【土木計画学研究・講演集 No.14(1) 1991年11月】

空間分布適合度指標SFIの改良

A DEVELOPMENT IN STATISTICAL INTERPRETATION OF SPATIAL FIT INDICATOR (SFI)

宮 本 和 明*、三 浦 良 平**

by Kazuaki MIYAMOTO and Ryohei MIURA

There are many kinds of quantitative models such as landuse models which analyze values by spatial units. But there is no effective indicator which shows the goodness of fit between estimated and observed distributions of the value. The authors proposed an indicator to represent spatial fit between two distributions, which we call SFI, in a former report. SFI is calculated by the transport cost of errors in the estimated distribution to reproduce the observed one. The aim of the present study is to develop statistical interpretation of the indicator. At first, we propose another three indicators which are included in a family of SFI. The first indicator can be used only in the case that all elements for analysis can be identified each other. Another three indicators are representative transport costs in the case that all elements are indifferential between each other. They are minimum, maximum and average case values, and the first one is the SFI of the former report. In this paper, interpretation of the four indicators as well as the way of statistical testing using them are discussed.

1．はじめに

土地利用モデルなどの空間を対象とした計量モデルにおいては、一般に空間を有限な小区画（以後ゾーンと呼ぶ）に分割し、ゾーンごとの人口や土地利用面積といった値の分析を目的としている。そのため、各モデルによる分析結果の評価、いいかえると、モデルによる予測値分布と実際値分布との適合の度合いを評価する際には、各ゾーンの値のみでなく、ゾーンの空間的な位置関係をも考慮する必要がある。しかるに、従来から用いられている相関係数に代表される評価指標は、各ゾーンを独立な個体として扱い、その空間的な位置関係は完全に無視している。これは空間的な計量モデルに関してだけの問題ではなく、一般に、２つの分布間の適合度を評価する適切な手法が存在していないことに起因する。

* 正会員　工博　横浜国立大学助教授　工学部建設学科（〒240　横浜市保土ヶ谷区常盤台156）

** 学生会員　横浜国立大学大学院　博士課程前期

著者達は、この問題意識にもとづいて、２つの分布間の空間的な適合度を示す指標として、空間分布適合度指標（ＳＦＩ：Spatial Fit Indicator）を提案してきている[1)2)]。本稿においては、先ず、既に発表してきた指標をより一般的に見直すことにより、場合に応じて使い分けるべき新しい指標を追加している。そしてさらに、従来の考え方においては、ＳＦＩは適合度の相対比較においてのみ用いられるに過ぎなかったが、本稿においては、その絶対的な評価として、統計的な検定指標として位置づける試みも行っている。なお、本研究においては、配分モデルの評価を行うことを対象としていることから、観測値分布と予測値分布の総計値は等しいことを前提にしている。なお、異なる場合においても各ゾーンの値を分布比率で置き換えることにより、本研究の考察がそのまま適用できるものである。

2．既存の適合度指標の問題点

相関係数に代表される既存の適合度指標は、(1)ゾーンの個々の位置、および(2)個々のゾーン間の位置

関係に関する情報はまったく考慮されない。そのため、実際値が大きなゾーンの予測値が大きければ、一般に高い相関係数となる。しかるに、その高い予測値が、例えば、隣のゾーンとなった場合は、相関係数はかなり低くなることが考えられる。このように相関係数は、当然のことながら、空間分布の適合度のごくわずかな側面を示すに過ぎない。空間分布の適合度の評価に着目した研究としては、清水ら[3]の研究があげられる。これは相関係数に「融通性」の概念を導入することにより位置情報を加味した評価を可能にするという新しい考え方を作り上げたものである。しかし、単一の指標値を与えるものではないことから、必ずしもそれを用いた評価が容易なものとはなっていない。著者達の研究においては、評価の視点が決まれば、２つの空間分布間の適合度を単一の指標値として表す方法を提案することにより、評価をより容易にし、さらに、その絶対的な評価につなげることをめざしている。

３．空間分布の適合度の考え方

ここでは、既に報告している空間分布適合度指標ＳＦＩを含めて、あらためて、空間分布間の適合度について考える。

本研究では、適合度を比較する二つの分布を便宜的に、観測値分布と予測値分布と呼ぶことにする。観測値分布は唯一の分布が与えられており、予測値分布はモデルの違い等により複数個与えられることが有り得る。なお、適合度自体は２つの分布間に定義されるものであり、どちらが基準値あるいは真値であるかは、適合度の計算に際しては、実際上は問題とはならない。また、便宜上、観測値分布の各ゾーン値の総計と予測値分布の各ゾーン値の総計は等しいものとする。いいかえると、予測値分布は何らかの配分モデルによる配分結果と考えることができる。さらに、各ゾーン値は人口あるいは世帯数といった非負の整数であるとする。なお、これらの条件は説明のためのものであり、本質的な前提ではない。

まず、個々の配分対象である個人あるいは世帯といった要素について考える。各要素は、本質的には個々に特定化しうるものであり、ある要素に関して考えると、予測されたゾーンの位置が観測されたゾーンの位置に近い方が予測精度が高いと判断できる。いいかえると、各要素の予測ゾーンと観測ゾーンの距離がモデルの予測性能の評価の基本要素と考えることができる。これを分布全体で考えると、各要素の予測ゾーンと観測ゾーンの距離の総計が、分布間の適合度の指標となると考えられる。ここで、予測ゾーンと観測ゾーンの距離は、予測結果から観測結果を再現するとした場合の要素の輸送費用とみなすこともできため、本研究ではこの値を「輸送費用」と呼ぶことにする。個々の要素が特定化できる場合は、この輸送費用が各要素ごとに唯一求められることから、その総計である総輸送費用も唯一値に確定する。本研究ではこれを「確定総輸送費用」と呼ぶことにする。

しかし、実際の計量モデルを用いた分析においては、個々の要素を特定化できる場合は極めて希であり、一般には、各要素は無差別として扱われる。その場合は、予測値分布から観測値分布を再現する輸送ケースは特定化する事はできなく、両分布の組み合わせによって決まる場合の数だけ可能性として存在することになる。そして、この場合の数に相当する総輸送費用が求められる。その中でも代表的な総輸送費用として、最小値、最大値、そして平均値が考えられる。最小値は、予測値分布を最も好意的にみた場合の総輸送費用であり、逆に、最大値は、最悪のケースの総輸送費用である。また、平均値は、確率的にみた期待値としての総輸送費用と考えることができる。本研究ではそれぞれを、「最小総輸送費用」「最大総輸送費用」「平均総輸送費用」と呼ぶことにする。なお、各総輸送費用は、予測値分布が与えられると各々唯一の値となるものである。

この内の最小総輸送費用に基づく指標が参考文献１）２）に報告したＳＦＩである。本稿においては、以上の他の３つの指標を含めて改めて、後に述べるように、空間分布適合度指標ＳＦＩを定義している。

４．ＳＦＩの計算方法

以上の考え方のもと、ＳＦＩを先に定義した総輸送費用として、改めて以下のように４つの指標として定義し直すこととする。

（１）確定ＳＦＩ（ＳＦＩＤ）

確定ＳＦＩは各要素が識別できる場合に、各要素

の観測ゾーンと予測ゾーンとの距離の総和で与えられる。これは、本来の意味での予測結果から観測結果を再現するための総輸送費用である。要素eの予測ゾーンをｉ（ｅ）、観測ゾーンをｊ（ｅ）、その距離を$d_{i(e)j(e)}$とおくと、ＳＦＩＤは以下のように求められる。

$$\mathrm{SFID}=\Sigma_e\ d_{i(e)j(e)} \tag{1}$$

（２）最小ＳＦＩ（ＳＦＩＬ）

最小ＳＦＩは先に述べた最小総輸送費用である。この求め方は、各ゾーンの予測値と観測値の差すなわち「誤差」を最も効率よく移動させて観測値分布を再現する場合の総輸送費用である。

すなわち、ゾーン数をｎとし、過大予測ゾーンｉの過大量をa_iとする。そして、過小予測ゾーンｊの過小量をb_jとする。ここで、過大予測ゾーン数と過小予測ゾーン数をそれぞれｌ，ｍとする。さらに、過大予測ゾーンを供給地、過小予測ゾーンを需要地とみなすことにより、供給地ｉから需要地ｊへの輸送量をx_{ij}とすれば、最適輸送量を求める問題は、次のような線形計画の輸送問題として定式化できる。

（供給制約）$\Sigma_j\ x_{ij} \leqq a_i$ (2)

（需要制約）$\Sigma_i\ x_{ij} \geqq b_j$ (3)

（非負条件）$x_{ij} \geqq 0\ \ (i=1,\cdot\cdot,l;j=1,\cdot\cdot,m)$ (4)

そして、ゾーンｉからゾーンｊへの輸送費用をゾーン間の距離の関数d_{ij}で表すと、総輸送費用を表す目的関数Ｓは、

$$S=\ \Sigma_{ij}\ d_{ij}x_{ij} \tag{5}$$

となり、この総輸送費用Ｓの最小値を線形計画法の輸送問題として求め、その値をＳＦＩＬとする。

（３）最大ＳＦＩ（ＳＦＩＭ）

最大ＳＦＩは先に述べた最大総輸送費用である。この求め方は、各ゾーンの予測値を最も効率悪く移動させて観測値分布を再現する場合の総輸送費用である。

すなわち、ゾーン数をｎとし、ｉゾーンの予測値と観測値をそれぞれA_i、B_iとする。最大輸送量を求める問題は、次のような線形計画の輸送問題として定式化できる。

（供給制約）$\Sigma_j\ x_{ij} \leqq A_i$ (6)

（需要制約）$\Sigma_i\ x_{ij} \leqq B_j$ (7)

（非負条件）$x_{ij} \geqq 0\ \ (i=1,\cdot\cdot,n;j=1,\cdot\cdot,n)$ (8)

そして、この時、総輸送費用を表す目的関数Ｔは、

$$T=\ \Sigma_{ij}\ d_{ij}x_{ij} \tag{9}$$

となり、この総輸送費用Ｔの最大値を線形計画法の輸送問題として求め、その値をＳＦＩＭとする。

（４）平均ＳＦＩ（ＳＦＩＥ）

平均ＳＦＩは、先に述べた平均総輸送費用である。このＳＦＩＥは各要素が識別できない場合に、すべての場合の数における総輸送費用の平均として、いいかえるとすべての場合を等確率として考えた確率的な期待値として与えられる。要素の総数をＰとおき、先の記号を用いると、すべての場合の数は、$[(P!)^2/((A_1!\cdot A_2!\cdots A_n!)(B_1!\cdot B_2!\cdots B_n!))]$と膨大なものとなり、実際上すべての場合を計算することは困難である。そこで、期待値計算として以下のように求める。予測値の要素１単位あたりのｉゾーンから全ゾーンへの輸送費用の平均$z_{i\cdot}$は、観測値分布の重み付き平均として、

$$z_{i\cdot}=\Sigma_j\ \frac{B_j}{P}\ \ d_{ij} \tag{10}$$

として与えられる。これに予測値分布の各ゾーンの値を乗じた総計が総輸送費用の期待値でありＳＦＩＥとなる。

$$\begin{aligned}\mathrm{SFIE}\ &=\ \Sigma_i\ A_i\cdot z_{i\cdot}\\ &=\ \Sigma_{ij}\ \frac{A_iB_j}{P}\ \ d_{ij}\end{aligned} \tag{11}$$

５．ＳＦＩの分布

（１）ＳＦＩを用いた評価

以上の４つのＳＦＩは、予測値分布の評価の視点にあった指標を選択し、それぞれ単独に計算することができる。そのため、予測値分布が複数あり、それらを比較評価して最良のものを見つけだすためには、各視点にあったＳＦＩを求めその値を評価し、最小のものが最も適合度が高い分布とみることができる。しかるに、一般には単一の予測結果の絶対評

価を行いたい場合も多く、SFIを統計的な有意性検定指標としてとらえることが必要である。そのためには、各SFIの母分布を知ることが必要となる。その際、各SFIの母分布形は解析的には導出不可能であり、また、要素数は先にも例をあげたように膨大となるため、直接分布を求めることができない。そのため、以下のように近似分布を求めている。

（2）SFIDの分布

SFIDは全ての予測値分布パターンに対する観測値分布からの総輸送費用の分布に相当し、中心極限定理を適用すると、その分布は正規分布に近似することができる。

すなわち、U_i（i＝1，2，･･･n）が互いに独立で、平均値μ_i、分散σ_iを有するとする。この時、

$$W_n = \Sigma_i U_i \quad (12)$$
$$M_n = \Sigma_i \mu_i \quad (13)$$
$$V_n = \Sigma_i \sigma_i \quad (14)$$

とおくと、nが十分大きな場合、中心極限定理より、W_nの平均と分散はそれぞれEXP（W_n）＝M_n、VAR（W_n）＝V_nとなる[4)5)]。

ここで、先の記号を用いると、観測値分布が与えられた場合の、SFIDは以下の平均μ_Dと分散σ_Dをもつ正規分布に近似される。ゾーンiからみた全てのゾーンへの要素1単位の輸送費用の平均をf_i、分散をg_iとおくと、

$$\mu_D = \Sigma_i B_i f_i \quad (15)$$
$$\sigma_D = \Sigma_i B_i g_i \quad (16)$$

となる。ただし、SFIDには最小値と最大値が存在することから両端部分においては正規分布とみなすことはできないが、本研究の目的からすると、実用上は十分近似できるものと考えられる。

SFIDの分布が与えられれば、予測値分布のSFIDの値を計算することにより、その予測結果の有意性が図1に示すように検定できることになる。

（3）SFILの分布

まず、図2にSFIDの分布と、ある予測値分布が与えられた場合の総輸送費用の分布、そしてその場合のSFIL、SFIM、そして、SFIEを示す。　SFILは基本的には、この図が示すようにSFIDを母分布としその中から部分要素を抽出した場合の最小値であることから、一種の極値分布であることが考えられる。ただしこの場合は、一般の極値分布の仮定である、母集団からランダムに同数要素を抽出する条件を満たさない。すなわち、予測値分布に依存して抽出される要素が特定化され、また、その要素数も予測値分布により異なるものである。しかし、この場合、要素数が十分大きくなると、SFIL分布に最小値が存在することから、極値分布のⅢ型すなわちワイブル分布の最小値分布に収束することが考えられる。なお、SFILの最小値は予測値分布が観測値分布に一致した場合であり0となる。そして、その収束速度等は不明であるが、本研究の対象とする土地利用分析等においてはその要素数が十分大きなことから、実用上はワイブル分布として近似されると考えられる。ただしそのパラメーターの値に付いては、解析的に求められないため、乱数を用いた数値計算が必要となる。またその際、分布形を確認する必要がある。

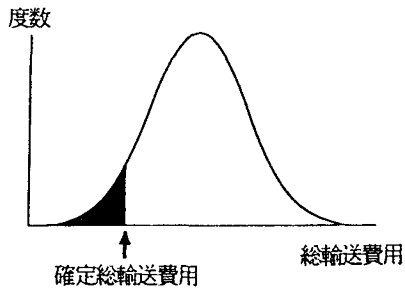

図1　SFIDの分布とそれに基づく検定

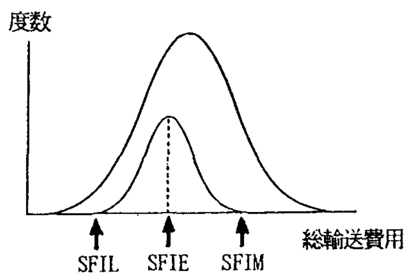

図2　SFIDの分布と予測値分布が与えられた場合の総輸送費用の分布

（4）SFIMの分布

SFIMの分布もSFILの分布同様、SFIDに最大値が存在することから、ワイブル分布の最大

値分布に近似されると考えられる。この場合も同様に、乱数を用いた数値計算によりパラメーターの値を求め同時に分布形の確認を行う必要がある。

（５）ＳＦＩＥの分布

ＳＦＩＥも前２者と同様の抽出要素の平均値の分布であり、抽出要素数は予測値分布により異なっている。しかしこの場合も、要素数が十分大きくなると、中心極限定理から、正規分布に近似できると考えられる。ただしこの場合は、ＳＦＩＤの場合と異なり、その平均と分散を観測値分布と予測値分布から直接求めることはできない。そのため、この場合も乱数を用いた数値計算により、そのパラメーターを求め、同時に、その分布形を確認する必要がある。

６．数値計算例

（１）数値計算の前提

以上の考察を確認するために、簡単な観測値分布とその予測値分布を作成し、それらを用いて、ＳＦＩＤ、ＳＦＩＬ、そして、ＳＦＩＥについて数値計算を行っている。それらはほとんど同様に以上の考察を裏付けるものであることから、本稿においてはそのうちの一つを例に示すこととする。

数値例で設定した観測値分布と予測値分布を図３に示す。

8	6	4
6	6	4
4	4	4

（観測値分布）

10	6	4
6	6	3
4	3	4

（予測値分布）

図３　数値計算例に用いた観測値分布と予測値分布

（２）ＳＦＩＤの分布

図３の観測値分布をもとに、予測値分布のありうるパターンでの総輸送費用を、乱数を用いてシミュレーションした結果に、先に示した中心極限定理から求めた正規分布を重ねたものが図４である。数値シミュレーションにおいては３０，０００の予測値パターンを発生させている。この図でみると、数値シミュレーション結果が中心極限定理から求めた正規分布に十分に集束していることがわかる。本例で用いた観測値分布は要素数およびゾーン数ともに少ないにもかかわらず、このように集束していることから、実際の適用においては、中心極限定理の適用が問題ないと考えられる。

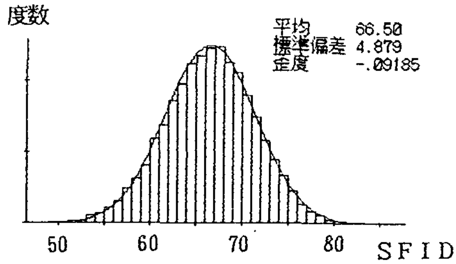

図４　ＳＦＩＤの分布（数値計算例）

（３）ＳＦＩＬの分布

ＳＦＩＬに関しても、図３の観測値分布のもとに有り得る予測値分布に対してＳＦＩＬを数値シミュレーションにより３５，０００ケース求めた結果の分布を図５に示す。この分布形はワイブル分布に近似できるとみなしてパラメータ推定を積率法とプロッティング法を用いて行った。その結果から得られた分布形をシミュレーションから求めた分布形に重ねている。図５に示すように両者はほぼ一致している。また、回帰分析の結果からも、ＳＦＩＬの分布はワイブル分布に近似できると判断できる。このように要素数そしてゾーン数ともに小数であるにもかかわらず、十分な近似ができることから、本研究の目的の範囲内では、実用上は、ＳＦＩＬの分布はワイブル分布に近似できると考えられる。

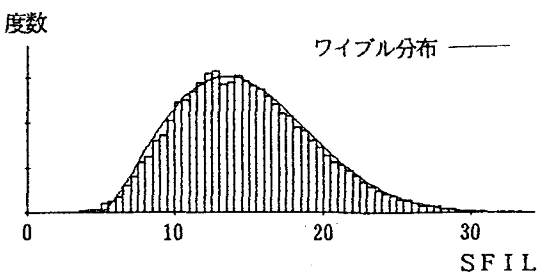

図５　ＳＦＩＬ分布とワイブル分布（数値計算例）

（４）ＳＦＩＥの分布

次に、まず、先の確率的な期待値計算による平均総輸送費用の求め方を確認するために、図３の予測値分布のもとで有り得る総輸送費用を数値シミュレーションに１０，０００ケース求めた結果の分布を図

６に示す。この場合の、確率的な期待値計算による平均総輸送費用は６４．８２であり、数値シミュレーションの結果とほぼ一致している。他の予測値分布を設定した場合においてもシミュレーションを実施しているが、同様の結果が得られている。

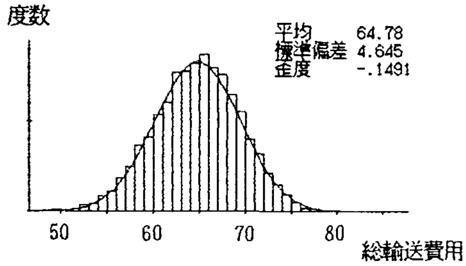

図６　予測値分布が与えられた場合の総輸送費用の分布（数値シミュレーションによる）

次に、有意性検定を行う際の基準となるＳＦＩＥの分布を求める。これも数値シミュレーションにより、予測値分布を１０，０００ケース発生させ、各ケースにおけるＳＦＩＥを期待値計算により求めている。その結果を図7に示す。

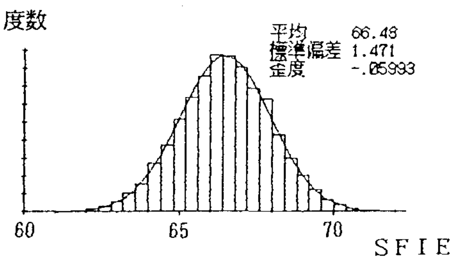

図7　ＳＦＩＥの分布（期待値計算による）

さらに確認のため、先の１０，０００の予測値分布のケースにおいて各３００の輸送パターンを発生させて求めた総輸送費用の平均値の分布形を求めている。その結果は、平均値６６．４６、標準偏差１．４９３であり、実用上、図７の分布と差がないと見なせる分布を得ている。

また、期待値計算により求める際のシミュレーションのケース数による平均値と標準偏差のそれぞれの変動についても数値シミュレーションにより求めている。その結果、この観測値分布の場合、平均値と標準偏差のそれぞれの標準偏差が、５０ケースで、０．２１３、０．１５４、１００ケースで、０．１３９、０．１０３、２００ケースで、０．０９７、０．０７７、さらに５００ケースでは０．０６５、０．０４８である。以上から、２００ケース程度の数値シミュレーションにより、本研究の目的にあった程度のＳＦＩＥの分布形を求めることができると判断される。

７．おわりに

本研究においては４つの空間分布適合度指標を提案し、各々の意味付けを行った。複数の予測分布間の相対評価においては、各指標はそれぞれ計算が容易であり、実用的に用いることができる。しかし、絶対的な評価を統計的検定を用いて行う場合は、ＳＦＩＤを除き、その母分布形の導出が数値シミュレーションを必要とする事から、容易なものとは言えない。その中では、ＳＦＩＥが比較的容易に母分布形が導出できること、また、平均値として最も代表的な指標と考えられることから、一般の評価においてはＳＦＩＥが最も有用なのではないかと現段階では考えている。

今後、さらに各指標の母分布特性をさらに詳細に検討し、その導出を容易なものとすることが必要である。また、各指標の利用目的についても、具体的な事例を用いて開拓していく必要性がある。たとえば、交通のＯＤ分布の再現性についても、有効な既存の評価指標は存在していないが、ＯＤ分布が２次元の空間分布であるという特性を考慮して、本研究で提案したＳＦＩをその評価に適用する事も可能である。また、物理的な問題として、海岸変形のシミュレーションの評価等にも用いられる可能性もあろう。ただし、ＳＦＩの実際問題での適用可能性を高めていくためには、各指標の各々の用途での算出プログラムを整備することが不可欠と考えている。

参考文献

1) 宮本和明、橋詰勝彦、後藤俊男：空間分布適合度指標ＳＦＩを用いた土地利用モデルの性能評価方法、土木計画学研究・講演集、No13、1990年
2) 三浦良平、宮本和明：空間分布適合度指標ＳＦＩの検討、土木学会第45回年次講演会概要集、1990
3) 清水英範、森山誠二、中村英夫：推定値の位置的ずれを考慮した土地利用モデルの適合度評価方法、土木学会第41回年次講演会概要集、1986
4) 森田優三：統計数理入門、pp. 183-186、日本評論社、1975
5) 竹内啓他：統計学辞典、pp. 21-26、東洋経済新報社、1989