

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2020-174553

(P2020-174553A)

(43) 公開日 令和2年10月29日(2020.10.29)

(51) Int.Cl.		F 1				テーマコード (参考)
AO1G 7/00	(2006.01)	AO1G 7/00	603			2G041
GO1N 27/62	(2006.01)	GO1N 27/62		V		

審査請求 未請求 請求項の数 13 O L (全 20 頁)

(21) 出願番号	特願2019-78180 (P2019-78180)	(71) 出願人	000000918 花王株式会社 東京都中央区日本橋茅場町1丁目14番1 0号
(22) 出願日	平成31年4月16日 (2019.4.16)	(74) 代理人	110000084 特許業務法人アルガ特許事務所
		(72) 発明者	未次 舞 和歌山県和歌山市湊1334 花王株式会 社研究所内
		(72) 発明者	出口 潤 和歌山県和歌山市湊1334 花王株式会 社研究所内
		(72) 発明者	藤松 輝久 栃木県芳賀郡市貝町赤羽2606 花王株 式会社研究所内

最終頁に続く

(54) 【発明の名称】 ダイズの収量予測方法

(57) 【要約】

【課題】ダイズの収量を早期に精度よく予測する方法を提供する。

【解決手段】ダイズから採取された葉サンプルから1以上の成分の分析データを取得し、当該データとダイズ収量との相関性を利用してダイズの収量を予測する、ダイズの収量予測方法であって、成分が2-ヒドロキシピリジン、コリン、クエン酸、グリセリン酸、グリシン、L-ピログルタミン酸、マロン酸、スクロース及びトレイトールから選ばれる1種以上である、方法。

【選択図】なし

【特許請求の範囲】

【請求項 1】

ダイズから採取された葉サンプルから 1 以上の成分の分析データを取得し、当該データとダイズ収量との相関性を利用してダイズの収量を予測する、ダイズの収量予測方法であって、成分が 2 - ヒドロキシピリジン、コリン、クエン酸、グリセリン酸、グリシン、L - ピログルタミン酸、マロン酸、スクロース及びトレイトールから選ばれる 1 種以上である、方法。

【請求項 2】

成分が、2 - ヒドロキシピリジン、グリシン、L - ピログルタミン酸及びスクロースから選ばれる 1 種以上である、請求項 1 記載の方法。

10

【請求項 3】

葉サンプルが、初生葉展開期から子実肥大期のダイズから採取される、請求項 1 又は 2 記載の方法。

【請求項 4】

葉サンプルが、初生葉展開期から開花期までのダイズから採取される、請求項 1 ~ 3 のいずれか 1 項に記載の方法。

【請求項 5】

分析データが、質量分析データである請求項 1 ~ 4 のいずれか 1 項に記載の方法。

【請求項 6】

葉サンプルから取得された成分の分析データを、前記 9 成分から選択される分析データを用いて構築された収量予測モデルと照合する工程を含む、請求項 1 ~ 5 のいずれか 1 項に記載の方法。

20

【請求項 7】

収量予測モデルが、前記 9 成分中の分析データうち、VIP 値が 6 位以上の成分について任意の 4 個のデータを用いる、請求項 6 に記載の方法。

【請求項 8】

収量予測モデルが、前記 9 成分の分析データのうち、VIP 値が 7 位以上の成分について任意の 5 個のデータを用いる、請求項 6 に記載の方法。

【請求項 9】

収量予測モデルが、前記 9 成分の分析データうち、VIP 値が 8 位以上の成分について任意の 6 個のデータを用いる、請求項 6 に記載の方法。

30

【請求項 10】

収量予測モデルが、前記 9 成分の分析データのうち、任意の 7 個のデータを用いる、請求項 6 に記載の方法。

【請求項 11】

収量予測モデルが、前記 9 成分の分析データのうち、VIP 値 1 位以下の成分のデータすべて、VIP 値 2 位以下の成分のデータすべて、又は VIP 値 3 位以下の成分のデータすべてを用いる、請求項 6 に記載の方法。

【請求項 12】

収量予測モデルが、OPLS 法を用いて構築されたモデルである請求項 6 ~ 11 のいずれか 1 項記載の方法。

40

【請求項 13】

収量予測モデルが、回帰分析手法を用いて構築されたモデルである請求項 6 ~ 11 のいずれか 1 項記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明はダイズの収量を早期に予測する方法に関する。

【背景技術】

【0002】

50

ダイズは、重要な穀物であり、日本を始め世界中で広く食されている。また他の代表的な穀物である米、麦、トウモロコシと異なり、タンパク質及び脂質の割合が高く栄養価も豊富である。そのため飼料や油脂原料としても重要であり、収量を増加させる技術の開発が行われている。

ダイズの生育期間は、品種や栽培条件によって若干異なるが、通常、播種から収穫まで4 - 5 か月という長期間を要する。よって、ダイズの収量を増加させる技術の開発において、収量評価を行うには栽培に多くの時間を必要とする。さらに、日本のような季節・気候条件では、収穫まで4 - 5 か月を要するダイズの栽培は年に1回が一般的である。屋外栽培での収量評価が年に1度しかできず収量増加技術の開発の障害となっていることから、早期に収量を予測する方法が求められてきた。また、実際の生産場面において、早期に収量を予測することができれば、生産者は安定した収量確保のために費用コストのかかる追加技術を投入すべきかどうかの判断を容易に下すことができる。

10

【0003】

これまでも生育中の植物体の発育状態と収量との相関性を利用した早期に収量性を評価する方法が種々検討されている。例えば、非特許文献1ではダイズの播種後40日程度に測定した主茎長と収量との相関 ($r = 0.51$) を利用する方法、非特許文献2では、播種後60 - 70日程度に測定した地上部の乾燥重量と収量との相関 ($r = 0.66$) を利用する方法が開示されている。また、非特許文献3及び4では、画像診断技術を用いて、圃場においてNDVI (正規化植生指標) やLAI (葉面積指数) 及び群落分光反射率を測定し、生育や収量性を評価する試みが開示されている。

20

【0004】

しかしながら、非特許文献1の方法は、比較的早期に収量を予測できる可能性がある一方で相関性が十分でなく、また非特許文献2の方法では相関性は向上するものの予測時期が栽培開始から2か月以上であり生育期間の半分が経過すること、及び地上部乾燥重を測定するため侵襲的であり、個体毎の予測因子と収量とを対応させたい場合の評価には向いていない。非特許文献3及び4の方法は非破壊で簡易的な測定であるといえるが、予測時期が開花期、すなわち播種後50日前後以降になること、また精度の面でも十分とは言えない。

【0005】

このほかイネにおいて、播種後15日程度の地上部から抽出される代謝物をGC - MSにより網羅的に測定し、それらのデータを用いてハイブリッドライス収量予測モデルを作成したことが報告されているが (非特許文献5)、この報告では、通常予測モデル構築の際に行われるクロスバリデーションというモデルの予測性評価が行われておらず、検証が十分とは云えない。また、侵襲的であり、個体毎の予測因子と収量とを対応させたい場合の評価には向いていない。

30

【先行技術文献】

【非特許文献】

【0006】

【非特許文献1】藤田与一ら、平成21年度「関東東海北陸農業」研究成果情報、「重粘度地帯におけるダイズ「エンレイ」の多収事例に基づく収量構成要素と生育指標」、http://www.naro.affrc.go.jp/org/narc/seika/kanto21/12/21_12_04.html

40

【非特許文献2】井上健一、高橋正樹、第229回日本作物学会講演会要旨集、2010、p50、「物質生産と窒素の蓄積から見たダイズの多収生育相」

【非特許文献3】長南友也ら、第245回日本作物学会講演会要旨集、2018、p83、「ダイズの簡易的な開花前生育診断技術」

【非特許文献4】渡邊智也ら、第245回日本作物学会講演会要旨集、2018、p84、「非破壊計測と畳み込みニューラルネットワークを利用したダイズの収量評価」

【非特許文献5】Dan,Z.et al., Scientific Reports,2016,6,21732

【発明の概要】

【発明が解決しようとする課題】

50

【0007】

本発明は、ダイズの収量を早期に精度よく予測する方法を提供することに関する。

【課題を解決するための手段】

【0008】

本発明者らは、ダイズの収量性評価について種々検討した結果、葉中に含まれる代謝物にはその存在量が収量と相関する成分があること、そして、播種後1か月程度という早期に展開葉を1枚採取し、葉中に含まれる成分を分析し、解析することで最終的な収量を個体レベルで評価できることを見出した。

【0009】

すなわち、本発明は、ダイズから採取された葉サンプルから1以上の成分の分析データを取得し、当該データとダイズ収量との相関性を利用してダイズの収量を予測する、ダイズの収量予測方法であって、成分が2-ヒドロキシピリジン、コリン、クエン酸、グリセリン酸、グリシン、L-ピログルタミン酸、マロン酸、スクロース及びトレイトールから選ばれる1種以上である、方法を提供する。

【発明の効果】

【0010】

本発明の方法によれば、ダイズの収量を早期に予測できる。これにより、例えば、収量確保のための追加技術投入の判断が容易となるほか、収量増加技術の開発の大幅な効率化を図ることができる。

【図面の簡単な説明】

【0011】

【図1】全125データを用いて構築されたOPLSモデルによる収量の予測値と実測値との関係を示す図。

【図2】図1のモデルにおけるVIP値1位以下、2位以下、3位以下、4位以下、5位以下及び6位以下のすべての成分データを用いてOPLS法により構築した各々のモデルの R^2 （図中では $R^2 Y$ と表示）値及び Q^2 （図中では Q^2 と表示）値を示す図。

【図3】図1のモデルにおけるVIP値1位及び2位、VIP値1位～3位及びVIP値1位～4位の成分データを用いてOPLS法により構築した各々のモデルの R^2 （図中では $R^2 Y$ と表示）値及び Q^2 （図中では Q^2 と表示）値を示す図。

【図4】図1のモデルにおけるVIP値上位1位～4位、2位～5位、3位～6位、4位～7位、5位～8位及び6位～9位の成分データを用いてOPLS法により構築した各々のモデルの R^2 （図中では $R^2 Y$ と表示）値及び Q^2 （図中では Q^2 と表示）値を示す図。

【図5】図1のモデルにおけるVIP値上位1位～5位、2位～6位、3位～7位、4位～8位及び5位～9位の成分データを用いてOPLS法により構築した各々モデルの R^2 （図中では $R^2 Y$ と表示）値及び Q^2 （図中では Q^2 と表示）値を示す図。

【図6】図1のモデルにおけるVIP値上位1位～6位、2位～7位、3位～8位及び4位～9位の成分データを用いてOPLS法により構築した各々のモデルの R^2 （図中では $R^2 Y$ と表示）値及び Q^2 （図中では Q^2 と表示）値を示す図。

【図7】図1のモデルにおけるVIP値上位1位～7位、2位～8位及び3位～9位の成分データを用いてOPLS法により構築した各々のモデルの R^2 （図中では $R^2 Y$ と表示）値及び Q^2 （図中では Q^2 と表示）値を示す図。

【発明を実施するための形態】

【0012】

本発明において、ダイズとは、マメ科の一年草である大豆（学名 *Glycine max*）を意味する。その品種はフクユタカ、エンレイ、里のほほえみ、湯上がり娘、リュウホウ、スズユタカ等、多岐にわたるが、本発明においてはそれに限定されるものではない。

【0013】

ダイズの出芽から落葉までの生育ステージは、VC：初生葉展開期（播種後7日前後）、R1-2：開花期（播種後50日前後）、R3-4：着莢期（播種後70日前後）、R

10

20

30

40

50

5 - 6 : 子実肥大期 (播種後 90 日前後) に分けられる (Fehr, W.R., Caviness, C.E., 1977. Stages of soybean development. Cooperative Extension Service, Agriculture and Home Economics Experiment Station, Iowa State University, Ames, Iowa)。本発明において、サンプルとして使用されるダイズの葉の採取時期は、葉が採取可能な初生葉展開期 (VC) から子実肥大期 (R5 - 6) までの間に行われればよく、好ましくは初生葉展開期 ~ R3 - 4 期、より好ましくは播種後 14 日 ~ R3 - 4 期、より好ましくは播種後 21 日 ~ R1 - 2 期、さらに好ましくは播種後 28 日 ~ R1 - 2 期のダイズが挙げられる。尚、上記各生育ステージにおける前後の日数幅は 10 日間以内が好適である。

或いは、ダイズの葉の採取時期は、播種後 7 日以上、好ましくは 14 日目以上、より好ましくは 21 日目以上、さらに好ましくは 28 日以上で、且つ好ましくは播種後 50 日より前、より好ましくは播種後 40 日より前、さらに好ましくは 35 日目より前であり得る。また、播種後 7 ~ 50 日目、好ましくは 14 ~ 40 日目、より好ましくは 28 ~ 35 日目であり得る。例えば、播種後 30 日 ± 3 ~ 5 日目のダイズから葉を採取するのが好適である。

【 0014 】

葉の採取部位は、特に限定されないが、例えば、最上位に展開する本葉の 1 または 2 葉齢古い本葉を構成する複葉 3 枚の内、中央の複葉を採取することが挙げられる。

【 0015 】

本発明において、取得される成分の分析データ (「成分データ」とも称する) としては、本発明で規定する成分を同定することができるものであれば特に限定されず、高速液体クロマトグラフィー (HPLC)、ガスクロマトグラフィー (GC)、イオンクロマトグラフィー、質量分析 (MS)、近赤外分光分析 (NIR)、フーリエ変換赤外分光分析 (FT - IR)、核磁気共鳴分析 (NMR)、フーリエ変換核磁気共鳴分析 (FT - NMR)、誘導結合プラズマ質量分析計 (ICP - MS)、ガスクロマトグラフや液体クロマトグラフと質量分析とを組合せた GC - MS、LC - MS 等の機器分析手段を用いて分析・測定されたデータが挙げられるが、好ましくは質量分析データであり、より好ましくは GC - MS による質量分析データである。

質量分析データとしては、精密質量 (「 m / z 値」)、イオン強度、保持時間等が挙げられるが、好ましくは精密質量の情報である。なお、精密質量とは、小数点以下 4 桁以上の精度にて測定された質量をさす。

【 0016 】

葉サンプルを、上記機器分析手段に適用するためには、分析手段に応じて適宜前処理されるが、通常、採取した葉はアルミホイルで包み直ちに液体窒素中で凍結して代謝反応を停止させ、凍結乾燥にかけて乾燥した後、抽出操作に供される。

抽出は、凍結乾燥した葉サンプルを、ピーズ粉碎機等を用いて粉碎した後、抽出溶媒を添加して攪拌することにより行われる。ここで用いられる抽出溶媒としては、メタノール、エタノール、ブタノール、アセトニトリル、クロロホルム、酢酸エチル、ヘキサン、アセトン、イソプロパノール、水等及びそれらを混合したものが挙げられる。分析手段として GC - MS を用いる場合には、メタノール / 水 / クロロホルムの混合溶媒等が好適に使用される。

【 0017 】

本発明において、分析される葉中の成分としては、2 - ヒドロキシピリジン (2 - hydroxypyridine)、コリン (choline)、クエン酸 (citric acid)、グリセリン酸 (glyceric acid)、グリシン (glycine)、L - ピログルタミン酸 (L - pyroglutamic acid)、マロン酸 (malonic acid)、スクロース (sucrose) 及びトレイトール (threitol) から選ばれる 1 種以上である。

当該 9 成分はダイズの代謝物質から選択抽出されたものであり、その選択方法は詳細には実施例に示すとおりであるが、概略すると、1) 2015 年 ~ 2017 年にかけて、土壌、品種、肥料を変えたダイズ 125 株を栽培し、2) それぞれ播種後 1 ヶ月前後に葉を

10

20

30

40

50

1枚採取し、3)メタノール/水/クロロホルムの混合溶媒(5:2:2, v/v/v)を用いて成分抽出を行い誘導体化を行った後、4)GC/MS分析を行い、得られた各成分由来ピーク情報(保持時間、質量情報)を既存のGC/MS質量スペクトルライブラリと照合することによりピークを同定し、5)分析データと対応する収量データとの相関解析を行い、収量と相関すると判定した成分を取得する、というものである。

【0018】

取得した125個の葉中9成分の分析データと対応する収量データとの相関解析の結果(各成分の分析データのピーク面積と収量との単相関係数 r 及び無相関の検定により p 値を算出)は後記表3に示すとおりであり、このうち、2-ヒドロキシピリジン、グリシン、L-ピログルタミン酸及びスクロースの4種は相関係数の絶対値 $|r|$ が0.51を超えている。したがって、本発明における分析対象成分としては、2-ヒドロキシピリジン、グリシン、L-ピログルタミン酸及びスクロースから選ばれる1種以上を含むのがより好ましい。

10

【0019】

したがって、ダイズの収量の予測手段は、予測したいサンプルについて、上記9成分、好ましくは2-ヒドロキシピリジン、グリシン、L-ピログルタミン酸及びスクロースから選ばれる1種以上の分析データを取得し、既知の収量とピーク面積との相関関係から収量値を推定することができる。

【0020】

また、上記9成分の分析データから複数をし、多変量解析手法を用いて構築された収量予測モデルと照合することにより、収量を予測することができる。

20

すなわち、播種から所定期間経過後のダイズの葉を採取し、分析サンプルを得、該分析サンプルを機器分析に供して機器分析データ得、該機器分析データを、収量予測モデルと照合することにより、当該ダイズの収量を予測することができる。

【0021】

収量予測モデルは、説明変数に例えば各成分の分析データのピーク面積値を、また目的変数に収量値を用いた回帰分析を行うことにより構築できる。回帰分析法としては、例えば主成分回帰分析、PLS(Partial least squares projection to latent structures)回帰分析、OPLS(Orthogonal projections to latent structures)回帰分析、一般化線形回帰分析の他、バギング、サポートベクターマシン、ランダムフォレスト、ニューラルネットワーク回帰分析等の機械学習・回帰分析手法等の多変量回帰分析手法が挙げられる。このうち、PLS法、PLS法の改良版であるOPLS法、或いは機械学習・回帰分析手法を用いるのが好ましい。OPLS法は、PLS法に比べ予測性は同じだが、解釈のための視覚化がより容易になる点が今回のような目的においては優れている。PLS法及びOPLS法は、共に高次元のデータから情報を集約し少数の潜在変数に置き換え、その潜在変数を用いて目的変数を表現する方法である。潜在変数の数を適切に選ぶことが重要であり、潜在変数の数を決めるのによく利用されるのがクロスバリデーション(交差検証)である。すなわち、モデル構築用データをいくつかのグループに分割し、あるグループをモデル検証に、その他のグループをモデル構築に用いて予測誤差を見積り、この作業を、グループを入れ替えながら繰り返して、予測誤差の合計が最小となる潜在変数の数が選ばれる。

30

40

【0022】

予測モデルの評価は、主に2つの指標で判断される。1つは予測精度を表す R^2 、もう1つは予測性を表す Q^2 である。 R^2 は予測モデル構築に使用したデータの実測値とモデルで計算した予測値との相関係数の2乗であり、1に近いほど予測精度が高いことを示している。一方、 Q^2 は、上記クロスバリデーションの結果であり、実測値と、繰り返し実施したモデル検証の結果である予測値との相関係数の2乗を表している。本発明のダイズ収量予測モデルにおいては、 $Q^2 > 0.50$ をモデル評価の基準とするのが好ましい。なお、常に $R^2 > Q^2$ となるため、 $Q^2 > 0.50$ は同時に $R^2 > 0.50$ を満たすこととなる。

50

以下に、上記9成分の質量分析データのピーク面積値と、子実収量を用いた種々のダイズ収量予測モデルを作成しその精度を検証した結果を示す。

【0023】

(1) 全ての成分情報を用いた収量予測モデルの構築

1データ当り9個の成分の分析データのピーク面積値と収量値を持つ全125個のデータマトリックスからOPLSモデル(図1)を構築した。なお、構築の際は、各成分の分析データのピーク面積値及び収量データはオートスケーリングにより平均0、分散1に変換した。 $R^2 = 0.56$ 、 $Q^2 = 0.55$ であり、高い予測性能を持つモデルといえる。なお、L-ピログルタミン酸の分析データのみで構築したモデルは $R^2 = 0.32$ 、 $Q^2 = 0.29$ であるため、複数の成分データを用いることでより高い精度のモデルを構築することができる。

10

上記モデルではVIP (Variable Importance in the Projection, 投影における変数重要性) 値とよばれる各成分に与えられるモデル性能への寄与度が算出される。

VIP値は、下記式1により求められる。

【0024】

【数1】

$$VIP = \sqrt{\sum_{a=1}^A (W_{ak}^2 * (SSY_{a-1} - SSY_a) * \frac{K}{SSY_0 - SSYA})}$$

A: 潜在変数の数

W: 説明変数のウエイト

SSY: 予測残差の分散

K: 成分数

20

【0025】

VIP値はその値が大きいほどモデルへの寄与度が大きく、相関係数の絶対値とも相関する。VIP値のリストを後記表4に示す。

【0026】

(2-1) VIP値が下位の成分由来ピークを用いたモデル

VIP値1位以下の成分の分析データすべて、VIP値2位以下の成分の分析データすべて、又はVIP値3位以下の成分の分析データすべて、を用いて構築されたOPLSモデルは $Q^2 > 0.50$ を満たし、収量予測モデルとして好ましい(図2)。VIP値4位以下及び5位以下の成分の分析データすべてを用いて構築されたOPLSモデルは何れも $Q^2 > 0.50$ を満たさなかった。

30

【0027】

(2-2) VIP値上位の成分の分析データを複数個用いたモデル

9個の成分の分析データから選ばれる任意の4個以上のデータを用いて構築されたモデルのうち、 $Q^2 > 0.50$ を満たすモデルは収量予測モデルとして好ましい。具体的には、以下の1)~4)に示すモデルが挙げられる。

1) VIP値上位から順に連続する4個の成分の分析データを用いたモデル

VIP値6位以上の成分についてVIP値上位から順に連続する4個の成分の分析データを用いて構築されたOPLSモデルは $Q^2 > 0.50$ を満たし、収量予測モデルとして好ましい。また、VIP値3位~6位の成分の分析データを用いて構築されたOPLSモデル $Q^2 > 0.50$ を満たしたことから、VIP値6位以上であれば、任意の4個の成分の分析データを用いることで、 $Q^2 > 0.50$ を満たす予測モデルが得られると考えられる。

40

例えば、VIP値上位1位~4位、2位~5位、3位~6位の成分の分析データを用いて構築されたOPLSモデルが挙げられる(図4)。

【0028】

2) VIP値上位から順に連続する5個の成分の分析データを用いたモデル

VIP値7位以上の成分についてVIP値上位から順に連続する5個の成分の分析デー

50

タを用いて構築された O P L S モデルは $Q^2 > 0.50$ を満たし、収量予測モデルとして好ましい。また、V I P 値 3 位 ~ 7 位の成分の分析データを用いて構築された O P L S モデル $Q^2 > 0.50$ を満たしたことから、V I P 値 7 位以上であれば、任意の 5 個の成分の分析データを用いることで、 $Q^2 > 0.50$ を満たす予測モデルが得られると考えられる。

例えば、V I P 値上位 1 位 ~ 5 位、2 位 ~ 6 位、3 位 ~ 7 位の成分の分析データを用いて構築された O P L S モデルが挙げられる (図 5)。

【 0 0 2 9 】

3) V I P 値上位から順に連続する 6 個の成分の分析データを用いたモデル

V I P 値 8 位以上の成分について V I P 値上位から順に連続する 6 個の成分の分析データを用いて構築された O P L S モデルは $Q^2 > 0.50$ を満たし、収量予測モデルとして好ましい。また、V I P 値 3 位 ~ 8 位の成分の分析データを用いて構築された O P L S モデル $Q^2 > 0.50$ を満たしたことから、V I P 値 8 位以上であれば、任意の 6 個の成分の分析データを用いることで、 $Q^2 > 0.50$ を満たす予測モデルが得られると考えられる。

例えば、V I P 値上位 1 位 ~ 6 位、2 位 ~ 7 位、3 位 ~ 8 位の成分の分析データを用いて構築された O P L S モデルが挙げられる (図 6、図 2)。

【 0 0 3 0 】

4) V I P 値上位から順に連続する 7 個の成分の分析データを用いたモデル

V I P 値上位から順に連続する 7 個の成分の分析データを用いて構築された O P L S モデルは $Q^2 > 0.50$ を満たし、収量予測モデルとして好ましい。また、V I P 値 3 位 ~ 9 位の成分の分析データを用いて構築された O P L S モデル $Q^2 > 0.50$ を満たしたことから、9 個の成分の内、任意の 7 個の成分の分析データを用いることで、 $Q^2 > 0.50$ を満たす予測モデルが得られると考えられる。

例えば、V I P 値上位 1 位 ~ 7 位、2 位 ~ 8 位又は 3 位 ~ 9 位の成分の分析データを用いて構築された O P L S モデルが挙げられる (図 7)。

【 0 0 3 1 】

本発明の態様及び好ましい実施態様を以下に示す。

< 1 > ダイズから採取された葉サンプルから 1 以上の成分の分析データを取得し、当該データとダイズ収量との相関性を利用してダイズの収量を予測する、ダイズの収量予測方法であって、成分が 2 - ヒドロキシピリジン、コリン、クエン酸、グリセリン酸、グリシン、L - ピログルタミン酸、マロン酸、スクロース及びトレイトールから選ばれる 1 種以上である、方法。

< 2 > 成分が、2 - ヒドロキシピリジン、グリシン、L - ピログルタミン酸及びスクロースから選ばれる 1 種以上である、< 1 > 記載の方法。

< 3 > 葉サンプルが、初生葉展開期から子実肥大期のダイズから採取される、< 1 > 又は < 2 > 記載の方法。

< 4 > 葉サンプルが、初生葉展開期から開花期までのダイズから採取される、< 1 > ~ < 3 > のいずれかに記載の方法。

< 5 > 分析データが、質量分析データである < 1 > ~ < 4 > のいずれかに記載の方法。

< 6 > 葉サンプルから取得された成分の分析データを、前記 9 成分から選択される分析データを用いて構築された収量予測モデルと照合する工程を含む、< 1 > ~ < 5 > のいずれかに記載の方法。

< 7 > 収量予測モデルが、前記 9 成分中の分析データうち、V I P 値が 6 位以上の成分について任意の 4 個のデータを用いる、< 6 > に記載の方法。

< 8 > 収量予測モデルが、前記 9 成分の分析データうち V I P 値が 7 位以上の成分について任意の 5 個のデータを用いる、< 6 > に記載の方法。

< 9 > 収量予測モデルが、前記 9 成分の分析データうち V I P 値が 8 位以上の成分について任意の 6 個のデータを用いる、< 6 > に記載の方法。

< 1 0 > 収量予測モデルが、前記 9 成分の分析データうち、任意の 7 個のデータを用い

10

20

30

40

50

る、< 6 > に記載の方法。

< 1 1 > 収量予測モデルが、前記 9 成分の分析データうち、VIP 値 1 位以下の成分のデータすべて、VIP 値 2 位以下の成分のデータすべて、又は VIP 値 3 位以下の成分のデータすべてを用いる、< 6 > に記載の方法。

【 0 0 3 2 】

< 1 2 > 収量予測モデルが、前記 9 成分の分析データうち、VIP 値上位 1 位 ~ 4 位、2 位 ~ 5 位又は 3 位 ~ 6 位の成分の分析データを用いる、< 6 > に記載の方法。

< 1 3 > 収量予測モデルが、前記 9 成分の分析データうち、VIP 値上位 1 位 ~ 5 位、2 位 ~ 6 位又は 3 位 ~ 7 位の成分の分析データを用いる、< 6 > に記載の方法。

< 1 4 > 収量予測モデルが、前記 9 成分の分析データうち、VIP 値上位 1 位 ~ 6 位、2 位 ~ 7 位又は 3 位 ~ 8 位の成分の分析データを用いる、< 6 > に記載の方法。

< 1 5 > 収量予測モデルが、前記 9 成分の分析データうち、VIP 値上位 1 位 ~ 7 位、2 位 ~ 8 位又は 3 位 ~ 9 位の成分の分析データを用いる、< 6 > に記載の方法。

< 1 6 > 収量予測モデルが、OPLS 法を用いて構築されたモデルである < 6 > ~ < 1 5 > のいずれかに記載の方法。

< 1 7 > 収量予測モデルが、回帰分析手法を用いて構築されたモデルである < 6 > ~ < 1 5 > のいずれかに記載の方法。

< 1 8 > 質量分析データが小数点以下 4 桁以上の精度である、< 5 > ~ < 1 5 > のいずれかに記載の方法。

【 実施例 】

【 0 0 3 3 】

1. 各栽培試験

2015 年から 2017 年に実施した屋外ポット栽培試験データについて詳述する。試験は全部で 5 試験実施した。

1) 2015 年ポット試験 (1) :

栃木県内にてポット栽培を実施した。土壌は国内の圃場土を用い、窒素 : リン酸 : カリウム = 3 : 6 : 6 (Kg / 10 a) となるように施肥を行い、土壌を耕耘機で耕耘した。土壌はこの耕耘後の土壌を用いた。ポットには 1 / 2000 a ワグネルポットを用い、上記土壌を 1 ポットあたり約 8 L 詰め、15 ポットを準備した。2015 年 6 月 25 日に 3 粒播きで各ポット内 2 カ所に播種した (1 ポットあたり 6 粒使用)。なお、品種は、「里のほほえみ」を用いた。初生葉展開期に 1 カ所につき 1 本に間引きし、各ポット 2 株立てとした。収穫は 11 月 9 日に実施した (播種後 137 日)。なお、収量予測用には 5 ポット 10 株を用いた。

【 0 0 3 4 】

2) 2015 年ポット試験 (2) :

和歌山県内にてポット栽培を実施した。土壌は国内の圃場土を用い、窒素 : リン酸 : カリウム = 1 : 6 : 6、3 : 6 : 6 及び 10 : 6 : 6 (Kg / 10 a) となるように施肥を行い、耕耘後の土壌を用いた (窒素量のみ異なる 3 種の施肥条件を設定した)。ポットには 1 / 2000 a ワグネルポットを用い、上記土壌を 1 ポットあたり約 8 L 詰め、各施肥条件でそれぞれ 15 ポット計 45 ポットを準備した。2015 年 7 月 1 日に 3 粒播きで各ポット内 2 カ所に播種した (各ポット 6 粒使用)。なお、品種は、「フクユタカ」を用いた。初生葉展開期に 1 カ所につき 1 本に間引きし、各ポット 2 株立てとした。収穫は、11 月 11 日に実施した (播種後 133 日)。なお、収量予測用には各 5 ポット 10 株を用いる計画だったが、1 株が欠株したため、計 29 株を用いた。

【 0 0 3 5 】

3) 2016 年ポット試験 :

栃木県内にてポット栽培を実施した。土壌は国内圃場土壌を用い、試験を実施した。ポットには 1 / 2000 a ワグネルポットを用い、上記土壌を 1 ポットあたり約 8 L 詰め、75 ポットを準備した。数日間静置後、2015 年同様に播種をおこなった。播種は、2016 年 7 月 1 日におこない、11 月 15 日に収穫をおこなった。なお、品種は「里のほほ

えみ」を用いた。また、収量予測用には23ポット、46株を用いた。

【0036】

4) 2017年ポット試験：

栃木県内にてポット栽培を実施した。土壌は、国内圃場土を用い、そこへ肥料として苦土石灰(協和)を 125 g/m^2 及びリン酸入り油かす(大栄物産)を 100 g/m^2 添加した土壌(1×)及び肥料を添加した土壌と未添加土壌を半量ずつ混和した土壌(0.5×)の2種類を用いた。また品種は、「里のほほえみ」、「フクユタカ」、「エンレイ」及び「湯上がり娘」の4品種を用いた。ポットには1/2000aワグネルポットを用い、上記土壌を1ポットあたり約8L詰め、各土壌で5ポットずつ、4品種で計40ポットを準備した。数日間静置後、品種ごとに4粒播きで各ポット内2カ所に播種した(各ポット8粒使用)。初生葉展開期に1カ所につき1本に間引きし、各ポット2株立てとした。播種は、2017年7月4日におこない、収穫は10月以降、成熟期に達し収穫適期と判断した株から順次収穫をおこなった。

10

【0037】

2. 葉のサンプリング

葉のサンプリングは、それぞれの栽培試験において播種後28~32日となる日の日中に実施した(おおむね10時~15時)。この際のダイズ生育ステージは年度、栽培条件、品種により若干異なるが概ね葉齢5~7程度であった。ここでいう葉齢は、初生葉を1とした際に最上位に展開した本葉が下から数えて何枚目かを数えた際の値とした。葉のサンプリングは、最上位に展開する本葉の1または2葉齢古い本葉を構成する複葉3枚の内、中央の複葉を採取した。しかしながら、中央の複葉が虫害等著しく損傷を受けている場合は、別の複葉を採取した。採取した葉はアルミホイルで包み直ちに液体窒素中で凍結し、代謝反応を停止させた。凍結サンプルは凍結状態を維持したまま実験室へ持ち帰り、凍結乾燥にかけて乾燥させた。この乾燥したサンプルを後述の抽出操作に供試した。なお、2015年及び2016年のポット栽培試験では1個体ごとに採取し、収量データは対応する個体のものを用いた。一方、2017年ポット栽培試験では1ポット毎、すなわち2個体毎にまとめてサンプリングを実施し、収量データは2個体の平均値を用いた。

20

また、葉をサンプリングした際の播種からの日数は以下のとおりである。

* 2015年ポット試験(1)：2015年7月25日(播種後30日)

* 2015年ポット試験(2)：2015年7月29日(播種後28日)

* 2016年ポット試験：2016年8月2日(播種後32日)

* 2017年ポット試験：2017年8月3日(播種後30日)

30

【0038】

3. 最終的な子実収量の測定

栽培試験後の各個体から全子実を回収し、80℃にて2~3日間乾燥させた。収量データはこの乾燥重量(g DW/個体)を用いた。2で既述したように2017年の試験における2個体(1ポット毎)の平均データは1つとしてカウントし、2015~17年のポット試験データは計125個となった。収量データは表2a~2cに示すように最小で0.9 g DW/個体、最大で42.5 g DW/個体であった。

【0039】

40

【表 2 a】

試験名	サンプル名	年度	条件	品種	乾燥子実重(g/株)
2015年ポット試験(1)	T1	2015(1)	窒素量:3	里のほほえみ	30.6
	T10	2015(1)	窒素量:3	里のほほえみ	27.1
	T2	2015(1)	窒素量:3	里のほほえみ	28.0
	T3	2015(1)	窒素量:3	里のほほえみ	28.4
	T4	2015(1)	窒素量:3	里のほほえみ	26.6
	T5	2015(1)	窒素量:3	里のほほえみ	29.3
	T6	2015(1)	窒素量:3	里のほほえみ	27.0
	T7	2015(1)	窒素量:3	里のほほえみ	25.3
	T8	2015(1)	窒素量:3	里のほほえみ	27.1
	T9	2015(1)	窒素量:3	里のほほえみ	23.8
2015年ポット試験(2)	N1-1	2015(2)	窒素量:1	フクユタカ	11.0
	N1-2	2015(2)	窒素量:1	フクユタカ	11.3
	N1-3	2015(2)	窒素量:1	フクユタカ	13.0
	N1-4	2015(2)	窒素量:1	フクユタカ	10.3
	N1-5	2015(2)	窒素量:1	フクユタカ	19.4
	N1-6	2015(2)	窒素量:1	フクユタカ	13.9
	N1-7	2015(2)	窒素量:1	フクユタカ	11.8
	N1-8	2015(2)	窒素量:1	フクユタカ	16.3
	N1-9	2015(2)	窒素量:1	フクユタカ	17.7
	N3-1	2015(2)	窒素量:3	フクユタカ	12.8
	N3-10	2015(2)	窒素量:3	フクユタカ	7.5
	N3-2	2015(2)	窒素量:3	フクユタカ	9.4
	N3-3	2015(2)	窒素量:3	フクユタカ	15.2
	N3-4	2015(2)	窒素量:3	フクユタカ	11.8
	N3-5	2015(2)	窒素量:3	フクユタカ	10.1
	N3-6	2015(2)	窒素量:3	フクユタカ	15.8
	N3-7	2015(2)	窒素量:3	フクユタカ	8.3
	N3-8	2015(2)	窒素量:3	フクユタカ	11.2
	N3-9	2015(2)	窒素量:3	フクユタカ	5.9
	N10-1	2015(2)	窒素量:10	フクユタカ	17.1
N10-10	2015(2)	窒素量:10	フクユタカ	11.0	
N10-2	2015(2)	窒素量:10	フクユタカ	13.5	
N10-3	2015(2)	窒素量:10	フクユタカ	8.3	
N10-4	2015(2)	窒素量:10	フクユタカ	13.3	
N10-5	2015(2)	窒素量:10	フクユタカ	14.6	
N10-6	2015(2)	窒素量:10	フクユタカ	10.8	
N10-7	2015(2)	窒素量:10	フクユタカ	9.8	
N10-8	2015(2)	窒素量:10	フクユタカ	15.3	
N10-9	2015(2)	窒素量:10	フクユタカ	17.0	

【 0 0 4 0 】

【表 2 b】

試験名	サンプル名	年度	品種	乾燥子実重(g/株)
2016年ポット試験	KG1	2016	里のほほえみ	30.3
	KG2	2016	里のほほえみ	33.0
	KG3	2016	里のほほえみ	29.8
	KG4	2016	里のほほえみ	23.1
	KG5	2016	里のほほえみ	26.8
	KG6	2016	里のほほえみ	24.9
	KG7	2016	里のほほえみ	28.9
	KG8	2016	里のほほえみ	28.4
	KG9	2016	里のほほえみ	23.0
	KG10	2016	里のほほえみ	24.4
	MG1	2016	里のほほえみ	29.5
	MG2	2016	里のほほえみ	35.4
	MG3	2016	里のほほえみ	31.9
	MG4	2016	里のほほえみ	32.2
	MG5	2016	里のほほえみ	27.0
	MG6	2016	里のほほえみ	29.6
	MG7	2016	里のほほえみ	26.8
	MG8	2016	里のほほえみ	34.9
	MG9	2016	里のほほえみ	34.8
	MG10	2016	里のほほえみ	35.7
	TC1	2016	里のほほえみ	22.9
	TC2	2016	里のほほえみ	29.3
	TC3	2016	里のほほえみ	22.2
	TC4	2016	里のほほえみ	21.9
	TC5	2016	里のほほえみ	21.1
	TC6	2016	里のほほえみ	25.7
	TC7	2016	里のほほえみ	21.9
	TC8	2016	里のほほえみ	26.8
	TC9	2016	里のほほえみ	28.6
	TC10	2016	里のほほえみ	32.1
	TC11	2016	里のほほえみ	28.6
	TC12	2016	里のほほえみ	25.8
	TC13	2016	里のほほえみ	32.9
	TC14	2016	里のほほえみ	29.6
	TC15	2016	里のほほえみ	16.5
	TC16	2016	里のほほえみ	18.7
	WK1	2016	里のほほえみ	37.1
	WK2	2016	里のほほえみ	29.3
	WK3	2016	里のほほえみ	31.9
	WK4	2016	里のほほえみ	31.7
WK5	2016	里のほほえみ	16.1	
WK6	2016	里のほほえみ	42.5	
WK7	2016	里のほほえみ	33.4	
WK8	2016	里のほほえみ	27.5	
WK9	2016	里のほほえみ	29.0	
WK10	2016	里のほほえみ	38.4	

10

20

30

40

【表 2 c】

試験名	サンプル名	年度	条件	品種	乾燥子実重(g/株)
2017年ポット試験	Y1/2_1	2017	1/2x 肥料	湯上がり娘	5.4
	Y1/2_2	2017	1/2x 肥料	湯上がり娘	4.3
	Y1/2_3	2017	1/2x 肥料	湯上がり娘	4.4
	Y1/2_4	2017	1/2x 肥料	湯上がり娘	5.4
	Y1/2_5	2017	1/2x 肥料	湯上がり娘	4.3
	Y1_1	2017	1x 肥料	湯上がり娘	13.2
	Y1_2	2017	1x 肥料	湯上がり娘	14.3
	Y1_3	2017	1x 肥料	湯上がり娘	9.3
	Y1_4	2017	1x 肥料	湯上がり娘	11.1
	Y1_5	2017	1x 肥料	湯上がり娘	12.4
	F1/2_1	2017	1/2x 肥料	フクユタカ	5.5
	F1/2_2	2017	1/2x 肥料	フクユタカ	7.1
	F1/2_3	2017	1/2x 肥料	フクユタカ	9.8
	F1/2_4	2017	1/2x 肥料	フクユタカ	7.9
	F1/2_5	2017	1/2x 肥料	フクユタカ	5.9
	F1_1	2017	1x 肥料	フクユタカ	14.2
	F1_2	2017	1x 肥料	フクユタカ	17.1
	F1_3	2017	1x 肥料	フクユタカ	17.0
	F1_4	2017	1x 肥料	フクユタカ	17.7
	F1_5	2017	1x 肥料	フクユタカ	5.1
	S1/2_1	2017	1/2x 肥料	里のほほえみ	6.1
	S1/2_2	2017	1/2x 肥料	里のほほえみ	8.0
	S1/2_3	2017	1/2x 肥料	里のほほえみ	10.3
	S1/2_4	2017	1/2x 肥料	里のほほえみ	8.7
	S1/2_5	2017	1/2x 肥料	里のほほえみ	8.7
S1_1	2017	1x 肥料	里のほほえみ	15.6	
S1_2	2017	1x 肥料	里のほほえみ	19.8	
S1_3	2017	1x 肥料	里のほほえみ	19.2	
S1_4	2017	1x 肥料	里のほほえみ	18.1	
S1_5	2017	1x 肥料	里のほほえみ	14.5	
E1/2_1	2017	1/2x 肥料	エンレイ	8.3	
E1/2_2	2017	1/2x 肥料	エンレイ	7.6	
E1/2_3	2017	1/2x 肥料	エンレイ	7.2	
E1/2_4	2017	1/2x 肥料	エンレイ	9.6	
E1/2_5	2017	1/2x 肥料	エンレイ	0.9	
E1_1	2017	1x 肥料	エンレイ	11.4	
E1_2	2017	1x 肥料	エンレイ	16.4	
E1_3	2017	1x 肥料	エンレイ	15.7	
E1_4	2017	1x 肥料	エンレイ	14.8	
E1_5	2017	1x 肥料	エンレイ	16.5	

10

20

30

40

【 0 0 4 2 】

4. 採取した葉の成分の抽出

凍結乾燥した葉サンプルは、スパーテルを用いて手作業にて可能な限り粉砕をおこなった。粉砕後、2 mL のチューブ（セーフロックチューブ，エッペンドルフ）に 10 mg を秤量し、直径 5 mm のジルコニア製ボール 1 つをチューブに加えて、ビーズ粉砕機（MM 400，Ret sch）にて 25 Hz で 1 分間さらに粉砕をおこなった。ここへ、メタノール（HPLC グレード，関東化学）、純水（milli Q，メルクミリポア）、クロロホルム（特級，関東化学）の混合溶媒（5：2：2，v/v/v）を 1 mL 加え、37 で

50

30分間抽出をおこなった。抽出後、2000×g程度の卓上遠心機(チビタン)にて5分間遠心分離をおこない、固形物と分離した。遠心分離後、上清600μLを1.5mLチューブに分注し、そこへ純水300μLを添加した。ボルテックスを用いて混合した後、再度5分間遠心分離をおこなった。遠心後の上清400μLを新たに1.5mLチューブに分注した。データのクオリティチェック及び保持時間の補正に用いたQCサンプルは遠心分離後の各サンプルから400μLとは別に200μLずつ集約し、それより400μLを別途1.5mLチューブに分注した。各サンプルに残存するメタノールをスピードバック(SC-210A, ThermoScientific)で10分間、濃縮をおこなった。濃縮後、残存した水分を液体窒素にて凍結し、一晚凍結乾燥をおこなった。

【0043】

5. 誘導体化

4で得られた抽出物に、5μLのミリスチン酸-d₂₇(Cambridge Isotope Laboratories)のピリジン溶液(50μg/mL)を内部標準(後述のライブラリ照合用)として添加した。さらに別の内部標準(サンプル間のピーク強度補正)としてリピトール(東京化成工業)のピリジン溶液(50μg/mL)を5μL加えた。さらに、100μLのメトキシアミン塩酸塩のピリジン溶液(20mg/mL)を添加し、37で90分間インキュベーションをおこなった(第1の誘導体化「メトキシム化」)。インキュベート後、卓上遠心機で不溶物と分離後、上清50μLをバイアルビンに分注した。次にGCシステムのオートサンプラー(Agilent 7693)に搭載の自動分注器を用いて、第2の誘導体化をおこなった。メトキシム化後のピリジン溶液が入ったバイアルに50μLのN-メチル-N-(トリメチルシリル)トリフルオロアセタミド(MSTFA: ThermoScientific)を分析前に自動で添加するようにプログラムを組み、加温部分にて37で30分間インキュベート後(第2の誘導体化「シリル化」)、そのままGC-MS分析をおこなった。なお、溶媒に用いたピリジンは脱水ピリジン(関東化学)を用いた。

【0044】

6. GC-MS分析

5で得られた1μLの分析サンプルを、スプリットモードでGC-MSに注入した(10:1, v/v)。GC-MS装置は、Agilent 7890B GCシステム(Agilent)、シングル四重極分析計Agilent 5977B GC/MSD(Agilent)ならびにAgilent 7693Aオートサンプラー(Agilent)を連結したものをを用いた。このシステムに0.25μmの膜厚をもち30mの分析カラムに10mのガードカラムが組み込まれた40m×0.25mm I.Dのヒューズドシリカキャピラリーカラム(ZORBAX DB5-MS+10m Duragard Capillary Column, 122-5532G, Agilent)を装着して用いた。注入温度は、250とし、カラムを通るヘリウムガスの流速は0.7mL/minに設定した。カラム温度は、1分間60で等温に保ち、次いで10/分で325まで上昇させ、そして10分間等温に保った。搬送ライン及びイオンソース温度は、それぞれ300及び230とした。イオンを(70.0kV)電子衝撃(EI)によって生成した上で、m/z50-600の質量範囲にわたって1秒当たり2.7回のスキャンを行ってデータを記録した。加速電圧は、5.9分の溶媒遅延後に作動させた。

なお、QCサンプルはサンプルの分析サンプルの5-6分析毎に1つ入れた。

【0045】

7. データ行列の作成

GC-MSで取得したデータは、netCDFファイル形式に変換し、ピーク検出及びアラインメントソフトウェアであるMetAlign(Wageningen Univ.)でデータ行列を作製した。サンプル間の各ピークの保持時間及びピーク強度の補正には、別途、内部標準として添加したリピトールを指標におこなった。

最終的にデータ行列はcsv形式で作成した。そこへ各サイズサンプルの収量値を統合し、データ解析に用いた。

10

20

30

40

50

【 0 0 4 6 】

8 . ピーク同定

GC - MS分析によって得られた分析データは、内部標準物質として添加したミリスチン酸d 27の保持時間を基準としてAgilent社製のFiehnメタボロミクスRTLライブラリ(以下、Fiehnライブラリ)を用いてピーク同定をおこなった。ピーク同定の基準はFiehnライブラリによるスペクトル一致率が70%以上の候補代謝物から保持時間の誤差が0.05分以内であることとした。

【 0 0 4 7 】

9 . 収量と関連する成分候補の探索

2016年データを用いて収量と関連する成分候補の探索を行った。表2に示す20サンプルを用いて、葉中成分由来のピーク及び各サンプルに対応する収量データのマトリックスデータを用いて解析をおこなった。同一と考えられる成分由来ピークを保持時間により整理・重複を除外したのち、各成分由来ピークデータと収量との関連解析をおこなった。

【 0 0 4 8 】

【表2】

サンプル名	年度	品種	乾燥子実重量(g/個体)
KG2	2016	里のほほえみ	32.97
KG4	2016	里のほほえみ	23.1
KG6	2016	里のほほえみ	24.87
KG8	2016	里のほほえみ	28.42
MG2	2016	里のほほえみ	35.39
MG4	2016	里のほほえみ	32.21
MG5	2016	里のほほえみ	26.97
MG6	2016	里のほほえみ	29.61
MG7	2016	里のほほえみ	26.8
MG9	2016	里のほほえみ	34.75
TC1	2016	里のほほえみ	22.87
TC4	2016	里のほほえみ	21.94
TC5	2016	里のほほえみ	21.06
TC12	2016	里のほほえみ	25.77
TC15	2016	里のほほえみ	16.52
WK1	2016	里のほほえみ	37.08
WK3	2016	里のほほえみ	31.91
WK4	2016	里のほほえみ	31.74
WK6	2016	里のほほえみ	42.53
WK9	2016	里のほほえみ	29

【 0 0 4 9 】

各ピークデータと収量との関連解析では、単相関係数r及び無相関の検定によりp値を算出した。また各成分由来ピーク情報(保持時間、質量情報)からFiehnライブラリと照合することで、収量と関連すると同定した成分の候補として9種の成分、すなわち、2-ヒドロキシピリジン、コリン、クエン酸、グリセリン酸、グリシン、L-ピログルタミン酸、マロン酸、スクロース及びトレイトールを選抜した。選抜した9種の成分を用いて、残りのサンプルも含め、2015-2017年の全サンプル125サンプルの解析を行った。

【 0 0 5 0 】

まず、全分析データを用いて、各成分相対量と収量との関連解析を行った。

既存技術では、ダイズの生育過程における指標を用いて、収量との関連性を解析した研

究が報告されている。例えば、非特許文献（「重粘度地帯におけるダイズ「エンレイ」の多収事例に基づく収量構成要素と生育指標」新潟農総研）では、播種後40日における主茎長が収量と相関し、その相関係数 r が0.51であることが示されており、また非特許文献6（「土壌理化学性の改善による麦跡作付地帯での大豆多収栽培体系の確立」福井県農試、福井県大）では播種後50-60日における根粒活性が収量と相関し、その相関係数が0.63であることが示されており、さらに別の既報（「物質生産と窒素の蓄積から見たダイズの多収生育相」福井県農試）では播種後60-70日における地上部乾物重が収量と相関し、その相関係数が0.66であることが示されている。これらの知見は、上述した播種後日数で特定の指標を評価することで一定の収量予測が可能であることを示している。しかしながら、本技術における播種後30日程度という予測時期はいずれの従来技術よりも早いことから、少なくとも相関係数の絶対値が0.51より大きい成分由来ピークは従来技術よりも優れた収量予測指標であると言える。

解析の結果、9種の成分の収量との相関は表3に示すようになり、相関係数の絶対値が0.51を超えるものとして、2-ヒドロキシピリジン、グリシン、L-ピログルタミン酸及びスクロースの4種が見いだされた。

【0051】

【表3】

成分	収量との相関 r	p
2-ヒドロキシピリジン	-0.52	0.000
コリン	0.01	0.945
クエン酸	-0.38	0.000
グリセリン酸	-0.10	0.288
グリシン	-0.58	0.000
L-ピログルタミン酸	-0.57	0.000
マロン酸	0.41	0.000
スクロース	0.57	0.000
トレイトール	-0.23	0.011

【0052】

10. モデル構築・評価

2つ以上の複数の成分の分析データを用いた収量予測モデルの構築には多変量解析手法を用い、解析ツールとしてSIMCA ver. 14 (Umetrics)を用いた。予測モデルは、説明変数に各成分の分析データのピーク面積値を、また目的変数に収量値を用いた回帰分析をおこなった。回帰分析はPLS法の改良版であるOPLS法でおこなった。

【0053】

予測モデルの評価方法は、主に2つの指標で判断される。1つは予測精度を表す R^2 、もう1つは予測性を表す Q^2 である。 R^2 は予測モデル構築に使用したデータの実測値とモデルで計算した予測値との相関係数の2乗であり、1に近いほど予測精度が高いことを示している。一方、 Q^2 は、上記クロスバリデーションの結果であり、実測値と繰り返し実施したモデル検証の結果である予測値との相関係数の2乗を表している。予測の観点から、少なくとも $Q^2 > 0.50$ であれば、そのモデルは良好な予測性を持つとされていることから (Triba, M. N. et al., Mol. BioSyst. 2015, 11, 13-19.)、 $Q^2 > 0.50$ をモデル評価の基準とした。なお、常に $R^2 > Q^2$ となるため、 $Q^2 > 0.50$ は同時に $R^2 > 0.50$ を満たすこととなる。

【0054】

10

20

30

40

50

10-1. 全データを用いたモデルの構築・評価

1 データ当たり9成分の相対量と収量値を持ち、全125個のデータマトリックスから、収量を予測するOPLSモデルを構築した。構築の際、各成分由来ピーク及び収量データはオートスケーリングにより平均0、分散1に変換した。モデル構築の結果、予測精度を示す $R^2 = 0.56$ 、予測性を示す $Q^2 = 0.55$ であった。結果を図1に示す。モデルの予測性能の基準として設定した $Q^2 > 0.50$ を満たしていた。この予測モデルにより、栽培1カ月程度の葉に含まれる成分組成を用いることで、一定の予測性能を持つモデルが構築でき、早期収量予測が可能であることが示された。

【0055】

10-2. VIP値の算出

10-1で構築したモデルではVIP (Variable Importance in the Projection, 投影における変数重要性) 値とよばれる各成分由来ピークに与えられるモデル性能への寄与度と与えられる。VIP値はその値が大きいほどモデルへの寄与度が大きく、相関係数の絶対値とも相関する。VIP値のリストを表4に示す。

【0056】

【表4】

成分	VIP値	VIP値rank
L-ピログルタミン酸	1.44	1
グリシン	1.35	2
2-ヒドロキシピリジン	1.24	3
スクロース	1.20	4
マロン酸	0.98	5
クエン酸	0.92	6
トレイトール	0.50	7
グリセリン酸	0.26	8
コリン	0.11	9

【0057】

10-3. VIP値を指標としたモデル構築 (2個以上の成分ピーク情報を用いたモデル)

10-1で構築したモデルへの各成分由来ピークの寄与度であるVIP値のランキング (10-2) を基に複数の成分でモデルを構築した。特に限定されるわけではないが、モデル性能の基準を便宜上 $Q^2 > 0.50$ とした。

【0058】

10-3-1. VIP値が下位の成分由来ピークを用いたモデル

VIP値1位以下、2位以下、3位以下、4位以下、5位以下及び6位以下のすべての成分データを用いてそれぞれOPLSモデルの構築をおこなった。その結果、 $Q^2 > 0.50$ を満たすのはVIP値1位以下すべて～3位以下すべての成分データを用いたモデルであり、VIP値4位以下の成分データすべてを用いても $Q^2 > 0.50$ とはならないことがわかった (図2)。

【0059】

10-3-2. VIP値上位の成分データを複数個用いたモデル

VIP値上位から順に複数個の成分データ、すなわち、VIP値1位及び2位、VIP値1位～3位及びVIP値1位～4位の成分データを用いてOPLSモデルの構築をおこなった。その結果、上位3成分では $Q^2 > 0.50$ を満たさず、上位4位まで使用するこ

10

20

30

40

50

とで $Q^2 > 0.50$ を満たすことがわかった。このことからモデル構築の場合は、少なくとも、9個の中から4つ以上の成分データを用いる必要があることが示唆された(図3)

すなわち、9個の成分データから選ばれる任意の4個以上の成分データを用いて構築したモデルのうち、 $Q^2 > 0.50$ を満たすモデルは予測性があると判断される。具体的には、以下のモデルが挙げられる。

【0060】

1) VIP値上位から順に連続する4個の成分データを用いたモデル

VIP値上位1位~4位、2位~5位、3位~6位、4位~7位、5位~8位及び6位~9位の成分データを用いて、それぞれOLSモデルの構築をおこなった。その結果、VIP値4位~7位の成分データを用いた際のモデルで初めて $Q^2 > 0.50$ を満たさなかった。それ以降 Q^2 は低下する傾向であった。このことから、VIP値6位以上であればその中から任意の4個の成分データを用いることで Q^2 の基準を概ね満たすが、VIP値4位以下から選ばれる任意の成分データ4個のみでは基準を満たさないことが示唆された(図4)。

10

【0061】

2) VIP値上位から順に連続する5個の成分データを用いたモデル

VIP値上位1位~5位、2位~6位、3位~7位、4位~8位及び5位~9位の成分データを用いて、それぞれOLSモデルの構築をおこなった。その結果、VIP値4位~8位の成分データを用いた際のモデルで初めて $Q^2 > 0.50$ を満たさなかった。それ以降 Q^2 は低下する傾向であった。このことから、VIP値7位以上であればその中から任意の5個の成分データを用いることで Q^2 の基準を概ね満たすが、VIP値4位以下から選ばれる任意の成分データ5個のみでは基準を満たさないことが示唆された(図5)。

20

【0062】

3) VIP値上位から順に連続する6個の成分データを用いたモデル

VIP値上位1位~6位、2位~7位、3位~8位及び4位~9位の成分データを用いて、それぞれOLSモデルの構築をおこなった。その結果、VIP値4位~9位の成分データを用いた際のモデルで初めて $Q^2 > 0.50$ を満たさなかった。このことから、VIP値8位以上であればその中から任意の6個の成分データを用いることで Q^2 の基準を概ね満たすが、VIP値4位以下から選ばれる任意の成分データ6個のみ、すなわち4位以下すべてでは基準を満たさないことが分かった(図6、図2)。

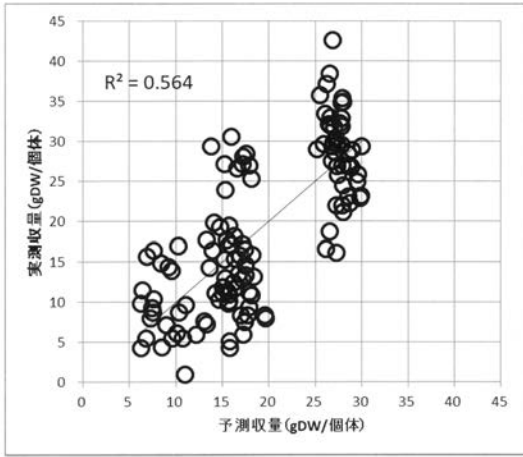
30

【0063】

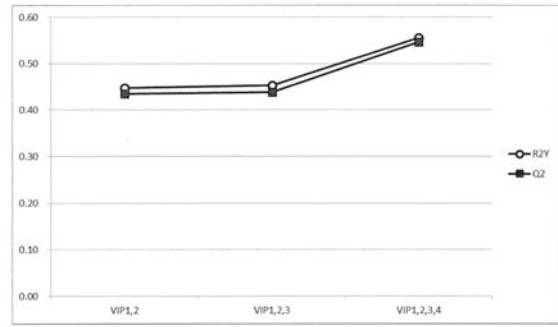
4) VIP値上位から順に連続する7個の成分データを用いたモデル

VIP値上位1位~7位、2位~8位及び3位~9位の成分データを用いて、それぞれOLSモデルの構築をおこなった。その結果、すべてのモデルで $Q^2 > 0.50$ を満たした。このことから、9個の成分データのうち任意の7個以上の成分データを用いることで Q^2 の基準を概ね満たすことが分かった(図7)。

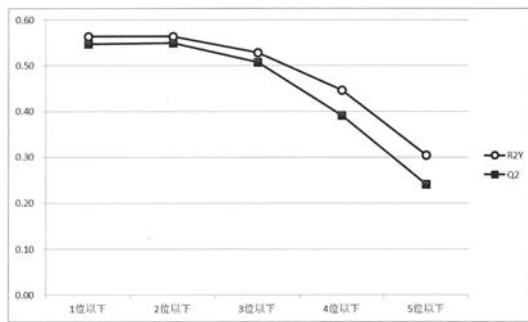
【 図 1 】



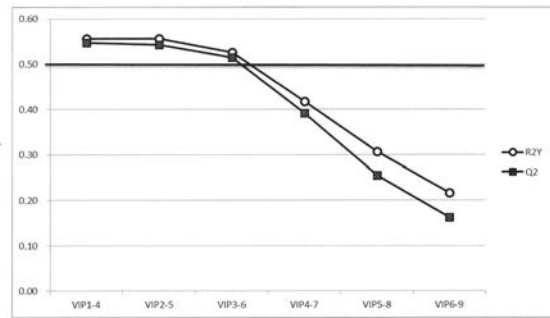
【 図 3 】



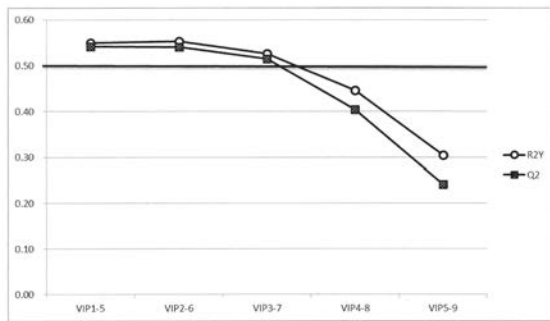
【 図 2 】



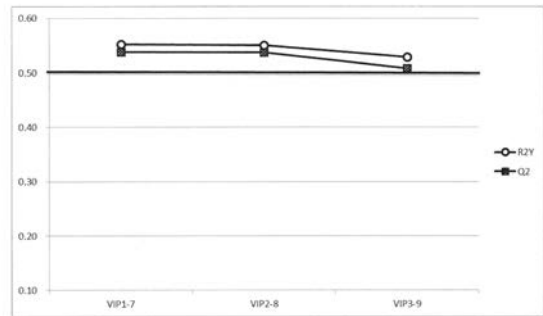
【 図 4 】



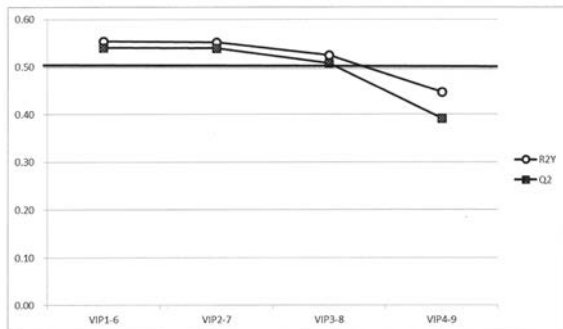
【 図 5 】



【 図 7 】



【 図 6 】



フロントページの続き

(72)発明者 遠藤 圭二

栃木県芳賀郡市貝町赤羽2606 花王株式会社研究所内

Fターム(参考) 2G041 CA01 EA04 EA06 FA06 HA01 LA07 LA08