

気候モデル再現性評価における変数冗長性を低減する幾つかの試み

*西井和晃¹・宮坂貴文¹・小坂優²・中村尚¹・横井覚³・高藪縁³・アジアメトリックWG
(1:東大院理, 2:ハワイ大 IPRC, 3:東大 CCSR)

はじめに

Reichler and Kim (2008, BAMS) や Gleckler et al (2008, JGR) などでは気候モデル再現性の総合評価を試みている。これらの研究では、気候モデルで再現された場から 20 個前後の変数を選び、それらの観測場からの根平均自乗誤差 (RMSE) を適当に組み合わせることで 1 つの数値による総合指標を定義している。しかし、この総合指標の問題点の 1 つは変数の冗長性である。例えば上層の西風と南風、及び地表面気温の 3 つの RMSE の場合、西風と南風の関連性は強いが、地表面気温との関連性は弱いため、この 3 変数の情報は実質約 2 変数分しかない。このため、このような任意抽出された変数の単純な RMSE の和では冗長性が避けられない。この研究では、冗長性を低減するための試みを幾つか紹介し、それらの有効性を比較する。

用いたデータ

24 の CMIP3 気候モデルの全球での 1 月気候平均場の 22 の変数を用いた。比較する観測として、ERBE(外向き赤外放射、短波放射)、CMAP(降水)、HadISST(海面水温)、ISCCP(雲量)、JRA25(風速、温度、地表気温、比湿)を用いた。

RMSE の自乗和

良い再現性を持つモデルは RMSE 変数空間において原点からの距離が小さいことが期待される (図 1)。冗長性を考えなければ単純に各変数の RMSE の自乗和を計算すると各モデルの総合指標が得られる。

SVD と Non-negative Matrix factorization (NMF)

RMSE を要素に持つ変数・モデル行列の次元縮約を考える。縮約後の新たな基底上で再現性評価を行うことにより、冗長性の低減が期待される。Gleckler らは特異値分解 (SVD) を提案している (但し実施せず)。が、新たな基底は負の要素もつものがあるため (例えば図 1 左の (2) ベクトル)、ある基底上での“負荷量”の大小が必ずしもモデルの良し悪しの評価に結びつかない (図 1 左)。

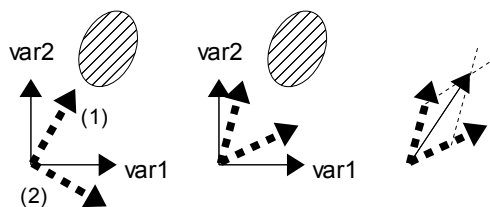


図 1 変数 var1, var2 の RMSE 位相空間。楕円はモデル分布を示す。(左) SVD による新しい基底(点線)。(中) NMF による新しい基底(点線)。(右) 新しい基底上でのモデルの RMSE ベクトル(実線)

主成分分析も次元縮約の一手法だが同様な問題がある。もし第 1 主成分の要素が全て正で、かつ、こ

の主成分が分散の大部分を説明できるならば、その因子負荷量に基づく総合指標を定義できる。しかし、これらの条件は一般に成り立つ保証はない。例えば本研究では、全分散の 80% 以上を説明するのに 6 個の主成分を要する (これはこのデータの情報量がおよそ 6 変数程度であることを示唆している)。

NMF (Lee and Seung 1999, Nature) は、非負行列を 2 つの非負行列に分解する手法である (図 1 中)。このため、新たな各基底上で負荷量が小さいモデルほど再現性の良いモデルと言える。本研究では 6 個の新たな基底上に各変数の RMSE を射影し、その値をモデル間分散で基準化した後、6 次元変数空間での各モデルの原点との距離を総合指標として用いた。

クラスタ解析

変数をクラスタ解析を用いて分別した。同じクラスタ内の変数の RMSE を自乗平均した後のクラスタ間での自乗平均を総合指標とした。横井ら (2009, 秋季気象学会予稿) も参照されたい。本研究では k-means 法を用い変数を 6 個のクラスタに分解した。

Total energy

Total energy (Talagrand 1981, Tellus) は東西風、南北風、温度、地表気圧、比湿夫々の自乗量から定義される (ただしデータの制約から本研究では地表気圧の寄与を計算していない)。よってこれをノルムとすれば上記変数の自乗誤差の和を物理的に意味のある重みで評価できる。しかしこれら以外の変数を考慮できない欠点を持つ。

総合指標間の比較

Gleckler らの Model Climate Performance Index (MCPI)、RMSE の自乗和 (S)、NMF (NMF) 及び、クラスタ解析 (CLSTR) による総合指標、Total Energy (TE) の間の相関係数を以下に示す。

	MCPI	S	NMF	CLSTR	TE
MCPI	1				
S	0.99	1			
NMF	0.97	0.99	1		
CLSTR	0.93	0.95	0.93	1	
TE	0.87	0.87	0.86	0.83	1

どの総合指標の間の相関係数も 0.8 以上で、どの指標も似た傾向を示すことが分かる。但し特定の変数しか採り入れられない TE は他の方法との相関がやや低い。しかし、どの総合指標が最も適切かについては今後の課題である。

謝辞

本研究は、環境省の地球環境研究総合推進費 (S-5) の支援により実施された。RMSE はデータ統合・解析システム (DIAS) 上で計算された。