

水稻平年収量等に関する研究会報告

平成13年 7 月

水稻平年収量等に関する研究会

目 次

平年収量等に関する研究会設置の背景	1
平年収量関係	1
1 現状と課題	1
2 算定方法の改善について	3
3 今後引き続き検討を要する事項について	4
収量予測関係	6
1 現状と課題	6
2 重回帰予測式を用いた予測手法の改善について	8
3 今後引き続き検討を要する事項について	10
参考資料	
1 平年収量の計算方法	14
2 収量予測時期別生育ステージ及び実測・推定項目	26
付属資料.....	27
1 水稻平年収量等に関する研究会開催要領	27
2 研究会における検討経過	29
3 研究会における委員からの発表テーマ一覧.....	30
別添資料：水稻平年収量等に関する研究会におけるシミュレーション結果（抜粋）	
別冊資料：水稻平年収量等に関する研究会における委員からの発表の概要	

平年収量等に関する研究会設置の背景

水稻の10 a 当たり平年収量は、生産量統計において作柄の良否等を分かりやすく表示する基準として、また、農業共済事業における基準収穫量の基礎あるいは米需給計画等農政の基礎資料として利用されることを目的として算定されており、従来からその精度向上に努力してきた。また、作況調査においては、ほ場調査の結果を基に、気象条件等を用いて的確な収量の予測に努めてきた。

近年、米の需給調整を含め、農家の総合的な経営・所得安定対策の構築が農政の重要な課題となっており、従来にも増して的確な平年収量の算定と適時・的確な収量予測情報の提供が求められている。

このような状況を踏まえ、より一層の精度向上を図る見地から、最新の研究成果や最先端の技術を活用し、いかなる改善が可能であるかについて、平成13年2月に学識経験者等で構成する「水稻平年収量等に関する研究会」を設置し、検討を進めてきた。

研究会においては、委員の知見に基づく意見・提案の発表を受け、可能な事項については模擬的計算（シミュレーション）により検証を行いつつ検討を行い、その結果を平年収量算定及び収量予測の精度向上のための提言として以下のとおり取りまとめた。

平年収量関係

1 現状と課題

(1) 現 状

水稻10 a 当たり平年収量は、「作物の栽培を開始する以前に、その年の気象の推移や被害の発生状況を平年並みとみなし、最近の栽培技術の進歩の度合や作付変動を考慮し、実収量のすう勢を基にして作成した、その年に予想される10 a 当たりの収量」と定義され、全国及び都道府県別に算定されている。

水稻平年収量の算定方法については、平成8年まで採用されていた単収と被害率のデータを用いた平方根重回帰式による方法に代わり、回帰手法として生物学・計量経済学・医学分野などで近年幅広く活用されているスプライン関数^(注1)を用いた方式が平成9年から採用されている。

この方式は、

ア 実単収が、気象によらず年次により傾向的に変化する部分（平年

収量部分)と気象による増減部分の和によって構成される。

イ 実単収のうち、年次により傾向的に変化する部分については、年次を説明変数とするスプライン関数を、気象による増減部分については、気象要因を説明変数とする式を用いている。

ウ これに過年次のデータを当てはめて係数を推定した上で、平年収量を推定する年次に係る気象データに平年値(計測期間の平均値)を代入して平年収量を推定する。

という方式である。

この場合、スプライン関数を用いることで、直線回帰や平方根回帰のように変化の方向を固定することなく、近年のすう勢を的確に反映することができる。

なお、気象データは、気象庁によるアメダス観測データである降水量、平均気温、最高気温、最低気温、日照時間、最大風速の6項目について、都道府県ごとに収穫日の前後各40日間(計81日間)^(注2)の平均データを用いている。また、重回帰分析で問題となる多重共線性^(注3)を回避するため、気象データの主成分分析^(注4)を行い、その主成分(気象指数)を説明変数としている。

なお、算定に用いるデータ数(年数)については、現行のアメダス観測体制が整った昭和54年以降のデータが用いられている。

(2) 課 題

現行の算定方法は、従来の方法に比べ、近年の10a当たり収量のすう勢を的確に反映することができる手法であるが、一層の予測精度の向上のために、検討すべき課題として、以下の提案があった。

ア 現行の算定方法においては、気象要素として、「降水量」「平均気温」「最高気温」「最低気温」「日照時間」「最大風速」を用いているが、作物学的知見からみると、より適切な気象データの選択・加工の方途があるのではないか。

イ 算定に用いる気象データは、いずれも収穫日の前後各40日間の平均値を採用しているため、短期間の大きな気象変動(低温、台風等)による収量への影響が、十分にモデルに反映できない場合があるのではないか。

ウ 現行の主成分(気象指数)を説明変数とする方法では、気象と収量との関係が曖昧になり、作物学的知見に基づくモデルの適否の判断や作物

学の研究成果等の反映が難しいため、直接、気象要素を説明変数とするモデルを作成すべきではないか。

エ 算定に用いることのできるデータ数が限られているので、この限られたデータをより有効に用いたモデルを作成すべきではないか。

2 算定方法の改善について

1の(2)の課題に対し、全県または代表的な県において以下の具体的な検証・検討を行った。

(1) 推定日射量の適用

水稻の光合成に不可欠な太陽エネルギーの量を説明する変数としては、「日照時間」よりも曇天時をも含めた「日射量」を利用することが適切であるとされており、また、日照時間のデータを用いて日射量を推定する計算手法は現在確立されているとの提案があったことから、日照時間から変換した『推定日射量^(注：5)』を適用することについて検討を行った。

この推定日射量は、実測された日射量との相関も極めて高く、平年収量の算定式において日照時間に代えて全ての県について検証した結果、ほとんどの県において予測精度の向上（GCV^(注：6)が小さくなること）がみられた。

よって、「日照時間」に代えて、『推定日射量』を用いることは有効であり、直ちに適用できる改善事項と考えられる。

(2) 補助変数の適用

短期間の気象変動（低温・台風等）による減収については、その要因を分析した上で、要因となった気象要素を『補助変数^(注：7)』として直接モデルに取り入れ、主成分（気象指数）と合わせて気象による収量の増減部分を説明することにより予測精度の向上を図ることが可能であり、以下の補助変数を適用すべきではないかとの提案があった。

台風等強い風雨による被害の主な要因である倒伏などを説明する変数としては、風圧の影響を表す「日最大風速の2乗値^(注：8)」を用いること。

出穂前の強い低温による減収を説明する変数として、「低温積算値^(注：9)」を用いること。

これらの補助変数を適用し、代表的な数県において検証を行った結果、いずれの場合においても予測精度の向上が見られた。

よって、台風等による減収については「日最大風速の2乗値」を、出穂前の一時的な低温による減収については「低温積算値」を、それぞれ『補助変数』として用いることは有効であり、直ちに適用できる改善事項と考えられる。

なお、同様に、雨を伴った風の被害、高温・寡照による登熟障害、いもち病等についても適切な補助変数を用いれば改善がさらに可能ではないかとの指摘があった。

3 今後引き続き検討を要する事項について

(1) 気象要素を直接用いる手法

作物学的知見に基づくモデルの適否の判断や各気象要素と収量との関係を明確にし、平年収量を高精度に推定するために、気象要素を直接説明変数として取り入れるべきとの提案があった。その場合、説明変数間の多重共線性の問題を回避するために、リッジ回帰^(注：10)の手法を加味したモデルを検討すべきとの提案があった。

全ての県についてリッジ回帰の手法を加味した検証を行った結果、ほとんどの県において、気象要素と収量の関係を示す係数の傾向について、作物学上の知見と矛盾が無いものとなり、予測精度についても現行とほぼ同程度、又はそれ以上となった。

しかし、現行方式以上に予測精度を高めるためには、説明変数に用いる気象要素（適用期間、データの加工を含む）の選定、回帰係数・各パラメータ（平滑化パラメータ^(注：11)、リッジ回帰の定数^(注：12)）の推定方法等について、さらなる検討が必要であると考えられる。

(2) 県を統合・分割して推定する手法

限られたデータを有効に活用し、予測精度の向上とモデルの妥当性を高めるため、「隣接県を統合するモデル」や「一つの県を小地域に分割したモデル」について検討すべきとの提案があった。

隣接する県を統合するモデルは、隣接する県で自然条件や栽培技術等の稲作条件が似ている場合は、気象の収量に対する影響は似たものになる（気象要素の回帰係数の値が近いものとなる）ことを前提に、限られ

たデータの効率的な活用を図るものである。

一部の県について検証した結果、隣接する県における気象要素に対する回帰係数の値は近いものとなった。

しかし、隣接県として、どのような範囲を設定するか等は、隣接する県間の気象条件・作付品種・栽培方法等の稲作条件に共通性があるかなどについて作物学的知見を踏まえ、さらなる検討が必要であると考えられる。

一つの県を小地域に分割するモデルは、現在、使用可能な県内地帯区分別のデータを利用し、分割した地域のそれぞれの収量・気象データを用いて各地域のモデルを求めるものである。

一部の県について、これら各地域のモデルを県合計として足し合わせ、県全体の平均値（収量、気象データ）を用いて得られたモデルと比較検証した結果、小地域における気象と収量との関係が明確となり、県全体の予測精度の向上がみられた。

しかし、分割した地域ごとの気象条件や局地的被害を反映し、気象と収量との関係をより明確かつ適切に表すためには、アメダスデータをメッシュ化して活用するなどの検討がさらに必要である。

注1：「スプライン関数」とは、小区間内で各々定義された多項式（3次多項式が用いられることが多い）による関数で、互いにできるだけ滑らかにつながっているものをいう。

2：「出穂日の前後各40日間（計81日間）」とした根拠は、水稻の収量との関係が一般的に大きいとされる期間は、出穂前については、穂首分化期～出穂期までの約30日間であるが、品種や気象条件等によって±10日以上之差がみられることもあり、このことを考慮したこと、出穂後については、水稻の登熟期間は約40～50日間であり、登熟歩合と出穂後40日間の平均気温に高い相関が確認されていることによる。

3：「多重共線性」とは、回帰分析において、説明変数間に強い相関関係があることをいう。この場合、回帰係数の推定が不安定になり、回帰係数の符号が予想されたものと逆に推定されることもある。

4：「主成分分析」とは、多変量解析の一手法で、多くの変量の値をできるだけ情報の損失を少なくして、相互に相関のない（互いに直交している）少数個の総合的特性値（主成分）に要約する手法である。

5：「推定日射量」は、日照率と日射率の関係を三次式で表した回帰式に、日照時間の観測値と大気外日射量、日照時間を与えて算出される日射量である。

6：「GCV (Generalized Cross Validation)」とは、予測誤差を推定することを意図した統計量で、小さい方が予測誤差が小さいとされる。

- 7 : 「補助変数」として用いる気象データ（加工したものを含む）は、主成分の要素に含めず、直接、説明変数として用いた。
- 8 : 「日最大風速の2乗値」は、風圧に比例し、台風等による倒伏を表す指標として有効であり、データは、出穂期から収穫期までの期間で最大を示した日最大風速の2乗値（一定以上の風速のみ）とした。
- 9 : 「低温積算値」は、危険期とされる減数分裂期の低温の程度を表す指標として有効であり、データは、日平均気温が20℃を下回った日について、その下回った値の出穂前15日間の積算値とした。
- 10 : 「リッジ回帰」とは、重回帰分析における多重共線性の問題に対処するため、リッジ回帰の定数（リッジパラメータ）をモデルに組み入れ、回帰係数の安定性を高める手法である。
なお、リッジ回帰により推定された回帰係数は、偏りをもつが分散は小さくなり、安定性が高くなる。
- 11 : 「平滑化パラメータ」とは、回帰式の当てはまり具合とスプライン関数の滑らかさの程度を調整するパラメータであり、このパラメータが大きくなるとスプライン関数は凹凸の少ない滑らかなものとなるが、回帰式の当てはまり具合は悪くなる。
- 12 : 「リッジ回帰の定数（リッジパラメータ）」とは、回帰式の当てはまり具合と回帰係数の安定性を調整するパラメータである。
なお、リッジパラメータが大きくなると、回帰係数の安定性は増すが回帰式の当てはまり具合は悪くなる。

収量予測関係

1 現状と課題

(1) 現 状

現在、水稻収穫量調査における収量予測は、各地域の水稻の生育段階に依じて、7月15日現在の生育情報調査時（早期栽培のみ）、8月15日現在（早期栽培及び早場地帯のみ）及び9月15日現在の作柄概況調査時、10月15日現在の予想収穫量調査時の各調査時期に行っている。

現行の収量予測方法は、各都道府県一本及び県内地帯区分ごとに、収穫期を迎えた地域では刈取実測によるが、収穫前の地域では、穂数、もみ数、粗玄米粒数歩合などの収量構成要素を掛け合わせて収量を予測している。

これらの収量構成要素については、調査日時時点で計測可能な項目は実測値を用い、実測不可能な項目は、収量構成要素ごとに過去の調査結果から、それが決定される時期の気象データ等を説明変数にして作成した重回帰予測式を用いて、これに本年の収量構成要素の実測値、気象の実況値及び調査日時点以降の気象データとしては気象平年値を入力するこ

とにより推定している。

なお、重回帰予測式は、各統計情報事務所において、推定値と実測値の相関係数がより高まることを主眼として、説明変数に用いる収量構成要素、気象要素、気象データの適用期間などをそれぞれ選択し作成している。

(2) 課 題

水稻収穫量調査結果は、需給計画や生産技術対策等の施策に必要な基礎資料として、早い予測時期から高い精度が求められている。

しかし、現行の収量予測においては、出穂直後に穂数・もみ数を実測した後は刈取りまで実測を行っておらず、出穂直後から刈取直前までの間の各予測時期において、登熟に関わる指標である「千もみ当たり収量」をいかに精度良く推定するかが課題となっている。

このため、検討すべき課題として、以下の提案があった。

ア 現在、各調査時期の千もみ当たり収量の推定において、粗玄米粒数歩合、玄米粒数歩合、玄米千粒重の推定値を掛け合わせて推定しているが、この方法による推定値は、誤差が積み重なる恐れがあるのではないか。

イ 出穂期以降の予測に当たっては、もみ数のような実測値が得られれば直ちに重回帰予測式の説明変数として用いるとともに、収量に関するトレンドを表すデータや作物学的な知見に基づいた気象データ等を活用すべきではないか。

ウ 重回帰予測式作成に用いることができる過年次のデータ数が限られていることから、データを有効に活用できる方法を検討すべきではないか。

エ 出穂直後の予測に当たっては、新たに出穂前約2週間のCGR(1日当たり乾物重増加量)と収量に関する研究成果を活用すべきではないか。

オ 登熟期の予測に当たっては、千もみ当たり乾物重の推移と収量に関する研究成果を活用すべきではないか。

カ 最近の収量予測に関する研究成果を踏まえて、収量予測モデル、リモートセンシング技術の活用などの新たな分野の収量予測手法について検討すべきではないか。

キ 気象予報の活用等の可能性を含めて、幅を持たせた予測方法もあるのではないか。

2 重回帰予測式を用いた予測手法の改善について

1の(2)の課題に対し、低温の影響を受けやすい東北の「宮城県」、早場地帯の「栃木県」、遅場地帯の「大分県」の3県を対象に、以下の具体的な検証・検討を行った。

なお、検証・検討は、現行の予測式と提案に基づいて作成した予測式について、相関係数、標準誤差、分散比等の指標を比較することにより行った。

(1) 千もみ当たり収量の推定方法の改善

出穂後における現行の収量予測方法は、単位面積当たり全もみ数の実測値に、重回帰予測式によりそれぞれ推定した収量構成要素(粗玄米粒数歩合^(注:13)、玄米粒数歩合^(注:14)、玄米千粒重^(注:15))を掛け合わせて収量を予測^(注:16)しているが、この方法によると誤差が積み重なる恐れがある。

このため、「千もみ当たり収量^(注:17)」を重回帰予測式で直接求め^(注:18)、それに単位面積当たり全もみ数の実測値を乗じて算出した収量の予測値を検証した結果、より実測値に近い結果が得られた。

(2) 予測式に用いる説明変数の改善

現行では、気象データのみを説明変数として用いている場合が多いため、気象以外の要因(地力、病・虫害、雑草、栽培技術の進歩等)の効果が評価されず、誤差の要因になっていることから、その解消に向けた改善が必要である。

このため、説明変数に調査日時点で実測しているもみ数を用いるとともに、栽培技術の進歩等の経年的な変化を表すデータとして10a当たり平年収量の適用について検討することが適当である。

また、気象データを用いる場合には、作物学的な知見に基づいて、気象要素の種類とデータ期間の選択、加工(指数化)について検討することが適当である。

ア 作況調査の実測データ等

(ア) 1㎡当たり全もみ数

水稻の収量には、気象要因の他に土壌条件や栽培条件等の要因

も複合的に影響していることから、調査日時点までの生育の履歴を表すデータを加味することは、より精度の高い予測のために有効と考えられる。

出穂直後の早い予測時期において生育の履歴を表すと考えられるもみ数は、西南暖地の早期栽培で7月上旬、早場地帯で8月上中旬、遅場地帯で9月上旬に概ね実測が可能であるが、説明変数としての利用は一部に止まっていることから、実測値である「1 m² 当たり全もみ数」を説明変数に用いることについて検証した結果、精度の向上が図られた。

ただし、近年、栽培品種や栽培技術が大幅に変更される等の理由により、「もみ数」が急激に変化している都道府県や地帯においては、今回の研究会で検証した昭和54年以降の22年間のデータでは有効な説明変数とはならない場合があったことから、このような場合には、データの適用を急激な変化があった以降のデータを適用して検討することが適当である。

(イ) 10 a 当たり平年収量

水稻の収量には、栽培技術の進歩、基盤整備の進展等の経年的な上昇要因と短期的な気象要因が複合的に影響していることから、気象データのみを用いた予測では経年的な要因は説明できず誤差の要因になるので、その解消に向けた改善のためには、各調査時期の収量予測において新たに収量に関するトレンドを表すデータを加味することが有効と考えられる。

このトレンドを表すデータとして、「10 a 当たり平年収量」を説明変数に用いることについて検証した結果、精度の向上が図られた。

イ 気象データ等

(ア) 推定日射量、低温積算値、日最大風速の2乗値

の2の(1)及び(2)と同様の視点から、生育期間の「推定日射量」、出穂前30日間の「低温積算値」、出穂期から収穫期までの「日最大風速の2乗値」を用いることは有効と考えられることから、これらのデータを説明変数に用いることについて検証した結果、精度の向上が図られた。

なお、説明変数として、倒伏による収量低下をもたらす強い風と雨の複合効果について、気象データを適切に加工して適用することが可能ではないかとの指摘があった。

(1) 気温日較差

登熟期間における高夜温が収量低下に影響した年もあり、このような現象を収量予測に反映させることは、より精度の高い予測のために有効と考えられる。

このため、高夜温による収量低下を表す説明変数として、出穂後30日間の「気温日較差」を用いることについて検証した結果、精度の向上が図られた。

なお、日射量が多い場合の高夜温は減収への影響は少ないものの、日射量が少ない場合の高夜温は減収に結びつくことから、今後、寡照と高夜温の複合効果について、気象データを適切に加工して適用することが可能ではないかとの指摘があった。

(3) 県内地帯区分別の全データサンプルを活用した予測

現在、予測式の説明変数及び目的変数として用いているデータサンプルは、各都道府県、又は県内地帯区分という予測の対象地域ごとに、実測データを平均する等により、1年につき1つのデータサンプルを作成して用いている。このため、いずれの予測式も過年次のデータサンプルは、それぞれ22年分の22個に限られ、自由度が大きくないことから、予測式に取り入れる説明変数が制約され精度向上の限界となっている。

この制約を解決する方途として、各年の県内地帯区分ごとのデータサンプル全てを活用して(22年×県内地帯区分数のデータサンプル数が得られる)一本の予測式に適用することにより、全ての県内地帯区分に共通して適用できる予測式を作成することについて検証した結果、県内地帯区分ごとの予測値、それらを面積で加重平均した県単位の予測値ともに精度向上が図られた。この方法は、データサンプル数が増加し、自由度が大きくなることから、有効な各種の説明変数を追加することが可能となるものである。

3 今後引き続き検討を要する事項について

(1) 出穂前約2週間のCGR(1日当たり乾物重増加量)からの収量予測

出穂前約2週間のCGRは、この時期がもみ数や花粉の充実度合が決定される時期であり、またこの時期の光合成の能力が登熟の良否を左右するものとなることから、この時期のCGRを把握することは、出穂直後の時期におけるより精度の高い収量予測のために有効と考えられる。

このため、今後、このデータを集積し、その適用の可能性を検証することが望ましいと考えられる。

(2) 千もみ当たり乾物重増加状況からの収量予測

登熟の進度と相対千もみ当たり乾物重(最終千もみ当たり乾物重に対するそのときの千もみ当たり乾物重の比)の間には、倒伏や突発的な災害がなければ、年次をこえて一定の関係が存在することが考えられる。登熟の進度を気温の関数である発育指数(DVI)^(注:19)などの指数で的確に表すことにより、出穂後のある時点での千もみ当たり乾物重の測定値から、最終千もみ当たり乾物重を予測し、それともみ数の実測値から収量を予測することも可能と考えられる。

このため、登熟期の収量予測において、発育指数と千もみ当たり乾物重の推移に関するデータを集積し、これらのデータに基づいた登熟歩合の推定の可能性を検証することが望ましいと考えられる。

(3) 最新の研究成果等に基づく収量予測

ア 収量予測システムの実用化

水稻の生育過程における生理的、生態的及び物理的プロセスの解析を基に組み立てられた発育予測モデルの活用は、適用範囲(地域、年代)の制約がなく生育の進度と収量の逐次予測が可能な方法として提案された。

この発育予測モデルによる進度予測は、東北農業研究センターの水稻冷害早期警戒システム、岩手県農業研究センターの水稻生育診断予測支援システム、栃木県の水稲生育診断システム予測事業等において1kmメッシュごとに行われており、当該年の水稻の生育監視や栽培指導等の実用面で活用されている。

しかしながら、収量予測に利用する場合には、短期的な気象災害、病虫害、呼吸と温度の関係等を反映させるというモデルの改善すべき

部分があることから、関連するデータを集積し検証することによって、早期予測が可能な実用化モデルとすることが望ましいと考えられる。

イ リモートセンシング技術を活用した収量予測

現状においては、リモートセンシングデータは、瞬時性、低頻度、方向性、大気状態による補正、データ品質等について十分な吟味が必要であることから、リモートセンシング情報単独で誤差 1 % 以内という精度の高い収量予測を行うことは難しいが、植生指数^(注：20)やストレス指数^(注：21)等の情報を広域的に把握できることから、回帰式や生育プロセスに基づいたモデルと組み合わせた活用方法について研究が進められている。

リモートセンシングデータを活用した収量予測の方法としては、坪刈調査結果と組み合わせて行う面的な広がり の推定や倒伏被害面積の把握に加えて、数カ所のモデル地区での生育状況のモニタリングによるその年の傾向の把握などを行い、これらの結果を従来からの方法で求めた予測値の補正に用いることにより予測精度向上へとつながり、将来的には収量予測に活用可能と考えられる。

(4) 幅を持たせた収量予測

予測時点以降の気象変動による収量の変動幅を評価する手法についても検討を進める必要があると考えられる。2 で作成した予測式により、説明変数として収穫期までの気象の実況値を入力した場合には、精度良く実収量を再現できることが検証されたが、一方、気象予報の精度の問題もあり、気象予報に基づく収量予測の精度は良くなかった。

このため、現在相当の幅を持たせて行われている気象予報を活用して、幅を持たせた収量を予測する可能性については、今後さらなる検討が必要と考えられる。

注13：「粗玄米粒数歩合」とは、全もみから粗玄米が得られる粒数の割合をいう。この指標は、全もみのうちどの程度稔実し、玄米粒、くず米粒となったかを表すものである。

14：「玄米粒数歩合」とは、粗玄米から玄米が得られる粒数の割合をいう。

15 : 「玄米千粒重」とは、玄米1,000粒当たりの重量をいう。

16 : 「掛け合わせて収量を予測」とは、全もみ数の実測値と、重回帰予測式によりそれぞれ推定した粗玄米粒数歩合、玄米粒数歩合、玄米千粒重を用いて、収量を以下の算式により予測することをいう。

$$\begin{aligned}\text{予測収量} &= \text{全もみ数} \times \text{粗玄米粒数歩合} \times \text{玄米粒数歩合} \times \text{玄米千粒重} \\ &= \text{全もみ数} \times \frac{\text{粗玄米粒数}}{\text{全もみ数}} \times \frac{\text{玄米粒数}}{\text{粗玄米粒数}} \times \frac{\text{玄米重}}{\text{玄米粒数}}\end{aligned}$$

17 : 「千もみ当たり収量」とは、1,000粒のもみから得られる玄米の重さ(収量)をいう。千もみ当たり収量は、登熟状態を総括的に表す指標として用いられる。

18 : 「重回帰予測式で直接求め」とは、過年次の千もみ当たり収量を目的変数、過年次の全もみ数、気象データを説明変数として作成した重回帰予測式に、本年の全もみ数の実測値、本年の気象データ(調査日以降は気象平年値)を入れて、本年の千もみ当たり収量を予測することをいう。

19 : 「発育指数(DeVelopmental Index; D V I)」とは、毎日の発育速度(DVR)の積算値で、発育の進度を表す。発育速度(DVR)は、気温等の関数で表される。

20 : 「植生指数」とは、リモートセンシングによって得られる複数波長の植被反射率を用いて算出する指標をいい、赤と近赤外の2バンドを使う正規化植生指数NDVIが最も簡易で一般的である。これは主に植被のグリーンバイオマスと相関が高いので、生育量のモニタリングや収量予測に利用されることが多い。

21 : 「ストレス指数」とは、熱赤外リモートセンシングによって得られる植被表面温度を気温や大気湿度に対して相対化したもので、水ストレス等による蒸散・光合成速度の低下程度を評価する指標である。

1 平年収量の計算方法

(1) 現行の計算方法

ア 基本的枠組み

収量 y を年次により傾向的に変化する部分（平年収量部分）と気象による増減の和によって表すことを考える。年次を表す変数を t 、説明変数とする気象要因を $x_1, x_2, \dots, x_m$ 、推定に用いるデータの期間を $1, 2, \dots, n$ 年の n 年、 i 年での t の値を T_i として、年次により傾向的に変化する部分を式(1)の $f_1(t)$ により、気象による増減を式(2)の $f_2(x_1, x_2, \dots, x_m)$ により表し、 y を表す関数 $f_3(t, x_1, x_2, \dots, x_m)$ を式(3)とする。

$$f_1(t) = c_0 + c_1 t + \sum_{i=1}^n d_i |t - T_i|^3 \quad (1)$$

$$f_2(x_1, x_2, \dots, x_m) = \sum_{j=1}^m a_j x_j \quad (2)$$

$$f_3(t, x_1, x_2, \dots, x_m) = f_1(t) + f_2(x_1, x_2, \dots, x_m) \quad (3)$$

ここで式(1)の $f_1(t)$ は、スプライン関数であるとしている。

このとき、 t に i 年の年次、気象要因に平年値を代入した式(3)の値を i 年の平年収量として考えることができる。

i 年における収量 y の実績を y_i 、気象要因 x_j の実績を x_{ij} とするとき、関数 f_3 における係数 c, d, a (c は c_0 と c_1 、 d は $d_1, d_2, \dots, d_n$ 、 a は $a_1, a_2, \dots, a_m$ を意味するものとする)は、次の式の E を最小にするものとする。

$$E = \sum_{i=1}^n \{y_i - f_3(T_i, x_{i1}, x_{i2}, \dots, x_{im})\}^2 + \lambda \int_{-\infty}^{\infty} \left\{ \frac{d^2 f_1(t)}{dt^2} \right\}^2 dt \quad (4)$$

ここで λ は、あらかじめ定めた正の定数で、平滑化パラメータと呼ばれる。

この式の第1項は、各年の実収量と関数 f_3 の値の差の2乗和で、これが小さいほど関数 f_3 の当てはまりが良いということになる。

第2項は、 f_3 の年次による傾向的変化の部分の「凸凹の大きさ」を λ 倍したもので、凸凹ペナルティと呼ばれる項である。 $f_1(t)$ の凸凹の大きさを $\int_{-\infty}^{\infty} \left\{ \frac{d^2 f_1(t)}{dt^2} \right\}^2 dt$ により評価すれば、第2項を入れることにより、この凸凹が小さいほど、つまり、 $f_1(t)$ が「滑らか」であるほど、この E は小さくなる。

単に第1項を小さくするのであれば、 a をどのような値にしても、第1項を0にするように c, d を選ぶことができるが、その場合、 $f_1(t)$ の滑らかさは全く保証されない。第2項を入れることにより、関数 f_3 の当てはまりの良さと、 $f_1(t)$ の滑らかさの両立が図られることになる。

また、 $f_1(t)$ が式(1)の形式で書けることを仮定せずに式(4)の E を常識的な条件の下で最小にする問題の解を求めると、 $f_1(t)$ が式(1)の形式で書けることが導かれる。

なお、式(4)の E を最小にする係数は次のようにして求められる。

まず、右辺の第2項の積分が収束しなければならない。つまり、 $f_1(t)$ は T_1 から T_n までの区間の外では1次関数にならなければならない。したがって、

$$\sum_{i=1}^n d_i = 0$$

$$\sum_{i=1}^n d_i T_i = 0$$

である。

この2式を用いて $d_1, d_2, \dots, d_n$ のうちの2個、例えば d_1, d_2 を $d_3, d_4, \dots, d_n$ によって表して E を $c_0, c_1, d_3, d_4, \dots, d_n, a_1, a_2, \dots, a_m$ の関数とする。 $f_1(t)$ は区間 $[T_1, T_n]$ の外では1次関数なので、凸凹ペナルティの積分区間は T_1 から T_n までとしていい。この E を各係数で微分し、それをすべて0とおく。その $n+m$ 個の連立1次方程式を解くと式(4)を最小にする係数を求めることができる。

また、 l は一般化クロスバリデーション(GCV)を用いて最も妥当な l を選択する。

実際の計算は、GCVPACK(注:20)を用いている。

イ 平滑化パラメータの最適化

こうした推定では、 λ (平滑化パラメータ)の値をどこにするかが問題となる。最も妥当な λ を選択することを最適化するというが、推定式の妥当性を調べるために、次の統計量がよく利用される。

クロスバリデーション(Cross-Validation)

$$CV = \sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n \cdot (1 - [H]_{ii})^2}$$

ここで y_i は i 年の推定値、 $[H]_{ii}$ はハット行列の対角要素である。

一般化クロスバリデーション (Generalized Cross-Validation)

$$GCV = \frac{SSE}{n \cdot \left(1 - \frac{\sum_{i=1}^n [H]_{ii}}{n} \right)^2}$$

ここで SSE は残差 2 乗和で $SSE = \sum_{i=1}^n (\hat{y}_i - y_i)^2$ である。

赤池の情報量基準 (Akaike's Information Criterion)

$$AIC = n \cdot \log \left(\frac{SSE}{n} \right) + 2 \sum_{i=1}^n [H]_{ii}$$

赤池の情報量基準を修正したもの (Corrected AIC)

$$AICc = \log \left(\frac{SSE}{n} \right) + \frac{n + \sum_{i=1}^n [H]_{ii}}{n - \sum_{i=1}^n [H]_{ii} - 2}$$

これらの統計量は、いずれも平滑化パラメータ λ の関数になるが、小さいほど推定式の妥当性が高いとされている。これらの統計量を最小にするパラメータを選択することを最適化するという。

このうち、 CV は、次のように書くこともできる (これが CV の本来の定義である)。

$$CV = \frac{1}{n} \sum_{i=1}^n (\hat{y}'_i - y_i)^2$$

ここで、 $\hat{y}'_i$ は、 i 年のデータを除いて、同じ方法により推定した t による y の i 年における推定値である。

CV (クロスバリデーション) は優れた統計量ではあるものの、幾つかの欠点があることも指摘されている。そこで、一般化クロスバリデーション (Generalized Cross-Validation, GCV) が開発された。一般化クロスバリデーションの「一般化」というのは、適用範囲が広い、というような意味ではなく、クロスバリデーションの特性を向上させた、という意味である。

また、 $AICc$ は、 AIC にみられる傾向、すなわち、平滑化パラメー

タの最適化を行うと凸凹が大きすぎる推定値が得られることがあるという傾向を弱めることを目的として開発されたものである。

なお、GCVとAICは異なった文脈から導かれた基準であるが、 n が大きい場合は、GCVと $\exp\left(\frac{AIC}{n}\right) \cdot \exp(AICc)$ はほぼ等しくなる。

現行の方法では、回帰式の妥当性を調べるための統計量としてGCVを用いている。GCVは、回帰式の妥当性を調べるための統計量として、その価値が理論的にも実証的にも十分に明らかになっていて、しかも、GCVの値の平方根をほぼ予測誤差の大きさとみなすことができるからである。

ウ 現行の推定に用いた気象要素

平年収量の算定は都道府県別に行われているが、用いた気象データは現行のアメダス観測体制が整った昭和54年から平成12年までの22年間の全国の約760のアメダス地点（4要素観測844地点から北海道の水稻栽培のない地域、伊豆諸島等を除外）における気象庁によるアメダス観測データであり、年別に都道府県ごとに平均して用いた。

データの種類は、日降水量、日平均気温、日最高気温、日最低気温、日照時間、日最大風速の6種類で、各観測地点のデータは、年別にその観測地点の属する都道府県の水稻出穂最盛日の40日前から40日後までの81日間の平均値（ただし、日最大風速については出穂最盛日以降の期間中の最大値）とした。

これらをそのまま気象による増減を表す関数 $f_2(x_1, x_2, \dots, x_m)$ の説明変数として用いたのでは、相互の相関が強く、回帰式の推定において多重共線性が問題になるため、6種類のデータを主成分分析し、第1、第2、第3主成分を用いた。

主成分分析に先立ち、各データとも期間中の平均を差し引き、それらを標準偏差で除した（標準化した）。したがって、各主成分とも、期間中の平均値は0になる。

計測は、説明変数の組み合わせにより数通りを行い、その説明変数としたのは、第1、第2、第3主成分及びそのべき乗（3乗まで）から選んだ数個のものである。

(2) 算定方法の改善について

ア 推定日射量の適用

現行の計算方法である式（3）を用い、前述の6種類の気象データのう

ち、日照時間に代えて、推定日射量を適用した。

イ 補助変数の適用

$$f_2'(x_1, x_2, \dots, x_m, x_{m+1}, \dots, x_{m+m'}) = \sum_{j=1}^{m+m'} a_j x_j \quad (5)$$

式(2)の $f_2(x_1, x_2, \dots, x_m)$ に、補助変数（気象指数の中では用いていないが収量に大きな影響を与えると思われる気象要因を示す変数）を加える。すなわち、 $f_2(t, x_1, x_2, \dots, x_m)$ の代わりに式(5)の $f_2'(x_1, x_2, \dots, x_m, x_{m+1}, \dots, x_{m+m'})$ を用いる、 $\{x_{m+1}, \dots, x_{m+m'}\}$ が補助変数である。すると、収量を予測するための式が式(3)から以下のものに代わる。

$$f_3'(t, x_1, x_2, \dots, x_m, x_{m+1}, \dots, x_{m+m'}) = f_1(t) + f_2'(x_1, x_2, \dots, x_m, x_{m+1}, \dots, x_{m+m'}) \quad (6)$$

今回、補助変数として、「日最大風速の2乗値」、「低温積算値」を用いた。その場合、補助変数の原データである日最大風速・日最低気温は、主成分分析及び主成分算出から除外した。

(3) 今後引き続き検討を要する事項について

ア 気象要素を直接用いる手法等

基本的枠組み

スプライン関数 f_1 は $[T_1, T_n]$ 区間において、基底である $n + 2$ 個の B - スプライン^(注：21) $g_1(t), g_2(t), \dots, g_{n+2}(t)$ を用いて

$$f_4(t) = \sum_{i=1}^{n+2} b_i \cdot g_i(t) \quad (7)$$

と表現できる。

ただし、 $g_1(t), g_2(t), \dots, g_{n+2}(t)$ は節点を $\{T_1, T_1, T_1, T_1, T_2, \dots, T_{n-1}, T_n, T_n, T_n, T_n\}$ としたときの3次のB - スプラインの基底の全てとする。

気象による増減を表すものとして f_5 を用いる。ここで $\{x_1, x_2, \dots, x_m\}$ は気象データそのもので、気象指数は用いていない。これによって、それぞれの気象データが回帰式において果たしている役割が明確になる。

$$f_5(x_1, x_2, \dots, x_m) = \sum_{j=1}^m a_j x_j \quad (8)$$

このとき、収量 y は次のような関数になる。

$$\begin{aligned} f_6(t, x_1, x_2, \dots, x_m) &= f_4(t) + f_5(x_1, x_2, \dots, x_m) \\ &= \sum_{i=1}^{n+2} b_i \cdot g_i(t) + \sum_{j=1}^m a_j x_j \end{aligned} \quad (9)$$

$[T_1, T_n]$ の外で $f_1(t)$ は1次関数なので、式(4)における凸凹ペナルティの積分区間を T_1 から T_n までとしてよいことに注意すれば、式(4)の E を最小にする b と a は次のように求めることができる。

まず

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n+2} \end{pmatrix}, \quad a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix},$$

$$G = \begin{pmatrix} g_1(X_1) & g_2(X_1) & \cdots & g_{n+2}(X_1) \\ g_1(X_2) & g_2(X_2) & \cdots & g_{n+2}(X_2) \\ \vdots & \vdots & \ddots & \vdots \\ g_1(X_n) & g_2(X_n) & \cdots & g_{n+2}(X_n) \end{pmatrix},$$

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix},$$

$$\text{また、} k_{ij} = \int_{T_1}^{T_n} \left(\frac{d^2}{dt^2} g_i(t) \right) \left(\frac{d^2}{dt^2} g_j(t) \right) dt \quad \text{とし、}$$

$$K = \begin{pmatrix} k_{11} & k_{12} & \cdots & k_{1,n+2} \\ k_{21} & k_{22} & \cdots & k_{2,n+2} \\ \vdots & \vdots & \ddots & \vdots \\ k_{n+2,1} & k_{n+2,2} & \cdots & k_{n+2,n+2} \end{pmatrix}$$

とすると、式(4)は、

$$E = (y - Gb - Xa)^t (y - Gb - Xa) + l b^t K b \quad (10)$$

と書くことができる。

この E を最小にする b と a は、 E を b と a で偏微分したものが 0 になるとして求められる。

$$\frac{\partial E}{\partial b} = \begin{pmatrix} \frac{\partial E}{\partial b_1} \\ \vdots \\ \frac{\partial E}{\partial b_{n+2}} \end{pmatrix} = -2 G^t (y - G b - X a) + 2 I K b$$

$$\frac{\partial E}{\partial a} = \begin{pmatrix} \frac{\partial E}{\partial a_1} \\ \vdots \\ \frac{\partial E}{\partial a_n} \end{pmatrix} = -2 X^t (y - G b - X a)$$

$\frac{\partial E}{\partial b} = O_{n+2}$ $\frac{\partial E}{\partial a} = O_m$ として、得られる b, a を $\hat{b}, \hat{a}$ とすると、

$$\begin{pmatrix} \hat{b} \\ \hat{a} \end{pmatrix} = \begin{pmatrix} G^t G + I K & G^t X \\ X^t G & X^t X \end{pmatrix}^{-1} \begin{pmatrix} G^t \\ X^t \end{pmatrix} y \quad \text{となる。}$$

このとき、 (n, n) 型行列

$$H = (G, X) \begin{pmatrix} G^t G + I K & G^t X \\ X^t G & X^t X \end{pmatrix}^{-1} \begin{pmatrix} G^t \\ X^t \end{pmatrix}$$

をハット行列という。

推定に用いた期間 n 年分の収量の推定値を表した n 次ベクトルを、 y とすれば、 $\hat{y} = Hy$ となる。

$[T_1, T_n]$ の外では $f_1(t)$ は 1 次関数であるが、その具体的な形は、 T_1 と T_n において $f_1(t)$ と $\frac{d}{dt} f_1(t)$ が連続であることから、B - スプラインによって表現した $f_4(t)$ を $[T_1, T_n]$ の外に延長することによって $f_1(t)$ を求めることができる。

これによって、式 (7) と式 (1) は数学的には同じものになるが、 $f_3(t)$ を用いて数値計算を行うと、高い精度の結果を得るためには、数値計算上の技巧を必要とする。G C V P A C K では、そうした技巧を充

分に取り入れ、しかも計算量を少なくするための工夫が加えられている。

一方、 $f_6(t, x_1, x_2, \dots, x_m)$ を用いて数値計算を行うと、この式をそのまま用いても高い精度の結果が得られる。このため、式(9)の $f_6(t, x_1, x_2, \dots, x_m)$ を基本にし、リッジ回帰の手法を加味する場合は式(11)、さらに、隣接県を統合して推定する場合は、式(12)、又は、式(13)を用いて検証した。

B - スプラインを用いる計算方法とこのプログラムは、竹澤委員の開発によるものである。

リッジ回帰を適用する方法

重回帰式の推定において、説明変数間の相関が高いと説明変数が少し動くだけで回帰係数が大きく変動し、誤差分散の小さい安定した推定ができないという多重共線性の問題が生じる。

平年収量の推定においては、気象による収量の増減を説明する変数（気象要因）に気象要素（単一の気象データ、又は、気象データを作物学的知見から加工したもの）を単独で用いる場合は、説明変数間の相関関係が強くなることが多く、新しい年次のデータを追加することにより回帰係数が大きく変動するなど、気象要因の係数の安定性が問題となる。

リッジ回帰は、回帰係数の2乗和に制約を設けて誤差分散の小さい安定した推定結果を得ようとするものであり、平年収量の推定にリッジ回帰を適用する場合は、気象要因の係数 $a_1, a_2, \dots, a_m$ に制約を設け、式(10)を次のようにする。

$$E = (y - Gb - Xa)^t (y - Gb - Xa) + lb^t Kb + ma^t a \quad (11)$$

ここで μ は、正又は0の定数で、リッジ回帰の定数（又はリッジパラメータ）という。 $\mu = 0$ のとき、この式は、式(10)に一致する。 μ が大きくなるにしたがって、気象要因の係数の2乗和を小さくしようとする傾向が強くなる。式(11)のEを最小にする b, a をそれぞれ $\hat{b}, \hat{a}$ とし、式(10)における場合と同様に求めると、次のようになる。

$$\begin{pmatrix} \hat{b} \\ \hat{a} \end{pmatrix} = \begin{pmatrix} G^t G + lK & G^t X \\ X^t G & X^t X + mI_m \end{pmatrix}^{-1} \begin{pmatrix} G^t y \\ X^t y \end{pmatrix}$$

ここで、 I_m は、 m 次の単位行列である。ハット行列は次のようになる。

$$H = \begin{pmatrix} G & X \end{pmatrix} \begin{pmatrix} G^t G + lK & G^t X \\ X^t G & X^t X + mI_m \end{pmatrix}^{-1} \begin{pmatrix} G^t \\ X^t \end{pmatrix}$$

リッジ回帰を適用した場合、G C V は平滑化パラメータ λ とリッジ回帰の定数 μ の関数になる。G C V で最適化する場合には、G C V を最小にする λ 、 μ を選択する。

なお、別添資料「水稻平年収量等に関する研究会におけるシミュレーション結果」では、リッジ回帰の定数 μ は、 k と表示している。

イ 県を統合・分割して推定する手法

隣接した県を統合して推定する方法

現行の方法ではそれぞれの都道府県の回帰式を独立に求めている。しかし、隣接する県で自然条件や栽培技術等の稲作条件が似ている場合は、気象要素が収量に及ぼす影響が類似していると考えられる。このことを利用すれば、各都道府県別の昭和54年以降のデータ（22年分）をより有効に利用することが期待される。

そこで、気象要因が収量に及ぼす影響を表す重回帰式の係数が隣り合う都道府県で近い値をとることを想定した回帰を行う。隣り合う2県を推定する場合には、最小にする対象の E を次のようにする。

$$\begin{aligned} E = & (y_1 - Gb_1 - X_1a_1)^t (y_1 - Gb_1 - X_1a_1) \\ & + (y_2 - Gb_2 - X_2a_2)^t (y_2 - Gb_2 - X_2a_2) \\ & + l_1 b_1^t K b_1 + l_2 b_2^t K b_2 \\ & + m(a_1^t a_1 + a_2^t a_2) \\ & + u(a_1 - a_2)^t (a_1 - a_2) \end{aligned} \quad (12)$$

ここで、 y_1 の1は県の番号で、例えば k 番の県の i 年の単収は、 y_{ki} であり、

$$y_k = \begin{pmatrix} y_{k1} \\ y_{k2} \\ \vdots \\ y_{kn} \end{pmatrix}$$

である。

ここで u は隣接する県間の気象要因の回帰係数の相違に対するペナルティの度合を定める定数（回帰係数調整パラメータ）である。

なお、別添資料「水稻平年収量等に関する研究会におけるシミュレーション結果」では、 u は s と表示している。

l 県を同時に推定する場合には、最小にする対象の E を次のようにする。

$$E = \sum_{k=1}^l \left\{ (y_k - G b_k - X_k a_k)^t (y_k - G b_k - X_k a_k) + l_k b_k^t K b_k \right\} + m \sum_{k=1}^l a_k^t a_k + u \sum_{k_1, k_2=1}^l e_{k_1 k_2} (a_{k_1} - a_{k_2})^t (a_{k_1} - a_{k_2}) \quad (13)$$

ただし、 $e_{k_1 k_2}$ は、 $k_1 \neq k_2$ の場合、 k_1 県と k_2 県が隣接しているときは 1、そうでないときは 0 とする。 $k_1 = k_2$ の場合は、0 とするとき、1 とするときがありうる。

なお、この式を隣り合う 2 つの県の場合に適用すると、式 (6) の v を $2v$ に代えた式になる。

また、最後の項は、

$$u \sum_{k=1}^l \left(a_k - \frac{1}{l_k} \sum_{K'=1}^l e_{kk'} a_{k'} \right)^t \left(a_k - \frac{1}{l_k} \sum_{K'=1}^l e_{kk'} a_{k'} \right)$$

としても類似した結果が得られると考えられる。ただし、

$$l_k = \sum_{K'=1}^l e_{kk'} \quad \text{とする。}$$

$l = 3$ のときで 3 県とも互いに隣接しているときに E を最小にする b_k, a_k を $\hat{b}_k, \hat{a}_k$ とすると、

$$\begin{pmatrix} \hat{b}_1 \\ \hat{a}_1 \\ \hat{b}_2 \\ \hat{a}_2 \\ \hat{b}_3 \\ \hat{a}_3 \end{pmatrix} = U^{-1} \begin{pmatrix} G^t & O_{n+2,n} & O_{n+2,n} \\ X_1^t & O_{m,n} & O_{m,n} \\ O_{n+2,n} & G^t & O_{m,n} \\ O_{m,n} & X_2^t & O_{m,n} \\ O_{n+2,n} & O_{n+2,n} & G^t \\ O_{m,n} & O_{m,n} & X_3^t \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

となる。ただし、

$$U =$$

$$\begin{pmatrix} G^t G + l_1 K & G^t X_1 & O_{n+2,n+2} & O_{n+2,m} & O_{n+2,n+2} & O_{n+2,m} \\ X_1^t G & X_1^t X_1 + (m+4u)I_m & O_{m,n+2} & -2uI_m & O_{m,n+2} & -2uI_m \\ O_{n+2,n+2} & O_{n+2,m} & G^t G + l_2 K & G^t X_2 & O_{n+2,n+2} & O_{n+2,m} \\ O_{m,n+2} & -2uI_m & X_2^t G & X_2^t X_2 + (m+4u)I_m & O_{m,n+2} & -2uI_m \\ O_{n+2,n+2} & O_{n+2,m} & O_{n+2,n+2} & O_{n+2,m} & G^t G + l_3 K & G^t X_3 \\ O_{m,n+2} & -2uI_m & O_{m,n+2} & -2uI_m & X_3^t G & X_3^t X_3 + (m+4u)I_m \end{pmatrix}$$

である。

隣接した県を統合した場合は、年によるデータの並びが第 1 の県の 1 , 2 , . . . , n 年、第 2 の県の 1 , 2 , . . . , n 年という形になっているので、これからある県の 1 年分を除いても別の県の同じ年が残ることになる。このため、ハット行列を用いて C V を簡単に求めることはできない。

C V は、まず、各年について、すべての県のデータからその年のデータを除いて回帰式を推定した後、その推定式の変数 t にその年の年次を表す値を代入して各県におけるその年の単収の推定値を算出する。こうして求めた k 県の i 年の推定値を $\hat{y}'_{ki}$ として次のように算出する。

$$CV = \sum_{k=1}^l \sum_{i=1}^n (\hat{y}'_{ki} - y_{ki})^2$$

G C V、A I C、A I C c は単純に定義できないので、この C V によってパラメータを最適化する。パラメータの数が増えると最適化のために必要な計算回数が飛躍的に増大するので、例えば、平滑化パラメータについてはあらかじめ各県単独で最適化したものを用いるなどの工夫が必要になる。

また県によって作付面積が異なるので E と C V において各県を平等に扱わず、作付面積の違いを何らかの形で反映させるようにすることも検討する必要がある。

県を分割して推定する方法

県をいくつかの地域に分割し、それぞれの地域で推定した場合、それぞれの最適化は G C V によって行う。県全体の平年収量を推定する方法

として、この県を分割して推定する方法と県全体をひとまとめにして推定する方法のどちらが妥当であるかを比較するには、C Vを用いる。県を分割して推定する場合、C Vの計算に用いる県全体としての実単収は、当該年1年を抜いて推定した当該年の各地域の実単収を当該年のそれぞれの地域の水稲の作付面積により加重平均した値とする。

ウ 採用する推定式とパラメータ値について

推定式(説明変数の選択を含む)が的確でデータ量も十分であるならば、その妥当性を判定するG C Vなどの指標が最小になる点を最も妥当なパラメータの値としてよい。また、そうした指標を最小にする推定式を最も妥当なものとして選択してよい。

しかし、それらが十分に満たされない場合は、G C Vなどが最小にならないところのほうが実態をより適切に反映した係数を得ることができるということもありうる。推定式を的確に選択することは実際には容易ではないこと、平年収量の推定においてはデータ量が限られていることから、採用するパラメータの値と推定式を決定するに当たっては、G C Vなどの指標に加え、作物学的な知見や品種、栽培技術の動向などを考慮する必要がある。

注20 : 「G C V P A C K」とは、

昭和60年頃にウィスコンシン大学統計学部で開発されて以来、平滑化スプラインと部分スプラインの計算を行うための標準的なプログラムとして世界的に高く評価されているものなので、信頼性が高い。

F O R T R A N 7 7 を使って書かれていて、数値計算上の様々な工夫が凝らされているので高速に実行できる。

G C V P A C K は、固定された機能を持つサブルーチン集なので、近隣の都道府県の影響を加味したり、回帰係数に条件を加えるなどの形に発展させることはできない。

21 : 「B - スプライン」とは、

B - スプラインは、特定の狭い領域でだけ正の値をとり、その他の領域では0になる関数の和によって、スプライン関数を表現する方法の代表的なものである。この関数のそれぞれを基底(Basis)と呼ぶことが、B - スプラインという言葉の由来である。

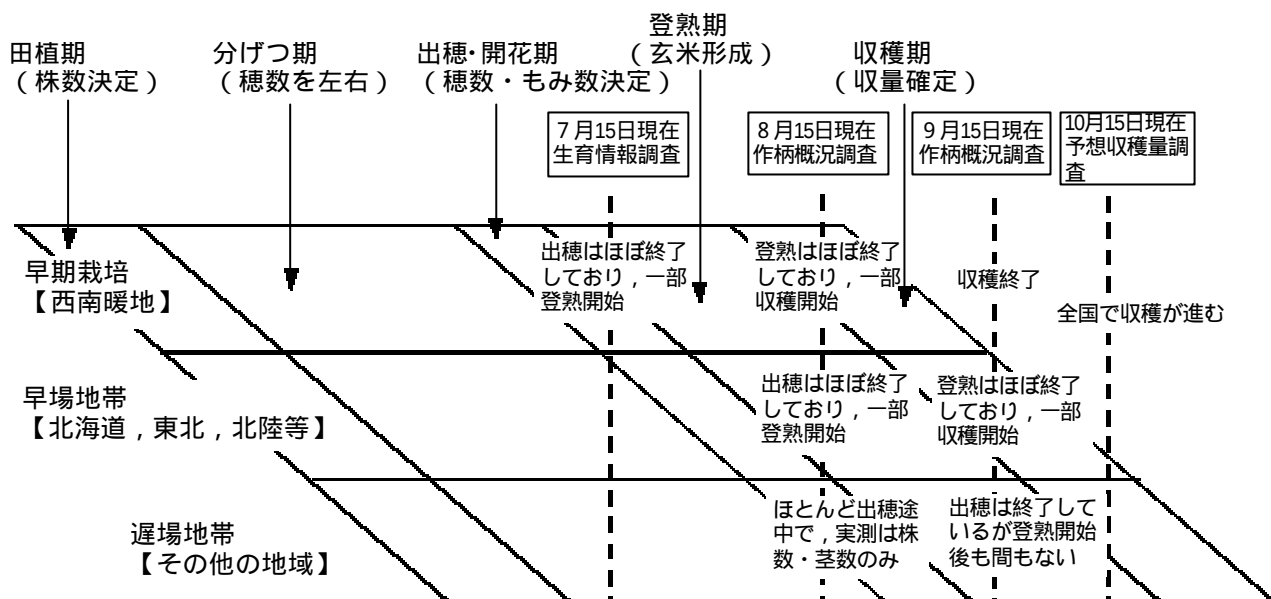
計算式をそのまま使って計算しても高い精度の推定値が得られる。

近隣の都道府県の影響を加味したり、回帰係数に条件を加えるなどの形への発展は、線形の連立方程式に帰着するものに限って可能である。

ハット行列を求めることができるので、前述する、C V、G C V、A I C、A I C cなどの統計量が得られる。従って、回帰式の選択や回帰式の性質を容易に調べることができる。

2 収量予測時期別生育ステージ及び実測・推定項目

○ 各収量予測時期と地帯別生育ステージの関係



○ 各収量予測時期別実測項目と推定項目の関係

調査時期	地帯区分	調査項目
7月15日現在 生育情報調査	早期栽培	実測項目：1 m ² 当たり全もみ数 推定項目：千もみ当たり収量(粗玄米粒数歩合、玄米粒数歩合、玄米千粒重)
	早場地帯	収量の推定は行わない
	遅場地帯	収量の推定は行わない
8月15日現在 作柄概況調査	早期栽培	実測項目：1 m ² 当たり全もみ数、一部の刈取調査 推定項目：一部の千もみ当たり収量(粗玄米粒数歩合、玄米粒数歩合、玄米千粒重)
	早場地帯	実測項目：1 m ² 当たり全もみ数 推定項目：千もみ当たり収量(粗玄米粒数歩合、玄米粒数歩合、玄米千粒重)
	遅場地帯	収量の推定は行わない
9月15日現在 作柄概況調査	早場地帯	実測項目：1 m ² 当たり全もみ数、一部の刈取調査 推定項目：一部の千もみ当たり収量(粗玄米粒数歩合、玄米粒数歩合、玄米千粒重)
	遅場地帯	実測項目：1 m ² 当たり全もみ数 推定項目：千もみ当たり収量
10月15日現在 予想収穫量調査	早場地帯	実測項目：ほぼ刈取調査収量
	遅場地帯	実測項目：1 m ² 当たり全もみ数、一部の刈取調査 推定項目：一部の千もみ当たり収量(粗玄米粒数歩合、玄米粒数歩合、玄米千粒重)

1 水稻平年収量等に関する研究会開催要領

1 目 的

水稻10 a 当たり平年収量の算定や作況調査における収量予測の精度向上を図るため、最先端技術及び研究成果の活用等について技術的見地から検討を行う。

2 位置付け

統計情報部長の私的研究会とするが、農業技術等広範な分野にわたることから、官房企画評価課（技術調整室）、生産局農産振興課、技術会議事務局企画調査課、食糧庁計画課等の協力を得ることとする。

3 研究会の構成等

- （１）「水稻平年収量等に関する研究会」（以下「研究会」という。）の委員は、別紙のとおりとする。
- （２）研究会は、必要に応じて、関係者からのヒアリングを行うことができるものとする。

4 公開方法

会議は非公開とするが、議事概要等については公開することを原則とする。ただし、ヒアリングにおける自由な発言を確保する必要がある場合等についてはこの限りではない。

5 スケジュール

平成13年 2 月13日に第 1 回研究会を開催し、1 箇月に 1 回程度開催。
平成13年夏頃を目途に報告取りまとめを行う。

「水稻平年収量等に関する研究会」委員

	所 属
秋田 重誠	滋賀県立大学環境科学部教授
井上 君夫	独立行政法人農業技術研究機構中央農業総合研究センター耕地環境部気象立地研究室長
井上 吉雄	独立行政法人農業環境技術研究所地球環境部生態システム研究グループ環境計測ユニット研究リーダー
岸野 洋久	東京大学大学院農学生命科学研究科教授
佐々木忠勝	岩手県農業研究センター農産部長
塩野 晃一	栃木県経営技術課専門技術員
高橋 弘	(株)三菱総合研究所地球環境研究本部長
竹澤 邦夫	独立行政法人農業技術研究機構中央農業総合研究センター農業情報研究部営農情報システム研究室主任研究官
鳥越 洋一	独立行政法人農業技術研究機構東北農業研究センター地域基盤研究部連携研究第1チーム長
浪岡 實	J A 全農営農総合対策部生産システム研究室技術主管
二宮 正士	独立行政法人農業技術研究機構中央農業総合研究センター農業情報研究部分散コンピューティング研究室長
畑中 孝晴	(社)農林水産先端技術産業振興センター理事長
平澤 正	東京農工大学農学部教授
堀江 武	京都大学大学院農学研究科教授
本郷 千春	千葉大学環境リモートセンシング研究センター助手
湯田 憲一	気象庁気候・海洋気象部気候情報課予報官（現福岡管区气象台技術部長）

（五十音順）

2 研究会における検討経過

研究会の開催月日		主 な 議 題	主 な 論 点
第 1 回研究会	2月13日	1 研究会の運営について 2 平年収量算定方式の説明及び精度向上のための改善の検討 3 作況調査（収量予測）方法の説明 4 その他	1 「日照時間」から変換した「日射量」の適用について検討が必要 2 収量に影響を及ぼす短期的な気象要因について分析し、算定の指標にする必要がある 3 データ数に制限があるため、全県分を同時に推定する等の手法はどうか 4 高位・中位等、幅を持たせた予測はどうか 5 収量予測にもトレンドを考慮した方が精度は良くなるのではないか 6 収量構成要素ごとに予測し、それを掛け合わせて収量予測を行うと、誤差が積み上がる恐れがあるのではないか 7 収量予測の変数に「もみ数」等の実測データを入れるべきではないか
第 2 回研究会	3月9日	1 季節予報の利用と今後の改善 2 気象要因と収量 3 気象情報の作成と活用 4 検討・検証した事項とシミュレーション結果及び今後の検討（平年収量関係） 5 平年収量算定方式の精度向上のための提案	1 都道府県ごとに回帰モデルを適用するのではなく、全国、または、広い地域でモデルを統一する検討が必要ではないか 2 気象要素のうち、最高気温、気温日格差、日射量の回帰係数が負となる場合、その回避方法の検討が必要ではないか 3 推定日射量の適用について、より多くの県でも検証が必要ではないか
第 3 回研究会	4月6日	1 水稻の生育・収量予測モデルの研究の現状 2 現行の算定方式の発展型の可能性について 3 検討・検証した事項とシミュレーション結果及び今後の検討（平年収量関係）	1 水稻の収量予測には、出穂前約2週間前の乾物重データが重要であると考えられるため、データを蓄積し、分析する必要があるのではないか 2 隣接県など、広い地域でモデルを統一する手法については、栽培方法や品種等に関する条件を考慮する必要があるのではないか
第 4 回研究会	4月27日	1 水稻冷害早期警戒システムの実践 2 水稻栽培モデルの実践 3 栃木県における水稻生育診断システム予測事業について 4 収量予測の現行手法及び改善のためのシミュレーション結果並びに当面の検討事項	1 その時点までの稲の履歴を表している実測データを活用するべきではないか（その実測データは穂数と考えて良いか。葉色指数はどうか） 2 各調査時に、もみ数等が実測できるところがあれば、できるだけ実測によるデータを活用した方が良いのではないか
第 5 回研究会	5月22日	1 収量予測へのリモートセンシング技術の活用について 2 リモートセンシング技術の現状と応用の可能性 3 現行収量予測手法の改善のためのシミュレーション結果及び当面の検討事項について 4 平年収量に関し検証した事項とシミュレーション結果及び当面の改善事項と検討事項について	1 現行モデルとリッジ回帰を加えた発展型モデルの比較において、トレンドの違いについて、さらに検討・検証が必要ではないか 2 収量予測において、将来的に、リモートセンシングデータを用いる方法を検討する必要があるのではないか
第 6 回研究会	6月15日	1 論点メモ（平年収量関係） 2 論点メモ（収量予測関係）	1 補助変数の適用において、さらに、気象要素・期間の検討が必要ではないか 2 「土壌」の記述について、土壌の影響は少なく、かなりの部分が気象要因で説明可能であるので、検討が必要ではないか 3 県を分割する方法においては、小地域ほど特有な条件を入れなければ精度の向上は図られないので、さらに検討が必要ではないか 4 予測時以降の気象変動による収量の変動幅を評価する手法について、引き続き検討が必要ではないか 5 リモートセンシングデータの活用方法について、引き続き検討が必要ではないか
第 7 回研究会	7月6日	「水稻平年収量等に関する研究会報告（案）」についての審議	

3 研究会における委員からの発表テーマ一覧

	開催日	発表者	発表テーマ
第2回	3月9日	湯田 委員	「季節予報の利用と今後の改善」
		平澤 委員	「気象要因と収量」
		井上（君）委員	「気象情報の作成と活用」
第3回	4月6日	堀江 委員	「水稻の生育・収量予測システムの研究の現状」
		竹澤 委員	「現行の算定方式の発展型の可能性について」
第4回	4月27日	鳥越 委員	「水稻冷害早期警戒システムの実践」
		佐々木 委員	「水稻栽培モデルの実践」
		塩野 委員	「栃木県における水稻生育診断システム予測事業について」
第5回	5月22日	井上（吉）委員	「収量予測へのリモートセンシング技術の活用について」
		本郷 委員	「リモートセンシングの現状と応用の可能性」