

微量元素含有量を用いた野菜類の分散型産地判別システム

佐藤 永欣[†] 上原 稔^{††} 玉岡 迅[†]
下村 講一郎[†] 山本 浩文[†] 上條 賢一[†]

食品の安全を確保するためトレーサビリティシステムが導入されているが、食品に付加した ID を追跡するためすり替え等が可能である。そこで、我々は野菜等の微量元素含有量の産地による違いを利用した産地判別システムの開発を行っている。土壌の微量元素含有量は土地によって異なり、植物の生育にともなって吸収される微量元素の量も土地や生育状況によって異なるため、産地判別に利用できる。本論文では農芸化学者により通常使われる方法よりも計算機に向けた産地判別手法を述べた後、農家からの出荷時に微量元素含有量を分析・蓄積し、産地偽装が疑われる野菜類の微量元素含有量と比較することで産地を判別するシステムについて述べる。

Distributed Geographical Origin Identification System for Green-groceries by Trace Element Compositions

NOBUYOSHI SATO,[†] MINORU UEHARA,^{††} JIN TAMAOKA,[†]
KOICHIRO SHIMOMURA,[†] HIROBUMI YAMAMOTO[†]
and KENICHI KAMIJO[†]

Food traceability systems are used to ensure food safety. However, switching foods is possible since they chases only IDs attached onto foods. Therefore, we have been developing a system to identify geographical origin of vegetables using their trace element compositions. Since trace, very small quantity element compositions in the soil differ from places, it differ that trace element compositions absorbed by plants along with their growth. So, they can be used to identify geographical origin. In this paper, we describe a computer friendly identification method rather than used by chemists, and we describe our system that identifies by comparison of two trace element compositions, measured in a farm and doubtful one's.

1. はじめに

ここ 10 年ほど、BSE や輸入野菜の残留農薬問題のような食品の安全性に関する問題が多発した。また、農産地のブランド価値を重視する傾向が強まり、安価な農産物をブランド産地の農産物と偽って販売する事件が発生した。2000 年の JAS 法改正により農産物の原産地表示の適正化が義務づけられたが、それ以降も年に数回、産地偽装事件が発覚している。

食品トレーサビリティを導入し、食品に対する信頼を取り戻そうという動きがある^{1),2)}。バーコード、RFID タグ等の形で食品のパッケージ等に ID を付加し、追跡することで、特定の食品の川上への追跡（トレース

バック、トレーシング）と川下への追跡（トレースフォワード、トラッキング）が実現される。しかし、食品そのものではなく、付加された ID を追跡していることに注意する必要がある。すなわち、パッケージや ID のすり替え、偽造等の可能性を排除できない。

一方、遺伝子分析や微量元素分析といった、農産物そのものの違いを利用して産地偽装を防止する試みもある。遺伝子分析は品種の違いを判別できるが産地は識別できない。微量元素分析は土壌の微量元素含有量の分布の地理的な違いを利用する。すなわち、植物は生育時に土壌中の微量元素を吸収するが、土壌中の微量元素含有量が異なれば、植物に吸収された微量元素の量も異なるため、野菜類が育った地域の異同を識別できる。また、土壌の微量元素含有量は施肥によっても変動すると考えられるため、農場や圃場のレベルで産地を判別できる可能性がある。

我々は農産物の微量元素含有量の違いを利用した産地判別システムを提案してきた³⁾⁻⁵⁾。本システムは

[†] 東洋大学植物機能研究センター
Plant Regulation Research Center, Toyo University

^{††} 東洋大学情報工学科
Department of Information and Computer Sciences,
Toyo University

農産物の微量元素含有量を農場からの出荷時にデータベースに蓄積し、流通経路上の農産物の微量元素含有量との相関係数を計算し、比較することで農産物の産地を判別する。農芸化学者による産地判別手法と違い、情報検索のアイデアに基づく本手法は、産地判別自体の曖昧さと人手の介入を排除し、規準を明確化した点で計算機による実現に向いている。

本論文の構成は以下のとおりである。2章では食品トレーサビリティシステムと微量元素による農産物の産地判別に関する関連研究について述べる。3章では相関係数を用いた産地判別手法を述べた後、実際のデータを用いて産地判別が可能か検討し、シミュレーションにより判別精度や誤判定の可能性を議論する。4章では本システムの設計と動作を述べ、プロトタイプによる産地判別に要する時間等の評価を示す。最後にまとめを述べる。

2. 関連研究

2.1 食品トレーサビリティ

青果類、牛肉を中心いくつかの食品トレーサビリティシステムが実験、または運用段階にある²⁾。牛肉のトレーサビリティは、BSE 問題を受けて制定された「牛の個体識別のための情報の管理および伝達に関する特別措置法」に基づいて運用されている。国産牛の個体識別と育成履歴の確認が可能である。平成 17 年度から農林水産省「ユビキタス食の安全・安心システム開発事業」²⁾として、いくつかのプロジェクトが補助金を受けている。

T-Engine Forum⁶⁾はIDとしてucode⁷⁾を利用し、生産から店舗までの各段階での追跡を行う実証実験を牛乳、豆腐、米、ジュース、青果で実施した。ucodeはRFID タグに記録されるが、バーコードも併用されている。全国農業協同組合中央会による全農安心システムでは、米や肉類のトレーサビリティをバーコードのラベルを貼布して実現している。生産段階から最終消費者への流通のすべての段階でのトレーサビリティの実現が特徴であり、加工工場等でも認証を受ける等して不正防止に留意している。

食品トレーサビリティは、流通、輸送用パッケージにバーコード、RFID タグ等機械可読な形でIDを付加することで実現される。生産、輸送、加工等の履歴はデータベースに格納され、ネットワーク越しに参照される。書き込み可能なRFID タグを利用すると履歴をRFID タグに記録できるが、書き込み可能なRFID タグは高価なため利用されていない。

農家から出荷された農産物は、集荷、選別、輸送、

加工、小分け等の段階を経て小売店に並ぶ。この間に複数の流通業者が関与する。集荷、選別の方法は農産物ごとに違い、共同選果場等での選別の方法によっては生産農家を特定できないこともある。加工、小分けの内容によっても追跡できない場合がある。たとえば、野菜を加工してサラダを作ると、サラダに対して新たにIDが割り当てられ、追跡が途切れる。コンテナや段ボール箱で輸送された農産物を小売用に小分け、包装する場合、IDが再割当てされることもある。一般に、IDの再割当てをともなう流通段階では、すり替え、虚偽報告等により偽装が可能である。

2.2 農産物の産地判別

農産物等の植物体内、植物製品の微量元素含有量を用いた産地判別の試みは、ワイン^{8),9)}、コーヒー¹⁰⁾、茶¹¹⁾等の産地や生産年が重要な製品のほかに、オレンジジュース¹²⁾、じゃがいも¹³⁾、長ネギ¹⁴⁾、ニンニク¹⁵⁾、タマネギ¹⁶⁾、玄米¹⁷⁾等で農芸化学分野の研究者を中心に行われてきた。このほかに、安定同位体比¹⁸⁾や有機物¹⁹⁾を用いた研究がある。

これらの研究では、測定・実験手法(実験系、システムと呼ばれる)を確立したうえで、主にクラスタ分析や散布図を用いて微量元素含有量等が産地により違うことを示している。一般に、農芸化学分野では実験系と統計処理後のデータが議論の対象であり、統計処理手法や産地判別のための情報システムは興味の対象ではない。また、産地判別に用いる元素の選択基準が明確でない、統計処理に試行錯誤をともなう、産地を判別できたと判断する基準が明確でない等の問題がある。このように、計算機による実現は考慮されておらず、人手の介入を前提とするプロセスもある。

サンプル収集と分析の時間的・経済的コストの問題により、広範囲から継続的にサンプルを収集し、農産物の微量元素含有量の地理的・時間的分布を明らかにした研究は存在しない。産地により微量元素含有量が異なること、微量元素含有量の傾向は大陸、国、州県等の単位で大まかに分類できることは示されているが、農場・圃場レベルの産地判別はまったく論されていない。日本国内で行われた研究¹⁴⁾⁻¹⁷⁾は原産地表示の適正化を視野にいれているが、特定国からの輸入品と国産品の識別が可能であることを示したにすぎない。

本論文では人手の介入を排し、産地判別の規準を明確化した計算機向きの手法として、相関係数の計算による産地判別手法を述べ、細粒度の産地判別が可能となることを実際のデータとシミュレーションを用いて示す。また、情報システムとしての産地判別システムについて述べ、評価を示す。

3. 多変量解析による産地判別法

3.1 産地判別手法の概要

ここでは相関係数の計算による、曖昧さと人手の介入を排した産地判別の基本的な手法を述べる。

産地数を c 、各産地を $d = 1, 2, \dots, c$ 、 d のサンプル数を k_d 、サンプルを $i = 1, 2, \dots, k_d$ で表し、 l 種類の元素を $e = 1, 2, \dots, l$ とする。微量元素含有量 t_{die} から産地と元素ごとの平均 $\bar{t}_{d.e}$ は次のように計算できる。

$$\bar{t}_{d.e} = \frac{\sum_i t_{die}}{k_d} \tag{1}$$

これを各産地の微量元素含有量の代表値とする。

次に規準化スコア n_{de} を産地を通算した微量元素含有量の平均 $\bar{t}_{..e} = \sum_d \bar{t}_{d.e}/c$ と標準偏差 $s_{..e} = \sqrt{\sum_d (\bar{t}_{d.e} - \bar{t}_{..e})^2 / (c - 1)}$ を用いて次式で計算する。

$$n_{de} = \frac{\bar{t}_{d.e} - \bar{t}_{..e}}{s_{..e}} \tag{2}$$

一般に、ベクトル $X = (x_1 x_2 \dots x_n)$ と $Y = (y_1 y_2 \dots y_n)$ のなす角 θ_{XY} は、 \bar{x} 、 \bar{y} を X 、 Y の要素の平均として、

$$\begin{aligned} \cos \theta_{XY} &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}} \\ &= r_{XY} \end{aligned} \tag{3}$$

である。相関係数 r_{XY} は $-1 \leq r_{XY} \leq 1$ を満たし、大きいほど X と Y が類似していることを示す。

産地 d の規準化スコアのベクトルを $N_d = (n_{de_1} n_{de_2} \dots n_{de_l})$ とし、式 (3) により d 相互間の相関係数を求め、相関係数が 1 に近い適当な閾値を超

えた場合に産地が一致すると判断する。ただし、産地が異なる野菜類の微量元素含有量が偶然類似する可能性と閾値の選択に留意する必要がある。

3.2 ブロッコリーを用いた産地判別の検討

6つの産地のブロッコリーを3つの販売店から調達した。産地と販売店の組合せは Ecuador (N), U.S. (M), California (K), 埼玉 (K), 埼玉 (M), 埼玉 (A) の計6種類である。Ecuador (N) はN社の冷凍食品、ほかは生鮮である。他のカッコ内は販売店の名称の頭文字である。産地は販売店による表示を採用し、埼玉県産は市町村レベルの情報がないため異なる産地として扱う。それぞれ5サンプルずつ微量元素含有量を誘導結合プラズマ質量分析装置により測定し、各元素の含有量の違いを検定するため一元配置分散分析を行い、有意差が得られたカドミウム (Cd), クロム (Cr), 鉛 (Pb), タリウム (Tl), バナジウム (V) の6種類を判別に使用することにした。

表1に式(1)に基づいて計算した各産地の微量元素測定結果の代表値を示す。これを式(2)により規準化し、式(3)により相関係数を求めた結果を表2に示す。Calif. (K)とU.S. (M)は類似しているが一致しているとはいえない。次にCalif. (K)の各サンプルCalif. (K)1~5の測定値から同様に相関係数を計算した結果を表3に示す。Calif. (K)との相関係数がおおむね0.999程度以上と高く、一致していると判断できる一方、U.S. (M)との間では0.9程度以下であり、類似しているが一致しているとはいえない。

次に、類似して変動した場合のシミュレーションを行った。Calif. (K)の微量元素含有量 $t_{d,calif}$ の標準偏

表 1 微量元素測定結果

Table 1 Analytical result for each trace element.

産地 (d)	Cd	Cr	Pb	Co	Tl	V
Ecuador (N)	5.4691	7.6026	9.3328	12.985	31.227	14.890
U.S. (M)	12.155	7.0151	5.6796	30.817	8.1998	8.6061
Calif. (K)	15.639	7.5646	14.147	29.030	11.720	7.8218
埼玉 (K)	12.375	8.8491	12.465	16.014	38.857	12.701
埼玉 (K)	22.278	47.863	72.581	31.888	27.066	99.437
埼玉 (A)	14.059	9.0484	13.420	10.945	8.8744	12.282

μg/100g 生重量

表 2 相関行列

Table 2 Correlation coefficient matrix.

	Ecuador (N)	U.S. (M)	Calif. (K)	埼玉 (K)	埼玉 (M)	埼玉 (A)
Ecuador (N)	1					
U.S. (M)	-0.5980	1				
Calif. (K)	-0.7991	0.8908	1			
埼玉 (K)	0.7863	-0.6176	-0.5615	1		
埼玉 (M)	-0.4242	-0.1031	-0.1145	-0.7153	1	
埼玉 (A)	-0.4745	-0.4074	-0.1193	-0.3022	0.7075	1

表 3 Calif. (K) の各サンプルと各産地の相関係数

Table 3 Correlation coefficient matrix among actual Calif. (K) samples and the other districts.

	Ecuador (N)	U.S. (M)	Calif. (K)	埼玉 (K)	埼玉 (M)	埼玉 (A)
Calif. (K) 1	-0.7977	0.8963	0.9998	-0.5709	-0.1072	-0.1250
Calif. (K) 2	-0.7956	0.8850	0.9998	-0.5497	-0.1255	-0.1197
Calif. (K) 3	-0.7979	0.8948	0.9999	-0.5667	-0.1111	-0.1237
Calif. (K) 4	-0.8119	0.8673	0.9986	-0.5450	-0.1174	-0.0843
Calif. (K) 5	-0.7900	0.9085	0.9991	-0.5738	-0.1113	-0.1440

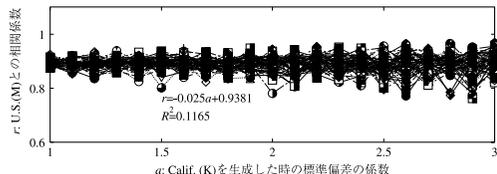


図 1 生成した Calif. (K) と実際の U.S. (M) との相関係数
Fig. 1 Correlation coefficients between samples simulated as Calif. (K) and actual U.S. (M).

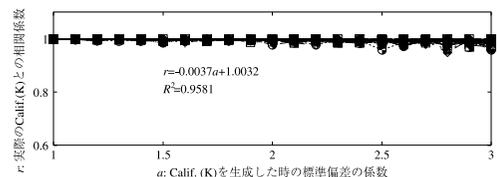


図 2 生成した Calif. (K) と実際の Calif. (K) との相関係数
Fig. 2 Correlation coefficients between samples simulated as Calif. (K) and actual Calif. (K).

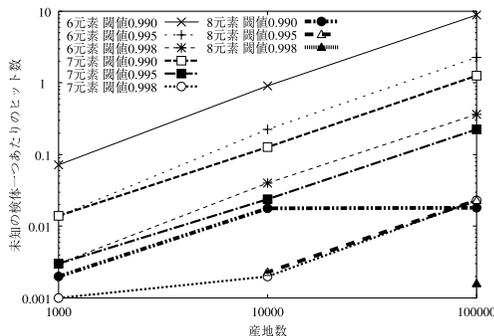


図 3 産地数, 元素数, 閾値とヒット数
Fig. 3 Relation among the number of producing districts, elements, thresholds and the number of hits.

差を $s_{d_{calif} \cdot e} = \sqrt{\sum_i (t_{d_{calif} ie} - \bar{t}_{d_{calif} \cdot e})^2 / (k_{d_{calif}} - 1)}$ とする. 平均 $\bar{t}_{d_{calif} \cdot e}$ と偏差を大きくした標準偏差 $s'_{d_{calif} \cdot e} = a s_{d_{calif} \cdot e}$ ($1.0 \leq a \leq 3.0$) を元に Calif. (K) を正規乱数を用いてシミュレートし, 同じ手順で相関係数を計算した. 規準化スコアの計算には $\bar{t}_{d_{calif} \cdot e}$ と $s_{d_{calif} \cdot e}$ を用いた. U.S. (M) との相関係数を図 1 に, Calif. (K) との相関係数を図 2 に示す. $a \leq 3.0$ ではシミュレートした Calif. (K) と実際の Calif. (K) の相関係数が U.S. (M) との相関係数よりほとんどの場合で大きく, 分散分析の結果有意差が認められた.

以上から, 各産地の野菜の微量元素含有量が相当変動したとしても産地判別は可能であり, あるサンプルに対して相関係数が 1 に近い閾値を超えた産地がそのサンプルの産地と判断してよいと考えられる.

3.3 産地数と判別精度

微量元素含有量による産地判別を実用化するためには, 微量元素含有量の広域にわたる調査と, 圃場, 施肥の状況等を考慮した細粒度の調査が必要である. しかし, これらに関する研究は存在しないため, 我々のデータに基づいてシミュレーションを行った. ここで

は閾値を超えることをヒットと呼ぶ.

まず, 産地 d , 元素 e の測定値の母平均 $\mu_{d \cdot e}$ をブロッコリーの実測データと同じ範囲になるよう一様乱数で生成し, 母標準偏差 $\sigma_{d \cdot e}$ を実測データの標準偏差の平均と標準偏差に基づき正規乱数で生成した. 次に d の i 番目のサンプルの e の測定値 t_{die} を $\mu_{d \cdot e}$ と $\sigma_{d \cdot e}$ に基づき正規乱数で生成し, その標本平均 $\bar{t}_{d \cdot e}$ と標本標準偏差 $s_{d \cdot e}$ を産地が既知のデータとした. 同様に判別対象の産地が未知のデータも生成した.

図 3 に元素数, 閾値と判別対象の産地数を変化させたときの, 産地が未知のデータ 1 件あたりのヒット数を示す. 産地数は農場または圃場の数を想定している. 日本の農家は約 200 万軒²⁰⁾ あるため, 同一作物・品種を栽培する農家は最大 10 万軒程度と仮定している. 未知, 既知のデータともに産地数は同じである. 産地数の 2 乗に比例してヒット産地数が増え, 閾値を上げればヒット産地数は減少する. 元素数を増やすと情報量が増えるためヒット産地数は急速に減少する. 図 4 に未知, 既知データともに 10 万件のときのヒット数の度数分布を示す. 元素数を増やし, 閾値を上げるとヒットしない例が増え, 産地数を増やすことが可能と考えられる. 次に, 母標準偏差を $\sigma'_{d \cdot e} = a \sigma_{d \cdot e}$ ($1.0 \leq a \leq 3.0$) とした未知データを用いて未知データ 1 件あたりのヒット数を測定したところ, a に対してほぼ一定であった.

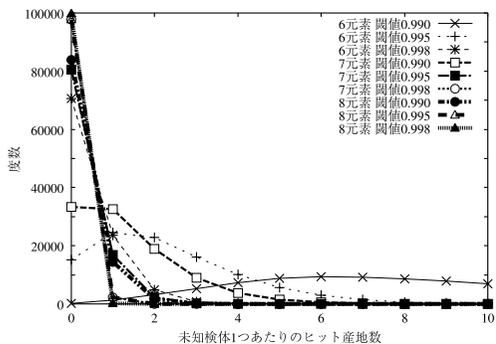


図 4 ヒット産地数の度数分布

Fig. 4 Frequency distribution for the number of hits.

以上から、提案手法は微量元素含有量の変動に対し頑健であり、農場・圃場レベルの細粒度の産地判別でも実用に耐えらると思えられる。

ところで、異なる産地の野菜類の微量元素含有量が偶然似てしまう可能性があるため、複数の産地がヒットしうる。産地数が増えればこの危険は大きくなる。閾値を上げすぎると第 1 種過誤の危険がある。また、相関係数が最大となる産地が真の産地とは限らない。さらに、現実の微量元素分布等から産地判別に用いる元素を簡単に増やせない可能性がある。

一方、トレーサビリティによる産地情報を微量元素含有量を用いて検証する場合、複数の産地が閾値を超えた場合でもトレーサビリティによる産地が含まれていれば、帰無仮説「産地が偽装されている」が棄却できなかったと考えられる。ただし、閾値を下げすぎた場合、第 2 種過誤の危険が増す。

産地を偽装して微量元素含有量による産地判別を欺くには、微量元素含有量が類似する野菜類が必要となる。しかし、微量元素含有量を人工的に調整するのは難しい。本来の産地から野菜類を調達すると産地偽装の意味がない。別の産地から調達するには、微量元素含有量の広域にわたる比較的详细なデータが必要となる。このため産地偽装は困難となる。以上から、現状ではトレーサビリティによる産地情報を微量元素含有量を用いて検証する手法が実現可能であるため、これを採用する。

4. 分散産地判別システム

4.1 設 計

提案手法は微量元素含有量を比較することで産地を判別するため、農家等からの出荷時に測定した微量元

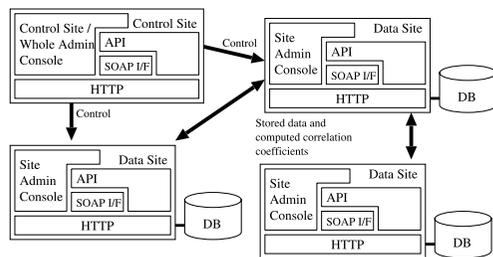


図 5 分散型産地判別システムの概要

Fig. 5 Overview of distributed geographical origin identification system.

素含有量をデータベースに蓄積する必要がある。1つ、または少数の組織が微量元素含有量のデータを全国規模で蓄積することはコスト等の問題から難しく、各産地の出荷団体等が各自、蓄積・管理するのが自然である。よって、各産地のサイトにデータベースを分散配置する。図 5 に概要を示すように本システムは以下の 2 種類のコンポーネントからなる。

Control Site (CS) 全体で 1 つだけ存在し、本システム全体の制御、Data Site (DS) が本システムに参加する際の認証、不正を行った DS の排除、不正に投入されたデータを排除するための広告等を行う。

Data Site (DS) 各産地の出荷団体等に配置され、農産物を出荷するときに測定した微量元素含有量をデータベースに蓄積する。産地判別の要求を受けたときに自身が蓄積するデータとの相関係数を計算する。

産地判別の実行は CS と DS の両方が要求できる。

まず、分散して蓄積された微量元素含有量から相関係数を分散計算する方法を述べる。DS の数を n 、 DS_i に蓄積された微量元素含有量のデータ数を m_i 、相関係数の計算対象の元素を l 種類とし、 e ($e = 1, 2, \dots, l$) で表す。 DS_i における j ($j = 1, 2, \dots, m_i$) 番目のデータの e の測定値を x_{ije} と表すと、 i における j 番目のデータは $X_{ij} = (x_{ij1} \ x_{ij2} \ \dots \ x_{ijl})$ と表せる。

相関係数の計算は規準化スコアの計算までと、相関係数自体の計算の 2 段階からなる。規準化スコアはデータの追加・削除がない限り再利用できる。規準化スコアは微量元素含有量のデータと 1 対 1 対応であるので各 DS で計算する。規準化スコアは e ごとに計算する。規準化スコアを計算する手順を以下に示す。平均の計算

(1) CS は DS_i に以下の式による各サイトにおける e の測定値の平均の計算を依頼する。

$$\bar{x}_{i \cdot e} = \frac{1}{m_i} \sum_{j=1}^{m_i} x_{ije} \tag{4}$$

第 1 種過誤：本来正しい帰無仮説を棄却してしまう誤り。
 第 2 種過誤：棄却すべき帰無仮説を棄却しない誤り。

(2) DS_i は $\bar{x}_{i.e}$ と m_i を CS に送信し, CS は e の全 DS の平均値を以下の式により計算する.

$$\bar{x}_{..e} = \frac{1}{\sum_{i=1}^n m_i} \sum_{i=1}^n m_i \bar{x}_{i.e} \quad (5)$$

標準偏差の計算

- (1) CS は $\bar{x}_{..e}$ を DS_i に配布する.
- (2) DS_i は自サイトにおける e の測定値の残差平方和 $d_{ie} = \sum_{j=1}^{m_i} (x_{ij.e} - \bar{x}_{..e})^2$ を求め, CS に送信する.
- (3) CS は e の測定値の全サイトでの標準偏差

$$s_e = \sqrt{\frac{1}{\sum_{i=1}^n m_i - 1} \sum_{i=1}^n \sum_{j=1}^{m_i} (x_{ij.e} - \bar{x}_{..e})^2}$$

$$= \sqrt{\frac{1}{\sum_{i=1}^n m_i - 1} \sum_{i=1}^n d_{ie}} \quad (6)$$

を計算する.

規準化スコアの計算

- (1) CS は s_e と $\bar{x}_{..e}$ を DS_i に配布し, 規準化スコアの計算を依頼する.
- (2) DS_i は規準化スコア $y_{ij.e} = (x_{ij.e} - \bar{x}_{..e})/s_e$ を計算し, $Y_{ij} = (y_{ij1} \ y_{ij2} \ \dots \ y_{ijl})$ とする.

次に相関係数の計算と産地判別の動作を述べる. DS が産地判別を行うときは CS を DS と読みかえる. 判別対象の規準化スコアの計算

- (1) CS は産地判別したい野菜類の微量元素測定結果 $U = (u_1 \ u_2 \ \dots \ u_l)$ の規準化を自分自身に依頼する.
- (2) CS は U からその規準化スコア $v_e = (v_e - \bar{x}_{..e})/s_e$ を計算し, $V = (v_1 \ v_2 \ \dots \ v_l)$ として返送する.

相関係数の計算

- (1) CS はすべての DS_i に V , V の産地情報と返送すべき相関係数の閾値 r_{th} を並列に送り, 相関係数の計算を依頼する.
- (2) DS_i は V と Y_{ij} の相関係数

$$r_{(Y_{ij}V)}$$

$$= \frac{\sum_{e=1}^l (v_e - \bar{v}) (y_{ij.e} - \bar{y}_{ij.})}{\sqrt{\sum_{e=1}^l (v_e - \bar{v})^2} \sqrt{\sum_{e=1}^l (y_{ij.e} - \bar{y}_{ij.})^2}} \quad (7)$$

を計算する. \bar{v} と $\bar{y}_{ij.}$ はそれぞれ V と Y_{ij} の各要素の平均であり, 各 DS が計算する.

- (3) DS_i は V のトレーサビリティによる産地により以下の該当する動作を行う.

- V の産地と一致する Y_{ij} を持つ場合, 該当する $r_{(Y_{ij}V)}$ が r_{th} を超えなかったときはその旨を最大の $r_{(Y_{ij}V)}$ と産地情報とともに返す. 超えた場合はその旨を $r_{(Y_{ij}V)}$ とともに返す.

- V の産地と一致する Y_{ij} を持たない場合, 最大となる $r_{(Y_{ij}V)}$ を産地情報とともに返す.

産地判別結果の妥当性の検証等を目的として, CS の要求があれば DS は上記のほかに相関係数の全計算結果を返す.

産地判別 CS はトレーサビリティによる産地情報と一致する DS_i の報告により, 以下の動作を行う.

- r_{th} を超えたと報告した場合, 産地表示が正しい旨ユーザに $r_{(Y_{ij}V)}$ とともに提示する.
- r_{th} を超えなかったと報告した場合, ユーザに産地が偽装されている旨提示し, 各 DS_i が返した $r_{(Y_{ij}V)}$ の中から最大となる X_{ij} の産地情報を $r_{(Y_{ij}V)}$ とともに推定される産地として提示する.

r_{th} は産地と判定される相関係数の下限であり, 低すぎれば第2種過誤の原因となるため慎重に決める必要がある. $r_{th} = 0.7$ としているが, 3.3 節のシミュレーション結果によれば 0.8 でも十分である. 産地判別に要求される地理的粒度にもよるが, 将来的には適切な r_{th} を現実のデータに基づいて自動的に決定する必要があるかもしれない. トレーサビリティの産地情報と微量元素含有量による産地判別結果が一致するとき, $r_{(Y_{ij}V)}$ は産地判別の結果の確からしさを示すと考えられ, 1 に近いほど確からしいといえる.

微量元素含有量のデータの蓄積が進み, 産地数が増えたときの本手法の有効性を, 年に数回程度確認する必要がある. しかし, 散布図やクラスタ分析等ではデータ数が多いため判断が難しい. セルフチェックのため, 本システムは DS に蓄積された微量元素含有量データどうしの相関係数の計算が可能である. 2 つの DS DS_i と DS_g のそれぞれ j 番目のデータ Y_{ij} と h 番目のデータ Y_{gh} の相関係数は以下のように計算できる.

$$r_{(Y_{ij}Y_{gh})}$$

$$= \frac{\sum_{e=1}^l (y_{ghe} - \bar{y}_{gh.})(y_{ij.e} - \bar{y}_{ij.})}{\sqrt{\sum_{e=1}^l (y_{ghe} - \bar{y}_{gh.})^2} \sqrt{\sum_{e=1}^l (y_{ij.e} - \bar{y}_{ij.})^2}} \quad (8)$$

相関行列を各 DS に分割する様子を図 6 に示す. 網掛け部分は DS_i 自身で計算可能であり, その他の部分は DS_i が DS_g から Y_{gh} を取得する必要がある.

PC クラスタ等を用いて, 集中的に蓄積測定値相互間の相関係数を計算することも考えられる. Y_{ij} , Y_{gh} をあらかじめ各計算ノードに配布しておけば高速に計算できる. Y_{gh} を転送しながら計算する場合は計算ノード間の帯域が十分広い必要がある. 微量元素含有量データは時間の経過とともに増えるのみのため,

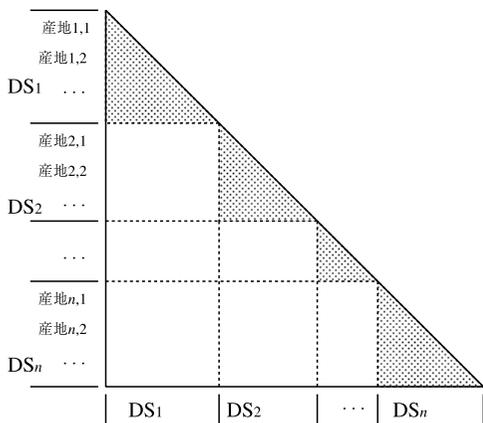


図 6 相関行列の各 DS への分割

Fig. 6 Division of correlation coefficient matrix into DSs.

Y_{gh} の差分のみを転送することでも転送時間を短縮できる。この場合、相関係数の計算を DS で行っても十分高速と考えられる。集中的に計算する場合でも DS からのデータの取得は差分の転送で十分である。本システムではこの計算の頻度が低いため、各 DS で分散して計算する方式を採用する。

4.2 プロトタイプによる計算時間の評価

本手法による産地判別は産地を判別したい野菜類の微量元素含有量と、データベースに蓄積された同じ作物・品種の野菜類の微量元素含有量の相関係数の計算を必要とする。単一の DS における計算量は、元素数を m 、差蓄積された微量元素含有量の数、すなわち産地数を n として $O(mn)$ と考えられる。相関係数の計算結果の通信に要する時間も同じと考えられる。元素数は 6 から 8 程度であり、簡単に増やせなかったり、増やしても有用とは限らないので定数と見なしてよい。蓄積された微量元素含有量データどうしの相関係数の単一の DS における計算時間は、図 6 の分割された四角形と三角形から $O(mn^2)$ と考えられる。

プロトタイプを Perl と C により実装し、相関係数の計算時間を測定した。データベースは PostgreSQL 7.4 を使用した。CS と DS、DS どうしの通信は SOAP を利用するが、効率を重視する場合は SOAP を通さずにデータを受け渡しできる。CS と DS を 1 サイトずつ用意し、実行時間を測定した。実行時間は起動時間、計算時間、回収時間の合計である。それぞれ、CS から DS への要求の送信、DS による計算、CS による計算結果の回収にかかる時間であり、逐次実行される。起動時間はすべての場合で約 0.75 秒であった。

規準化スコアの計算には全 DS に蓄積された微量元素含有量の平均の計算が必要である。この実行時間は

表 4 全 DS の平均の計算時間
Table 4 Computing time of averages in all DSs.

産地数	計算時間 [s]	回収時間 [s]
1,000	0.48	0.72
10,000	1.04	0.72
100,000	6.66	0.72

表 5 1 つのサンプルと各産地の測定値との相関係数の計算時間
Table 5 Computing time of correlation coefficients between a sample and accumulated samples.

産地数	計算時間 [s]	回収時間 [m:s]
1,000	0.56	0:01.49
10,000	1.82	0:08.96
100,000	15.13	1:05.67

表 6 蓄積測定値相互間の相関係数の計算時間

Table 6 Computing time of correlation coefficients between all accumulated samples in the same site.

産地数	計算時間 [m:s]	回収時間 [m:s]	転送サイズ
100	0:00.18	0:00.25	10.54 KB
1,000	0:03.07	0:03.07	1.17 MB
10,000	5:29.00	5:08.73	528.1 MB

表 7 蓄積測定値相互間の相関係数の計算時間 (他サイトから転送)
Table 7 Computing time of correlation coefficients between each accumulated sample with transferring from another site.

転送産地数	産地数	計算時間 [m:s]	回収時間 [m:s]	転送サイズ
100	100	0:01.60	0:00.41	110 KB
100	1,000	0:02.43	0:00.78	1.04 MB
100	10,000	0:12.25	0:06.82	10.15 MB
1,000	100	0:04.49	0:00.91	1.04 MB
1,000	1,000	0:09.49	0:06.86	10.63 MB
1,000	10,000	1:01.13	0:52.77	102.9 MB
10,000	100	0:38.67	0:06.95	10.41 MB
10,000	1,000	1:20.56	0:56.58	107.1 MB
10,000	10,000	9:21.75	8:54.45	1.008 GB

表 4 に示すように十分短い。CS から DS に産地判別を要求したときの実行時間を表 5 に示す。産地数に比例して計算時間と結果回収時間が延びる。1 サイトに蓄積される産地数を 10,000 程度と仮定すると、産地判別は数秒で実行できる。ここで、回収時間は相関係数の計算結果をすべて転送して測定した。通常の産地判別で、相関係数が最大となる結果もしくは産地判別の結果のみを返す場合、回収時間はすべての場合で 0.8 秒以内であった。表 6 に蓄積された微量元素含有量データどうしの相関係数 (図 6 の小さな三角 1 つ) の実行時間と計算結果回収で転送されるサイズを示す。おおむね $O(mn^2)$ である。表 7 に Y_{gh} を転送する場合 (図 6 の四角 1 つ) の実行時間と転送サイズを示す。産地数あたりの計算時間が表 6 の 2 倍となり、 Y_{gh} を転送する時間が必要である。表 6 と表 7 は結

果のサイズが大きいため SOAP を通さずに結果を回収した。

5. ま と め

本論文では、我々が提案した農産物の微量元素含有量を蓄積・比較することによる分散型産地判別システムについて述べた。判別対象と蓄積された農産物の微量元素含有量の相関係数を計算し、閾値を超えた場合に両者の産地が一致すると判断する。本手法は人手の介入を排し、判別規準を明確化した点で農芸化学者による手法よりも計算機に向いている。本手法の有効性を実際のデータとシミュレーションに基づいて示し、産地が増えた場合や細粒度での判別の可能性等の考察を行った。複数の産地が閾値を超えうが、実用上はトレーサビリティによる産地情報を検証できれば十分である。また、分散型産地判別システムのプロトタイプを評価した。1 サイトに 10,000 件のデータが蓄積されていた場合、数秒で産地判別が可能である。

今後の課題として、農産物の微量元素含有量の広域にわたる詳細な調査と、産地判別に用いる元素の選択の自動化があげられる。また、実用システムの実装、複数の DS での性能評価も重要である。現在、判別対象の農産物と同じ作物・品種の全蓄積データが計算対象である。相関係数が高くなると予測されるものだけに限定できれば、相関係数の計算を依頼する DS の数を減らすことができると考えられる。

謝辞 本研究は文部科学省産学連携推進事業（平成 15 年度～平成 19 年度）により実施された。

参 考 文 献

- 1) 社団法人農協流通研究所：青果物のトレーサビリティ導入ガイドライン (2003).
- 2) 農林水産省：トレーサビリティ関係。
<http://www.maff.go.jp/trace/top.htm>
- 3) 佐藤永欣, 上原 稔, 玉岡 迅, 下村講一郎, 山本浩文, 上條賢一：微量元素分析による野菜類の分散産地判別システムの検討, マルチメディア, 分散, 協調とモバイルシンポジウム, pp.773-776 (2005).
- 4) Sato, N., Uehara, M., Tamaoka, J., Shimomura, K., Yamamoto, H. and Kamijo, K.: Initial Design of Distributed Identification System for Geographical Origin by Trace Element Analysis, *Proc. 8th Int. Workshop on Network-Based Info. Syst. (NBIS2005)*, pp.79-83 (2005).
- 5) Sato, N., Uehara, M., Tamaoka, J., Shimomura, K., Yamamoto, H. and Kamijo, K.: Target Selection by Similarity Preserve Hash in Distributed System for Geographical Origin Identification of Vegetables, *Proc. 20th Int. Conf. on Advanced Info. Networking and Applications (AINA2006)*, pp.173-178 (2006).
- 6) T-Engine Forum: T-Engine Forum.
<http://www.t-engine.org/>
- 7) Ubiquitous ID Center: ucode.
<http://www.uidcenter.org/japanese/uid.html>
- 8) Baxter, M.J., Crews, H.M., Dennis, M.J., Goodall, I. and Anderson, D.: The Determination of the Authenticity of Wine from Its Trace Element Composition, *Food Chemistry*, Vol.60, No.3, pp.443-450 (1997).
- 9) Kallithraka, S., Arvanitoyannis, I., Kefalas, P., El-Zajouli, A., Soufleros, E. and Psarra, E.: Instrumental and Sensory Analysis of Greek Wines; Implementation of Principal Component Analysis (PCA) for Classification According to Geographical Origin, *Food Chemistry*, Vol.73, No.4, pp.501-514 (2001).
- 10) Anderson, K.A. and Smith, B.W.: Chemical Profiling to Differentiate Geographic Growing Origins of Coffee, *J. Agric. Food Chem.*, Vol.50, No.7, pp.2068-2075 (2002).
- 11) Fernández-Cáceres, P.L., Martín, M.J., Pablos, F. and González, A.G.: Differentiation of Tea (*Camellia sinensis*) Varieties and Their Geographical Origin According to their Metal Content, *J. Agric. Food Chem.*, Vol.49, No.10, pp.4775-4779 (2001).
- 12) Simpkins, W.A., Louie, H., Wu, M., Harrison, M. and Goldberg, D.: Trace Elements in Australian Orange Juice and Other Products, *Food Chemistry*, Vol.71, No.4, pp.423-433 (2000).
- 13) Anderson, K.A., Magnuson, B.A., Tschirgi, M.L. and Smith, B.: Determining the Geographical Origin of Potatoes with Trace Metal Analysis Using Statistical and Neural Network Classifiers, *J. Agric. Food Chem.*, Vol.47, No.4, pp.1568-1575 (1999).
- 14) 有山 薫, 堀田 博, 安井明美：ネギの産地判別のための無機元素測定法の確立と予備的検討, 分析化学, Vol.52, No.11, pp.969-978 (2003).
- 15) 安野正博, 木下恵子, 小塚大生, 木谷祐亮, 安井義徳, 田窪 健, 斉藤春絵, 森田正晶：農産物原産地表示適正化を担保する分析評価方法について—ニンニクにおける国産品と輸入品(中国)の比較, 農林水産消費技術センター調査研究報告, No.22, pp.77-86 (1998).
- 16) 田口陽嗣：無機元素組成・含有量調査による青果物の産地判別—ICP-MSによる青果物中の無機元素分析及びタマネギの産地推定, 農林水産消費技術センター調査研究報告, No.26, pp.1-8 (2002).

- 17) 安井明美, 進藤久美子: 玄米中の無機元素組成による産地判別, 分析化学, Vol.49, No.6, pp.406-410 (2000).
- 18) Martin, G.J., Guillou, C., Martin, M.L., Cabanis, M.-T., Tep, Y. and Aerny, J.: Natural Factors of Isotope Fractionation and the Characterization of Wines, *J. Agric. Food Chem.*, Vol.36, No.2, pp.316-322 (1988).
- 19) McDonald, M.S., Hughes, M., Burns, J., Lean, M.E., Matthews, D. and Crozier, A.: Survey of the Free and Conjugated Myricetin and Quercetin Content of Red Wines of Different Geographical Origines, *J. Agric. Food Chem.*, Vol.46, No.2, pp.368-375 (1998).
- 20) 総務省統計研修所(編): 日本の統計, 総務省統計局 (2006). <http://www.stat.go.jp/data/nihon/>

(平成 17 年 11 月 28 日受付)

(平成 18 年 5 月 9 日採録)



佐藤 永欣 (正会員)

昭和 51 年生まれ。平成 16 年東洋大学大学院博士後期課程修了。博士(工学)。同年東洋大学植物機能研究センター研究助手。分散情報検索等の研究に従事。IEEE, 電子情報通信学会各会員。

信学会各会員。



上原 稔 (正会員)

昭和 39 年生まれ。平成 7 年慶應義塾大学大学院博士課程博士(工学)。平成 10 年東洋大学工学部情報工学科助教授。平成 17 年同教授。分散計算, プログラミング言語等に興味を持つ。ACM, IEEE 各会員。

を持つ。ACM, IEEE 各会員。



玉岡 迅

昭和 31 年生まれ。東京大学大学院博士課程修了。農学博士。平成 10 年より東洋大学生命科学部助教授。微生物分類, 機器分析, 生物情報学。日本農芸化学会, 生物工学会各会員。



下村 講一郎

昭和 26 年生まれ。昭和 56 年九州大学大学院薬学研究科博士課程修了。薬学博士。国立医薬品食品衛生研究所等を経て平成 13 年東洋大学生命科学部教授。植物の生産する抗酸化物質に関する研究等に従事。日本薬学会, 日本農芸化学会, 日本植物細胞分子生物学会等各会員。



山本 浩文

昭和 35 年生まれ。昭和 63 年京都大学大学院薬学研究科博士後期課程修了。薬学博士。長崎大学薬学部助教授を経て平成 15 年東洋大学生命科学部教授。植物生化学, 植物代謝工学等の研究に従事。日本薬学会, 日本植物細胞分子生物学会, 植物化学研究会等各会員。



上條 賢一

昭和 24 年生まれ。平成 6 年京都大学理学博士(地球物理学)。現在, 東洋大学大学院生命科学研究科生命科学専攻教授。情報科学, 地球情報学, 生命情報学における複雑系の研究に従事。電子情報通信学会, 人工知能学会, 日本気象学会, 日本測地学会各会員。

研究に従事。電子情報通信学会, 人工知能学会, 日本気象学会, 日本測地学会各会員。