

## 当たるシミュレーションを作る：データ同化の考え方

上野玄太

統計数理研究所

データ同化とは、要するに観測データにモデルをあてはめることである。モデルとは例えば、最も単純には直線を想定しておけばよい。2次元の散布図のデータに対して、最小二乗法により、最適な傾きと切片を求める操作は学部生の演習でもお馴染みの方法だろう。直線のかわりにシミュレーションをデータにあてはめる手法のことをデータ同化という。

しかしそもそも、シミュレーションに関しては従来から脈々と受け継がれている基本理念がある。すなわち、シミュレーションとは、確立された物理学の基本法則にのっとり物理の素過程を明らかにするための道具であり、観測データとは独立に進めるべきものである、というものだ。さらに続けて、データとつぎ合わせるにしても、シミュレーションの計算がすべて完了してから行うべきものとも考えられている。この独立性を保っておかないと、どこが素過程として本質的な部分かがわからなくなってしまうというのがこの理念の根拠である。ところが、データ同化とは、そういった理念とは反する姿勢の手法である。

理念に反してまでデータ同化を行うのにはそれなりの理由がある。その理由とは煎じ詰めると、シミュレーション結果は観測データを正確に再現していない点である。その原因は当然シミュレーションの不備にある。シミュレーションを走らせる際の初期条件・境界条件、モデリングの際に無視した異スケールの物理、1,2次元性などの仮定、グリッドの大きさ、経験的公式、などがシミュレーションの不備の内訳である。

ところが、正確に再現しないと価値がないシミュレーションというものが存在する。天気予報が好例だ。そこで、従来のシミュレーションの理念には反するが、データを参考にしながらシミュレーションを修正し、現象の正確な再現を図るのがデータ同化のねらいである。理念上ではやってはいけないことではあるが、データのカンニングのプロセスをシミュレーションの実行に抱き合わせることで、現象の正確な再現という意味で価値がある「エセ」シミュレーションの結果を作り出すわけだ。

そのようなカンニングをどうやって行うのか。まず、自分はシミュレーションコードを持っているとする。いくらシミュレーションが正確でないとはいえ、信頼できる方程式なども含まれているわけだし、データにもいろいろ誤差は入り得るだろうから、データを丸のみするのもどうかと思う。そこで、データを参考にして、シミュレーションのどの部分を修正すべきかの計画を立てるのが第一歩である。次に、その計画を実行するための手法、すなわちどうやってカンニングをするのかについての吟味にうつる。そして最後に、どの程度までカンニングを認めるのが妥当か、という問題も生じてくる。

本講演では、このカンニング手法の基礎として、データ同化におけるモデル、アルゴリズムについて、ベイズ統計を基礎として述べることにする。