

生態環境計測学 2017.11.22 の質問・補足

回答者 : 植山

1. 最適化と逆解析に関して説明がほしい。

[Answer]

入力データ（原因）をモデル（過程）に入力して出力（結果）を得る使用法が通常のモデルの使用法だとすると、逆解析とは、通常はモデル（過程）の出力（結果）となるデータをモデルに入力して入力データ（原因）に該当するデータをえる解析のことをさす。例えば、ある地点の大気 CO₂ 濃度は、発生源からの CO₂ と吸収源に取り込まれた CO₂、そして拡散・輸送された CO₂ の量によって決まる。通常、吸収・放出源からの CO₂ 収支（原因）と大気輸送モデル（過程）による CO₂ の移動を計算すると、各場所での CO₂ 濃度（結果）を推定できる。逆解析を用いると、大気 CO₂ 濃度（結果）と大気輸送モデル（過程）から、吸収・放出源の空間分布を得ることができる。つまり、結果から原因を類推するようなモデルの利用法を逆解析という。解析によって得られた結果は、検証が不可欠である。例えば、上記の CO₂ 吸収・発生源の地理的分布であれば、観測で得られた CO₂ フラックスとレンジや傾向が一致しているかや、衛星観測により推定された CO₂ フラックスと一貫した傾向が得られているかなど、多面的なデータとの検証が行われている。

最適化とは、入力データ（原因）と出力（結果）に該当するデータをモデルに入力することで、モデル内のパラメータ（過程）を決定する解析のことをさす。与えられた入力データから出力を最も説明することのできるパラメータを類推する技法である。最適化であれば、得られたパラメータが観測により得られるパラメータと同程度であるかなどを検証する必要がある。

2. モデルの検証における決定係数や二乗平均平方根誤差の解釈について説明がほしい。

[Answer]

モデルには様々な不確実性が伴うため、モデルによる結果を実際に観測されたデータと比較して、精度の検証を行う必要がある。決定係数 (R^2) は、モデルによる結果が実際の現象の変動を何%説明できているかをあらわす指標である。 R^2 が高ければモデルは実際の変動を多く説明できており、低ければモデル結果と実際の変動には関連が薄い、或いは関係がないことをあらわす。二乗平均平方根誤差 (RMSE; Root Mean Square Error) は、両者の平均的なずれの絶対値を表す指標である。例えば、モデル結果と観測結果の RMSE が 10 W m^{-2} であったなら、モデルによる結果は $\pm 10 \text{ W m}^{-2}$ 程度誤差を含んでいることを示している。どのくらいの R^2 や RMSE が要求されるかはケースバイケースであり、高い精度が要求される場合は高い R^2 かつ低い RMSE が要求されるが、概算を知りたいだけであれば、 R^2 や RMSE はそれほど良い値でなくてもよい場合もある。

3. モデルを使用する際の外挿性能の悪さとは。

[Answer]

経験式により関係をモデル化する場合、経験式の構築に利用した条件においては、高い精度を得ることができるが、経験式の構築に利用していない条件においては著しく精度が悪くなることがある（図 1）。このように、経験式の構築に用いたレンジを越えて外挿して適用する際には、大きな不確実性が伴うことに注意が必要である。図 1 では、説明変数 x の範囲が外挿になった例を示しているが、説明変数 x が内挿となる範囲であってもそのほかの条件が経験式を作った条件と著しく異なっている場合も経験式はうまくいかない場合がある。例えば、ある森林で気温が $20\sim 25^{\circ}\text{C}$ の環境で作った光-光合成曲線は、同じ森林の $0\sim 5^{\circ}\text{C}$ の環境や、別の森林の同じ温度環境下で成り立つとは限らない。

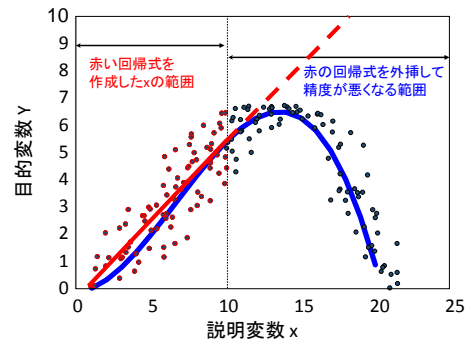


図 1 赤い点を元に作成された直線回帰式と説明変数 x がとりうる全ての範囲で目的変数 Y がとる関係（青点、青線）。限られた範囲で作成された経験式は、その範囲においては現象を近似できているが、経験式の構築に用いていない条件では、関係を正しく記述できおらず回帰式の外挿性能が悪いことが分かる（赤い破線）。