ACL(\mu)$	$ACL(\sigma)$
INFOODS Starchy roots & tubers	$\#c=17(7)$ $ACL(\mu_N) > ACL(\mu_C)$	$\#c=17(7)$ $ACL(\sigma_N) > ACL(\sigma_C)$
INFOODS Nuts & seeds	$\#c=5$ $ACL(\mu_N) \sim ACL(\mu_C)$	$\#c=5$ $ACL(\sigma_N) \sim ACL(\sigma_C)$
INFOODS Vegetables	$\#c=8$ $ACL(\mu_N) \sim ACL(\mu_C)$	$\#c=8$ $ACL(\sigma_N) < ACL(\sigma_C)$
INFOODS Fruits	$\#c=16$ $ACL(\mu_N) > ACL(\mu_C)$	$\#c=16$ $ACL(\sigma_N) \subset ACL(\sigma_C)$
INFOODS Meat	$\#c=11$ $ACL(\mu_N) > ACL(\mu_C)$	$\#c=11$ $ACL(\sigma_N) \subset ACL(\sigma_C)$

Database	$ACL(\mu)$	$ACL(\sigma)$
Synecoculture Na	$\#c=1$, Ds=-0.123 $ACL(\mu_N) < ACL(\mu_C)$	$\#c=1$, Ds=-0.489 $ACL(\sigma_N) < ACL(\sigma_C)$
Synecoculture K	$\#c=1$, Ds=0.4835 $ACL(\mu_N) > ACL(\mu_C)$	$\#c=1$, Ds=1.051 $ACL(\sigma_N) > ACL(\sigma_C)$
Synecoculture Mg	$\#c=1$, Ds=0.4526 $ACL(\mu_N) > ACL(\mu_C)$	$\#c=1$, Ds=1.1283 $ACL(\sigma_N) > ACL(\sigma_C)$
Synecoculture Ca	$\#c=1$, Ds=0.1896 $ACL(\mu_N) > ACL(\mu_C)$	$\#c=1$, Ds=-0.176 $ACL(\sigma_N) < ACL(\sigma_C)$
Synecoculture Na, K, Mg, Ca	$\#c=4$, Ds=0.2508 $ACL(\mu_N) > ACL(\mu_C)$	$\#c=4$, Ds=0.3784 $ACL(\sigma_N) > ACL(\sigma_C)$

Figure 2. *Actual confidence level of in cultura/natura diets intake distribution mean value from INFOODS database (starchy roots & tubers). The color represents the component name, and the marker shape specifies the l_1 value. Top: in cultura diet. Bottom: in natura diet.*
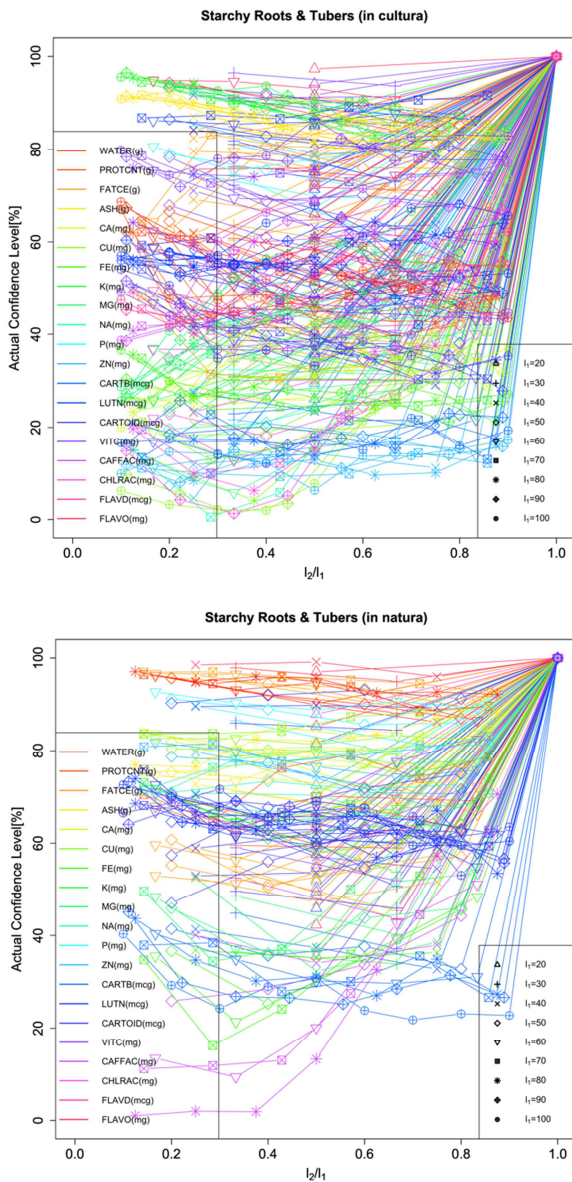
Figure 3. *Actual confidence level of in cultura/natura diets intake distribution variance from INFOODS database (starchy roots & tubers). The color represents the component name, and the marker shape specifies the l_1 value. Top: in cultura diet. Bottom: in natura diet.*
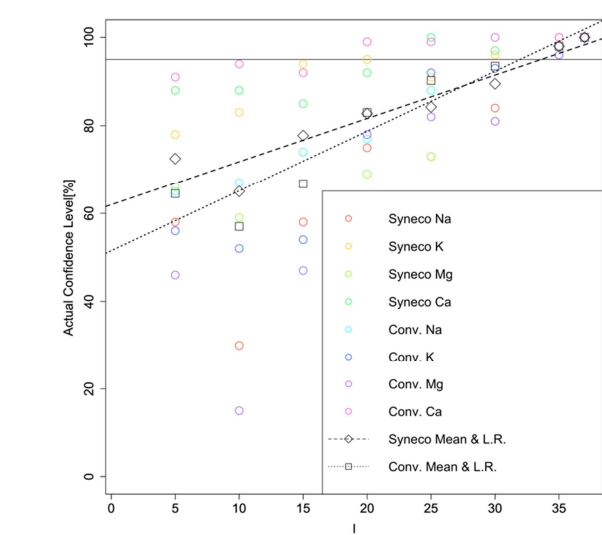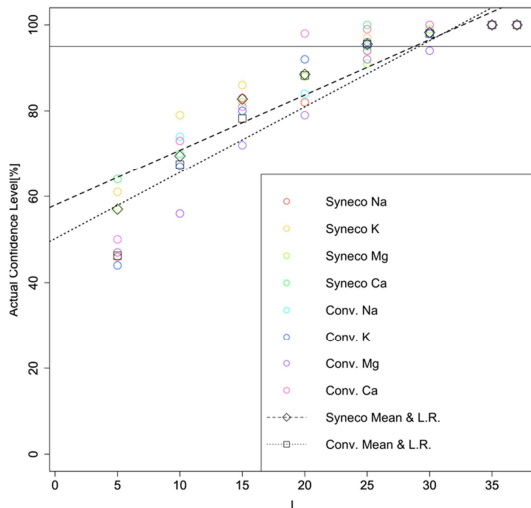
Figure 4. *Actual confidence level of in cultura/natura diets intake distribution variance from Synecoculture database and corresponding conventional data. Black plots are the mean plots of 4 minerals with liner regression. Top: mean value. Bottom: Variance.*
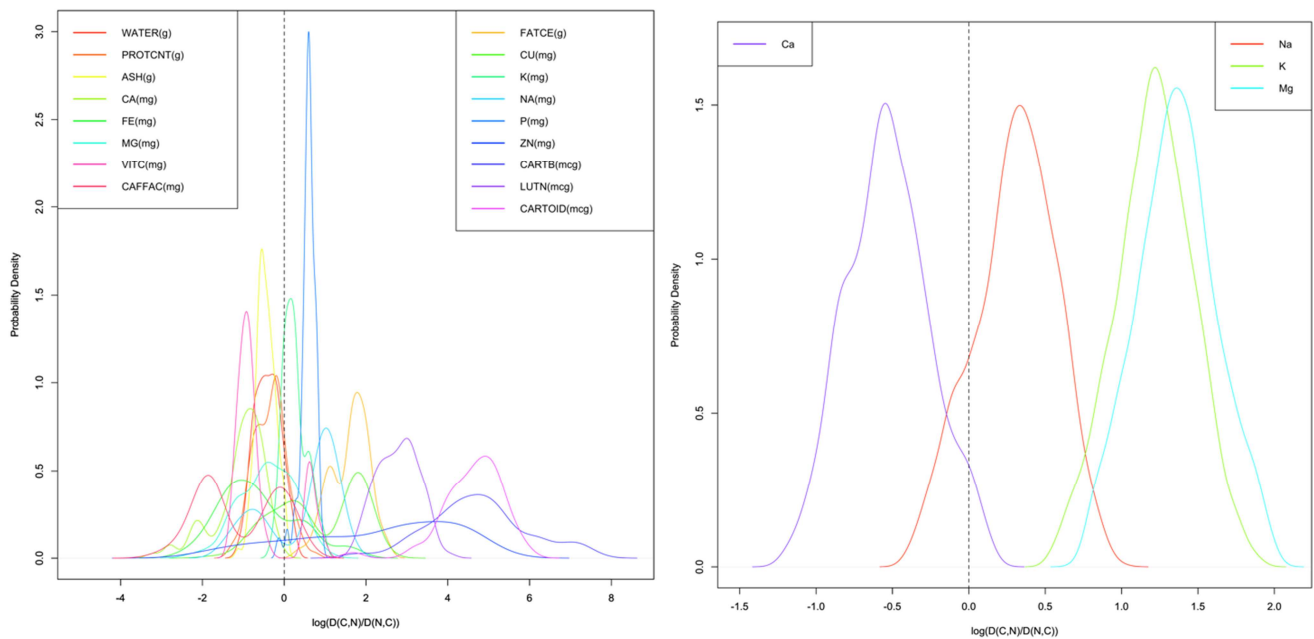
4.4. KL-divergence Ratio Between in cultura and in natura Diet Intake Distributions

The distribution of KL-divergence ratio $D(C,N)/D(N,C)$ was investigated. The results are partially shown in Figure 5. Since KL-divergence is an asymmetric measure between 2 distributions, the ratio represents the characteristic of *in cultura/natura* in reference to each other. The summary of the $\log(D(C,N)/D(N,C))$ is listed in Tables. 5. #c+ and #c- represents the number of components with the value $D(C,N) > D(N,C)$ and $D(C,N) < D(N,C)$, respectively. The number of phytochemicals is shown with parentheses. $mean[\log(D(C,N)/D(N,C))]$ was taken as the mean value for all components. Although specific food components could not commonly be separated between the groups #c+ and #c-, the overall tendency inclines to $D(C,N) > D(N,C)$.

This means that *in cultura* diet is more distant to *in natura* diet than the other way round, when the intake distribution is considered as information input to metabolism. This result could be applied in considering the coevolution of human metabolism and the development of agriculture, which can be formalized as a long-term transition from *in natura* to *in cultura* diet during 10,000-12,000 years of agricultural history [20].

Table 5. *Mean values of KL-divergence ratio $\log(D(C,N)/D(N,C))$.*

Database	#c+	#c-	$mean[\log(D(C,N)/D(N,C))]$
INFOODS Starchy roots & tubers	9(3)	8(1)	1.0517804059 $D(C,N) > D(N,C)$
INFOODS Nuts & seeds	2	3	0.15762788 $D(C,N) > D(N,C)$
INFOODS Vegetables	1	7	-1.645527125 $D(C,N) < D(N,C)$
INFOODS Fruits	10	6	0.3252658806 $D(C,N) > D(N,C)$
INFOODS Meat	8	3	0.7351448273 $D(C,N) > D(N,C)$
Syneco Na, K, Mg, Ca	3	1	0.620709825 $D(C,N) > D(N,C)$

Figure 5. *Density distributions of KL-divergence ratio $log(D(C,N)/D(N,C))$. Left legend lists the components with $D(C,N) < D(N,C)$ in average, while right legend with $D(C,N) > D(N,C)$. Left: INFOODS starchy roots & tubers. Right: Synecoculture.*
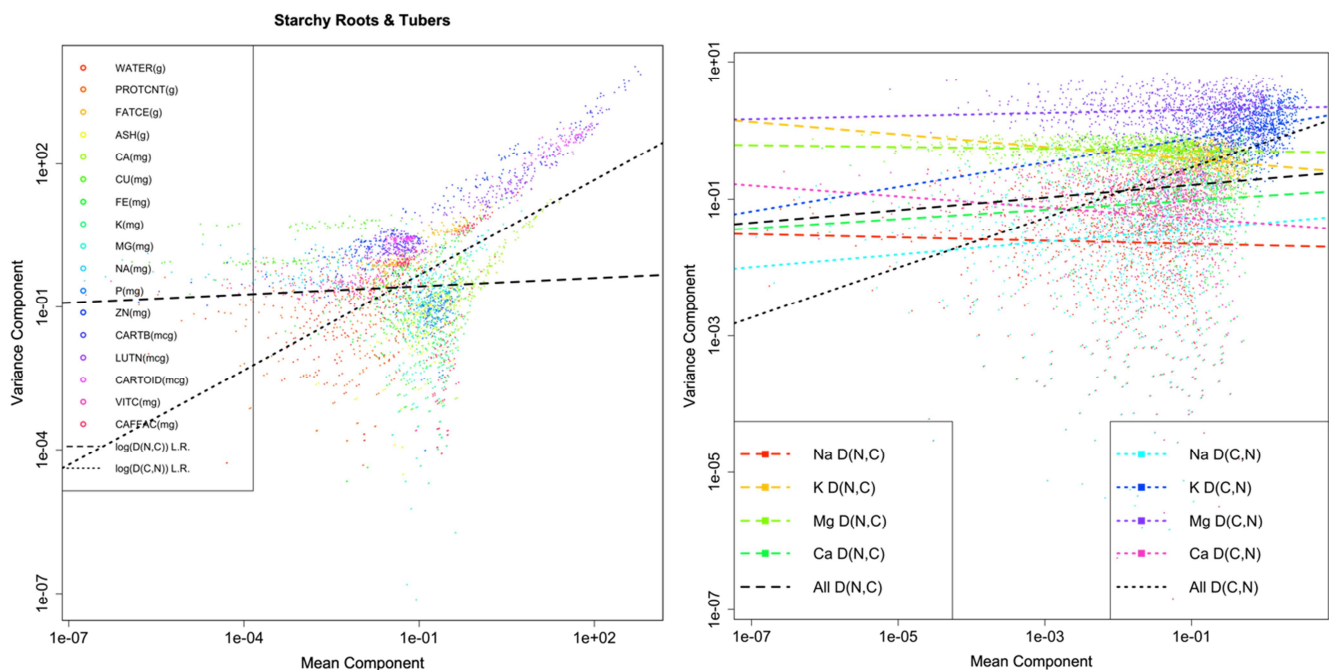
Figure 6. *Mean vs. variance components of KL-divergence $D(C,N)$ and $D(N,C)$. Linear regression was performed on double-logarithmic scale. Left: INFOODS starchy roots & tubers. Right: Synecoculture.*

4.5. Mean and Variance Components of KL-divergence Between in cultura and in natura Diet Intake Distributions

We investigated the relation between the mean components $D(C,CN)$, $D(N,NC)$ and variance components $D(CN,N)$, $D(NC,C)$ in the equations (3) (4). Significant linear regressions were obtained in double logarithmic scale. The slope $S(C,N)$ of linear regression between $log(D(C,CN))$ and $log(D(N,NC))$, and the slope $S(N,C)$ of linear regression between $log(D(N,NC))$ and $log(D(NC,C))$ were calculated with p<0.05 (except for Syneco Na $S(C,N)$ with p=0.112 and Syneco Mg $S(N,C)$ with p=0.312).

The results were partially depicted in Figures 6 and summarized in Table 6. The function $sgn(\ )$ was used to judge the value of the slope between positive (+) and negative (-) value. In total, the relation $S(C,N) > 0$ and

$S(C, N) > S(N, C)$ were dominant in both INFOODS and Synecoculture database. This implies that the variance components of KL-divergence $D(C, N)$ is positively related with the mean components, and the degree is higher in $D(C, N)$ than $D(N, C)$. This fact addresses a novel subject in food component analysis, to target the high order statistics as distinctive factor between *in cultura/natura* diets and investigate possible physiological effect it exerts on our metabolism. More concretely, the effect of variance component may relate to the foraging behavior, that are originally inseparable with diet for wild animals and hunter-gatherers [49].

Table 6. *Slopes of linear regression between mean and variance components of KL-divergence $D(C, N)$ and $D(N, C)$ in double logarithmic scale.*

Database	$sgn(S(C,N))$, $sgn(S(N,C))$, $S(C,N) - S(N,C)$
INFOODS Starchy roots & tubers	+, +, 0.61132 $S(C, N) > S(N, C)$
INFOODS Nuts & seeds	+, +, 0.47342 $S(C, N) > S(N, C)$
INFOODS Vegetables	+, +, -0.08483 $S(C, N) < S(N, C)$
INFOODS Fruits	+, +, 0.06162 $S(C, N) > S(N, C)$
INFOODS Meat	+, +, 0.31234 $S(C, N) > S(N, C)$
Synecoculture Na	+, -, 0.11128 $S(C, N) > S(N, C)$
Synecoculture K	+, -, 0.26542 $S(C, N) > S(N, C)$
Synecoculture Mg	+, -, 0.043176 $S(C, N) > S(N, C)$
Synecoculture Ca	-, +, -0.16262 $S(C, N) < S(N, C)$
Synecoculture Na, K, Mg, Ca	+, +, 0.27282 $S(C, N) > S(N, C)$

5. Conclusion

We defined the *in cultura/natura* diets based on the culture condition in view of defining a category of sustainable diet *in natura*, taking both health and environmental factors into account.

Through statistical invariance analysis, common invariant features of INFOODS and Synecoculture database were investigated. With respect to the *in cultura/natura* distinction, some results could derive meaningful insights that this distinction differentiates invariant features of intake distributions, therefore the component statistics could serve as markers to support the definition of *in natura* diet in comparison to *in cultura* diet.

Among the invariant features that distinguish between *in cultura/natura* diets, analysis of R_σ indicated that culture condition affected more than genetic profile when component distributions were projected to intake distributions.

The result of $R_\sigma(\sigma_C > \sigma_N)$ coincides with the result of actual confidence level $ACL(\mu_N) > ACL(\mu_C)$, with the view that *in cultura* diet is more diversely differentiated by culture condition and difficult to estimate the population distribution

from limited samples. Inversely, *in natura* diet might be more consistent in terms of intake distribution even the food system is limited.

Although actual confidence level showed some complexity reduction in *in natura* diet, the convergence of lower bound within $l_1 \leq 100$ sometimes exceed linear relation with l_2/l_1 in INFOODS database (Figures 2 and 3). This suggests that empirical study with metabolomics is still encouraged to augment plausibility of the invariance analysis in larger population.

Especially, further incorporation of phytochemicals will introduce more link to biodiversity, environmental quality as well as health protective factors. This would ameliorate the quality of statistical measures as integrated markers of health and environment.

Another recommendation that the results can provide is the recognition of high order statistics as significant parameter. In most of food science, and even in cellular biology with highly controlled nutrition condition in cell culture, little has been investigated on the effect of nutrient fluctuation to biological activity. While our metabolism has been continuously exposed to intermittent intake of foods through foraging behavior, current studies only focus on the stable input of nutrition with regular frequency. With this respect, current state of knowledge on the health effect of food components seems to be biased by the stable input of *in cultura* diet. If we look back at the Paleolithic diet to recover lost aspects of our health [20], *in natura* diet in relation to the variance components might be a possible alternative for the investigation. As the dynamic aspect of our metabolic response is considered to be essential in the adaptive functioning [50], large fluctuation response of homeodynamics will be a key issue for further investigation of *in natura* diet.

Finally, how the asymmetry of $D(C, N)/D(N, C)$ between *in cultura/natura* diets affects our metabolism in physiological and evolutionary scale remains to be investigated. Theoretically, environmental drift would be imprinted in genetic profiles through the selection within phenotypic plasticity [51]. How adaptation to *in cultura* diet that led to achieve today's civilization has been affecting our health? This is a question that could relate to the metabolic consequence of agricultural development associated with the self-domestication of humans by diet transition. Through which, we may clarify the trade-off between *in cultura* and *in natura* diets with both health and environmental perspectives.
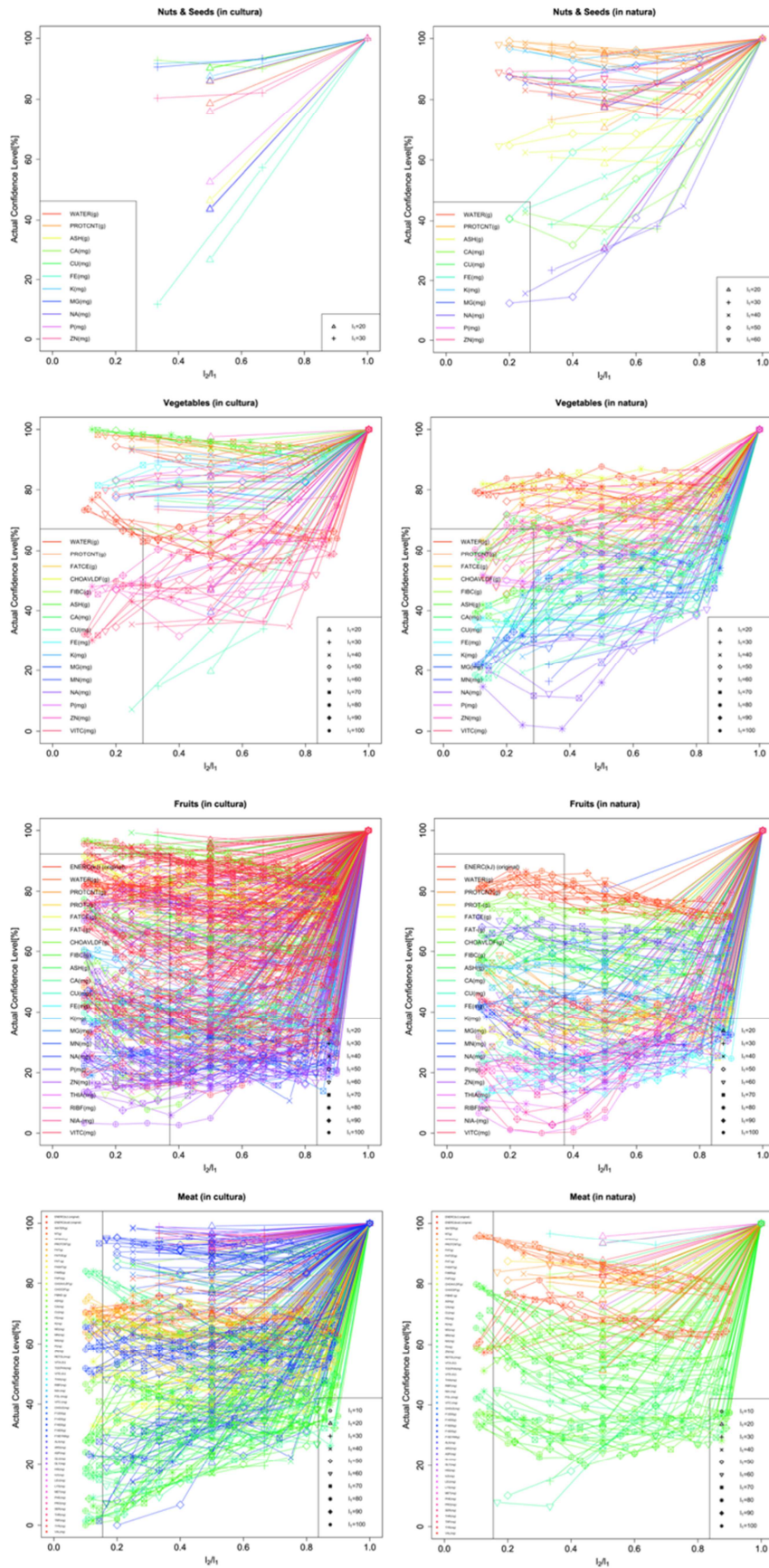
Acknowledgements

Appendix

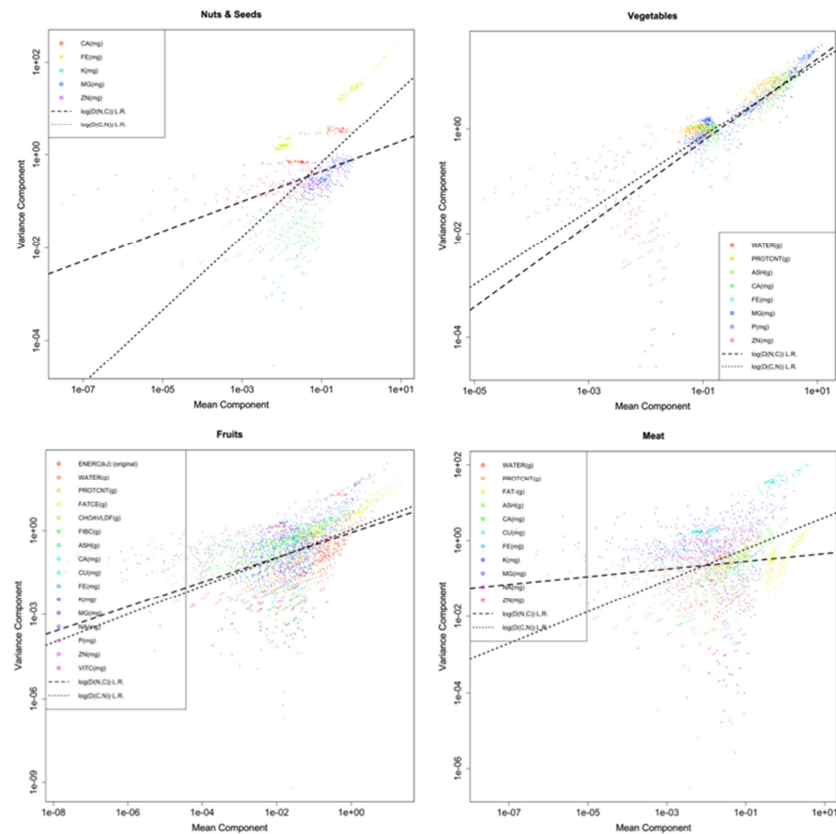Supplement figures for Figures 2, 3, 5, 6 are shown below.
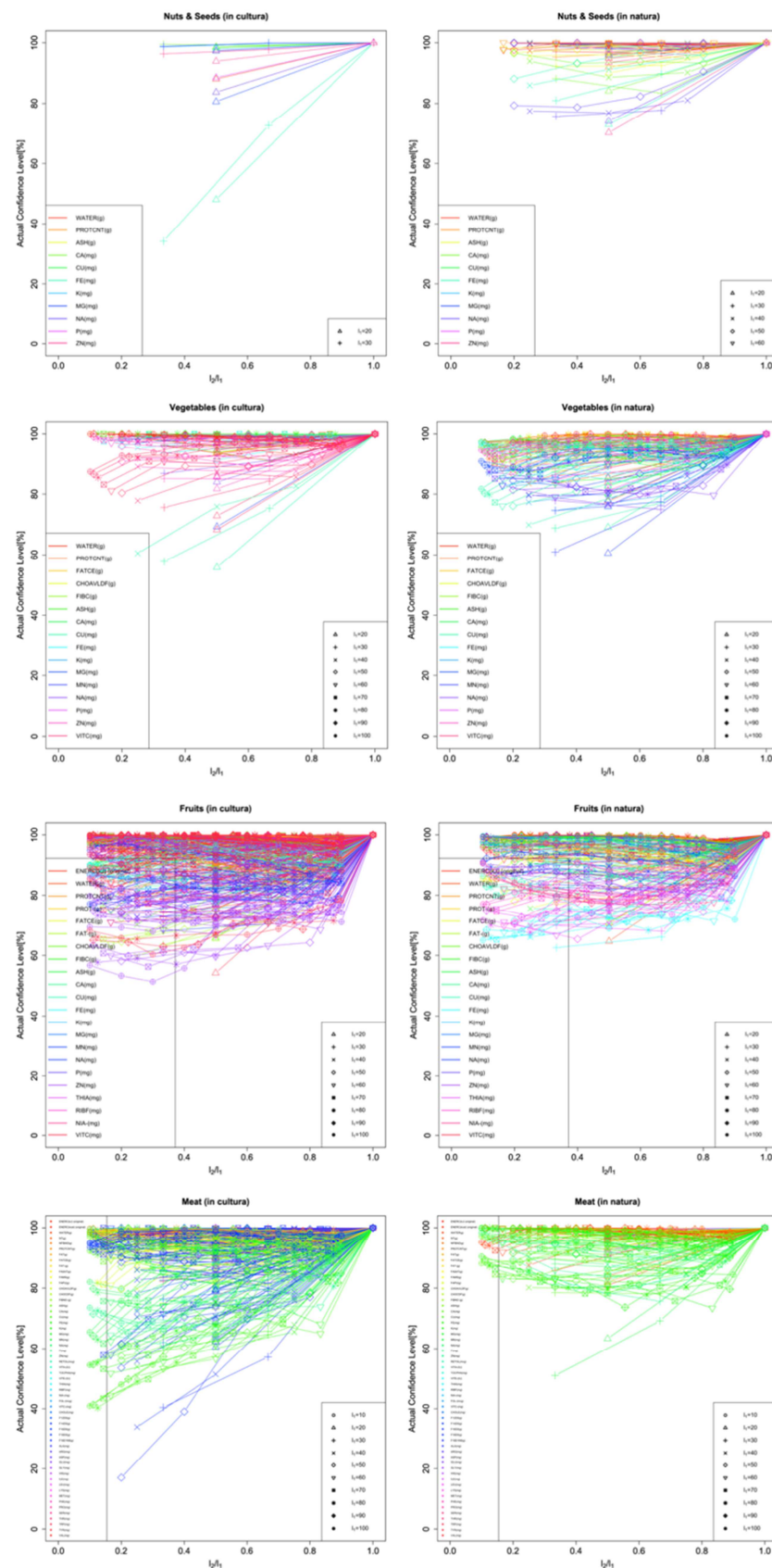
Figure 7. *Supplement Figures for Figure 2. Actual confidence level of in cultura/natura diets intake distribution mean value from INFOODS database. From top to bottom, nuts & seeds, vegetables, fruits, meat. The color represents the component name, and the marker shape specifies the l_1 value. Left figures: in cultura diet. Right figures: in natura diet.*

Figure 8. *Supplement Figures for Figure 3. Actual confidence level of in cultura/natura diets intake distribution variance from INFOODS database. From top to bottom, nuts & seeds, vegetables, fruits, meat. The color represents the component name, and the marker shape specifies the* l_1 *value. Left figures: in cultura diet. Right figures: in natura diet.*
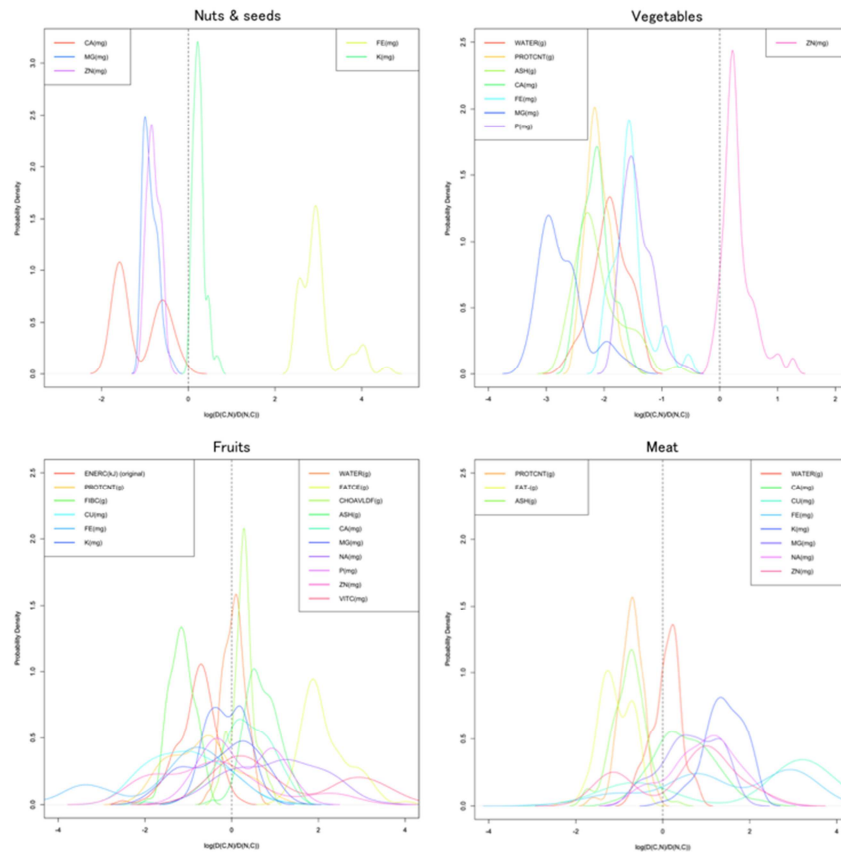
Figure 9. Supplement Figures for Figure 5. Density distributions of KL-divergence ratio $log(D(C,N)/D(N,C))$. Left legend lists the components with $D(C,N) < D(N,C)$ in average, while right legend with $D(C,N) > D(N,C)$. From top left to bottom right: INFOODS nuts & seeds, vegetables, fruits, meat.

Figure 10. Supplement Figures for Figure 6. Mean vs. variance components of KL-divergence $D(C,N)$ and $D(N,C)$. Linear regression was performed on double-logarithmic scale. From top left to bottom right: INFOODS nuts & seeds, vegetables, fruits, meat.

References

[1] David Tilman & Michael Clark, "Global diets link environmental sustainability and human health" Nature, vol. 515, 27 Nov 2014, pp. 518-522.

[2] Kung (2008) Kung et al., "Deaths : Final Data for 2005" National Vital Statistics Reports, NVSS, CDC, vol 56, Number 10.

[3] T. Furukawa and N. Izumida (2004) "Japanese Healthcare Expenditures in a Comparative Context", The Japanese Journal of Social Security Policy, Vol 3, No2.

[4] WHO (2003) WHO Technical Report Series 916, "Diet, Nutrition and The Prevention of Chronic Disease" Geneva 2003.

[5] Harvard (2009) Harvard School of Public Health, "The Global Burden of Non-communicable Diseases", A Report by the World Economic Forum and the Harvard School of Public Health, 2011.

[6] Barnosky (2012) A. D. Barnosky et al., "Approaching a state shift in Earth's biosphere", Nature, Vol 486, pp.52-58, 7.

[7] Scenarios for Global Biodiversity in the 21st Century, Henrique M. Pereira, et al., Science 330, 1496 (2010).

[8] IAASTD (2009) "International assessment of agricultural knowledge, science and technology for development (IAASTD): synthesis report with executive summary: a synthesis of the global and sub-global IAASTD reports." http://www.unep.org/dewa/agassessment/reports/IAASTD/EN /Agriculture at a Crossroads_Synthesis Report (English).pdf

[9] A. Catharine Ross et al., Ed. "Modern Nutrition in Health and Disease" Lippincott Williams & Wilkins, 1992 pages, (2012/12/22).

[10] Craig WJ, "Phytochemicals: guardians of our health." J Am Diet Assoc. 1997 Oct; 97 (10 Suppl 2): S199-204.

[11] British Nutrition Foundation, "Plants: Diet and Health" Wiley-Blackwell, (2008/6/9).

[12] World Health Organization, "Trace elements in human nutrition and health", 361 pages, 1996 http://www.who.int/nutrition/publications/micronutrients/9241 561734/en/

[13] Alejandro Cifuentes (Ed) (2013), "Foodomics: Advanced Mass Spectrometry in Modern Food Science and Nutrition," Wiley Series on Mass Spectrometry (Book 52), Wiley.

[14] Timothy Johns, "The Origin of Human Diets & Medicines" Univ of Arizona Pr (1996/09) 356 pages.

[15] Hiroaki Kitano "A robustness-based approach to systems-oriented drug design" Nature Reviews Drug Discovery 5, 202–210 (1 March 2007) | doi: 10.1038/nrd2195.

[16] Liu RH, "Health benefits of fruit and vegetables are from additive and synergistic combinations of phytochemicals." Am J Clin Nutr. 2003 Sep; 78 (3 Suppl): 517S-520S.

[17] Masatoshi Funabashi, David Chavalarias, Jean-Philippe Cointet, "Order-Wise Correlation Dynamics in Text Data" Complex Networks, Volume 207 of the series Studies in Computational Intelligence pp 161-171.

[18] R. Hursel et al., "The effects of catechin rich teas and caffeine on energy expenditure and fat oxidation: a meta-analysis" obesity reviews, 2011 International Association for the Study of Obesity 12, e573–e581.

[19] William Rumpler et al., "Oolong Tea Increases Metabolic Rate and Fat Oxidation in Men" The Journal of Nutrition, 2001 American Society for Nutritional Sciences, pp.2848-2852.

[20] Artemis P. Simopoulos ed., "Metabolic Consequences of Changing Dietary Patterns" S Karger Pub (1996/07) 226 pages.

[21] Dong Wang et al., "Dietary patterns and hypertension among Chinese adults: a nationally representative cross-sectional study" BMC Public Health 2011, 11: 925.

[22] Shuqian Liu et al., "Prevalence of Diabetes and Impaired Fasting Glucose in Chinese Adults, China National Nutrition and Health Survey, 2002" Preventing chronic disease, VOLUME 8: NO. 1, A13 JANUARY 2011.

[23] Yan-Ping Li et al., "Disease Risks of Childhood Obesity in China" BIOMEDICAL AND ENVIRONMENTAL SCIENCES 18, 401-410 (2005).

[24] Kuhnlein (2013) H. V. Kuhnlein et al.(Ed) "Indigenous Peoples' Food systems & well-being", Food and Agriculture Organization of the United Nations, Center for Indigenous Peoples' Nutrition and Environment, Rome, 2013.

[25] Carol J. Pierce Colfer ed., "Human Health and Forests: A Global Overview of Issues, Practice and Policy" Routledge, (2012/5/4) 400 pages.

[26] Vaclav Smil and Kazuhiko Kobayashi, "Japan's Dietary Transition and Its Impacts" The MIT Press (2012/8/24) 244 pages.

[27] Artemis P. Simopoulos ed., "Plants in Human Health and Nutrition Policy" S Karger Pub (2003/04) 138 pages.

[28] M. Leonti et al., "Wild Gathered Food Plants in the European Mediterranean: A Comparative Analysis" Economic Botany, 60(2), 2006, pp. 130-142.

[29] A. Trichopoulou "Nutritional composition and flavonoid content of edible wild greens and green pies: a potential rich source of antioxidant nutrients in the Mediterranean diet" Food Chemistry 70 (2000) 319-323.

[30] Djoulde Daman Roger et al., "Nutritional properties of "Bush Meals" from North Cameroon's Biodiversity" Pelagia Research Library, Advances in Applied Science Research, 2012, 3 (3): 1482-1493.

[31] Aurélie Cotillard et al., "Dietary intervention impact on gut microbial gene richness" Nature 500, 585–588 (29 August 2013).

[32] Stephanie L. Schnorr et al., "Gut microbiome of the Hadza hunter-gatherers" Nature Communications 5, 3654, 15 April 2014.

[33] M. Funabashi et al., "Foundation of CS-DC e-laboratory: Open Systems Exploration for Ecosystems Leveraging" Proceedings of Complex Systems Digital Campus World e-conference, 2015.

[34] Slatkin, M. (2009). "Epigenetic Inheritance and the Missing Heritability Problem". Genetics 182 (3): 845–850.

[35] Manuel J. Reigosa et al. ed., "Allelopathy: A Physiological Process with Ecological Implications" Springer Netherlands, (2006/2/28) 638 pages.

[36] M. Funabashi, "In Natura Diet: Statistical Invariance Analysis and Synecoculture Experiment" proceedings of the 11th International Food Data Conference, Hyderabad, India, 2015.

[37] Max Weber "Die 'Objektivitat' sozialwissenschaftlicher und sozialpolitischer Erkenntnis" GRIN Verlag, (2009/8/24).

[38] Putman (1984) R. J. Putman and S. D. Wratten, "Principles of Ecology" University of California Press.

[39] Quintana-Murci, L. et al., (2007). "Immunology in natura: Clinical, epidemiological and evolutionary genetics of infectious diseases". Nature Immunology 8 (11): 1165–1171.

[40] M. Funabashi, "IT-Mediated Development of Sustainable Agriculture Systems–Toward a Data-Driven Citizen Science" Journal of Information Technology and Application in Education (JITAE), 2013, 2(4), 179-182.

[41] K. Yoshida, Y. Shimizu-Yoshida and M. Funabashi, "Functional and Compositional Characteristics of Conventionally, Organically and Naturally Grown Cabbage and Carrots." Proceedings of the 4th International Conference on Food Studies, Prato, Italy, on October 20-21, 2014.

[42] K. Yoshida and M. Funabashi, "Taste Analysis on Conventionally, Organically and Naturally Grown Cabbage" proceedings of the 1st IMEKOFOODS conference in Rome, Italy, on 12-15 October 2014.

[43] K. Yoshida, "Categorical Mapping of Food Compositions" proceedings of the 11th International Food Data Conference, Hyderabad, India, 2015.

[44] M. Funabashi, "Comparison of major minerals uptake between conventional and non-tillage/non-fertilizer/non-chemical vegetables in Japan" proceedings of the Nutrition Society, 72 (OCE5), E313 (2013).

[45] S. Kullback, R. Leibler, "On information and sufficiency", Annals of Mathematical Statistics, vol. 22, 1951, p. 79-86.

[46] S. Amari and H. Nagaoka, "Method of information geometry" in Translations of Mathematical Monograph, Vol. 191 (Oxford University Press, Oxford, UK, 2000), 206 pp.

[47] FAO/INFOODS(2013). FAO/INFOODS Food Composition Database for Biodiversity Version 2.1-BioFoodComp2.1. FAO, Rome.

[48] Ministry of Education, Culture, Sports, and Technology-Japan (2010) Standard Tables of Food Composition in Japan.

[49] R. Hughes ed., "Diet Selection: An Interdisciplinary Approach to Foraging Behaviour" Wiley-Blackwell (1993/6/7) 232 pages.

[50] F. E. Yates "Order and Complexity in Dynamical Systems: Homeodynamics as a Generalized Mechanics for Biology" Mathl. Comput. Modelling Vol. 19, No. 6-8, pp. 49-74, 1994.

[51] Kaneko K, Furusawa C "An evolutionary relationship between genetic variation and phenotypic fluctuation" J Theor Biol. 2006 May 7; 240(1): 78-86.