データ同化と陸域生物圏モデル

加藤知道^{*, **, ***, †, ††}

(^{*}英国ブリストル大学地球科学科 QUEST 研究室
^{**}独立行政法人 海洋研究開発機構 地球環境変動領域
^{***}フランス気候環境科学研究所)

Data assimilation for terrestrial biosphere model

Tomomichi KATO^{*, **, ***, †, ††}

^{*}QUEST, Department of Earth Sciences, University of Bristol, Wills Memorial Building,
Queen's Road, Bristol, BS8 1RJ, UK

^{**}Environmental Biogeochemical Cycle Research Program, Research Institute for Global Change, Japan Agency for
Marine-Earth Science and Technology, 3173-25 Showa-machi, Kanazawa-ku, Yokohama-city,
Kanagawa, 236-0001, Japan

^{***}Laboratoire des Sciences du Climat et de l' Environnement, CEA-CNRS-UVSQ, Bât. 709.
Orme des Merisiers. F-91191 Gif sur Yvette. France

Abstract

Data assimilation, which minimizes the difference between modeled and observed statuses statistically along with objective criteria, has been applied frequently for terrestrial biosphere models in recent years. This approach enables highly accurate estimation, gap-filling of missing or sparsely distributed data, optimizing model parameters and initial state. The Markov chain Monte Carlo method is a useful data assimilation approach because of its relative simplicity in programming. However, this approach has been applied mainly for plot-scale simulation, which does not have any lateral interaction with other distant plots and does not require a large computational load. Ensemble Kalman filter, which has advantages in the simplicity of coding and the affordability of computational load despite the unavoidable reduction in the numbers of assimilated data points and optimized parameters, has been demonstrated recently in many field sites. The adjoint (4D-VAR) method enables us to estimate the uncertainty of parameters chronologically forward and backward with moderate computational load, but we also need to find a way to avoid the problems induced by the non-linearity of terrestrial biosphere models. In addition to the expectation for high accuracy simulations, those data assimilation approaches would also offer great potential for producing useful bi-products, e.g. objective analysis data sets, and supplemental evidence of the changes in physiological and ecological responses to environmental variations by looking at the chronological variation in their responses. While much greater effort should be made to overcome the variety of operational problems in those data assimilations, greater attention has been paid to improving terrestrial biosphere modeling through multi-disciplinary research fields, including ecological, climatological and global environmental studies.

Key words: Adjoint method, Ensemble Kalman filter, Markov chain Monte Carlo method, Parameter optimization.

キーワード: アジョイント法, アンサンブル・カルマンフィルター, パラメータ最適化, マルコフ連鎖モンテカルロ法

2012年7月27日 受付, 2013年1月16日 受理

[†]Corresponding Author: tomomichi.kato@lscce.ipsl.fr

^{††}Laboratoire des Sciences du Climat et de l' Environnement, CEA-CNRS-UVSQ

1. はじめに

陸域生態系は、その生物活動によって、大気との間で、エネルギーや、水、CO₂などの物質を交換しており、地球全体や地域の気候を決めているひとつ

の大きな要因である。観測の限界を超えて、過去・現在・将来の陸域生態系内もしくは、気候-陸域生態系間のエネルギー・物質循環を調べるためには、陸域生物圏モデルと呼ばれる、コンピューターシミュレーションモデルを利用することは不可欠である。70年代から発達して来たエネルギー・水についての陸面過程モデル (SiB, Sellers *et al.*, 1986; SiB2, Sellers *et al.*, 1996; BATS, Dickinson *et al.*, 1981; BATS2, Dickinson *et al.*, 1998 など)に加え, 90年代より急激に開発が進んだ炭素循環を推定する陸域生態系モデル (LPJ-DGVM, Sitch *et al.* 2003; SIM-CYCLE, Ito and Oikawa 2002 など)は, モデル単独の研究の興隆もさることながら, 将来の気候変化の予測を可能にする大気大循環モデル(GCM)に組み込まれて (Cao and Woodward 1998; Cox *et al.* 2000 など), 気候システム研究にも利用されており, まさに地球環境の研究にはなくてはならない存在になっている。しかし, また一方で, その予測の精度を上げることが重要な課題になっている。

陸域生物圏モデルを含む, コンピューターシミュレーションモデルでは, 実験や理論から導き出された多くのパラメータを使用しているが, それが必ずしも現実の観測と一致する結果を生むわけではない。多くのモデル開発者は, 経験と直感によって, いくつかのパラメータの値を少しずつ変えて何度もシミュレーションを繰り返し, 観測との誤差を埋めていく「チューニング」を行い, 最適と思われるパラメータを決定していることが多い。しかしながら, このいわば「恣意的」な方法では, 開発者の先入観が混入し, シミュレーションには多くの誤差が含まれている可能性が高い。そこで, これらのモデルシミュレーションにまつわる人為的誤差の問題を解決するために, 「データ同化」という手法を用いることがある。これは, 数値モデルに観測データを挿入(同化)することで, より統計的に観測に近い状態量の推定, データの時空間的な補間, パラメータの最適化を行う手法である。従来のモデルチューニングを主観的・場当たりのなものだとすれば, データ同化は客観的・系統的なモデルチューニングと言うこともできる。このデータ同化手法は, 特に, 20世紀中ごろから気象予報業務を中心に, 地球物理学の分野において急速に発展しており, 大気・海洋大循環モデルに盛んに用いられている (Ghil *et al.* 1997)。

一方で, 近年, これらデータ同化手法は, 陸域生物圏モデルにも用いられ, より最適なパラメータや状態量を得る努力がなされている。そこで, 本総説では, それらデータ同化手法の概説と, その陸域生物圏モデルへの適用例の紹介をする。

2. データ同化とは

2.1 データ同化の定義

データ同化は, 「数値シミュレーションに実測データを埋め込み, 馴染ませること」を意味する(細川, 2009)。数値シミュレーションでは, 初期条件, 境界条件, パラメータなどをあらかじめ与えることが必要であるが, たいていの場合, それらをはじめから正確に決めることはできない。しかも, 与える条件・パラメータによって, シナリオ(計算結果)は, 大きく変化する(破線, 図 1)。一方で, データ同化は, 観測データ(星印, 図 1)を, 数値シミュレーションモデルに埋め込み, 馴染ませていくことで, 現象を尤もらしく表現する推定値(実線, 図 1)を得ることができる。

データ同化の目的は, 大きく分けて二つある(細川, 2009)。一つ目は, 「実測データを用いて数値シミュレーションモデルの精度・性能を改善する」ことである。つまり, 図 1 のように, シミュレーションに対して, 観測データを同化することで, よりよい初期条件やパラメータのセットを導き出し, 現在・過去・将来のモデル推定精度を向上させることである。たとえば, ある地点の陸域生態系シミュレーションをする際に, 同地点における CO₂ フラックスの観測データが得られたと想定する。その際に, 植物の光合成・フェノロジー, 土壌の根圏深度・呼吸の温度依存性などのパラメータや, 植物・土壌の炭素蓄積量の初期値を, 系統的に繰り返し変化させ, 観測データをより良く再現するパラメータ・初期値およびシミュレーション結果を得ることである。二つ目の目的は, 「物理法則を表現するシミュレーションモデルを用いることで, 観測の不足を補ったり, 観測誤差を修正したりする」ことである。一般的に,

図 1 データ同化の概念。

野外観測においては、欠測の発生や、機器の劣化による観測精度の低下や、観測地点の分布に偏りがある場合が多く、時間的・空間的に均質・連続な観測データを取得することは難しい。一方で、数値シミュレーションでは、時間的・空間的に偏りの無い連続データを出力することは容易だが、逆に、利用した物理過程(物理法則)のみによって挙動が支配されているため、必ずしも推定結果が現実と合うとは限らない。そこで、データ同化によって、モデルを存在する観測データに対して最適化させ、物理過程に則って尤もらしい数値シミュレーションを行うことで、観測データの不足を補うことができる。つまり、時間・空間スケールにおいて離散的な(飛び飛びの)観測データから、観測のない時間や場所をもカバーする均質・連続な状態量を提供することが可能である。たとえば、陸域生物圏モデルのシミュレーションでは、入力値として時空間的に一定間隔に配列された客観解析(再解析と呼ばれることもある)気候データと呼ばれるデータ(NCEP/NCAR, CRU, ERA40, JRA-25 など)を利用することが多いが、それらは世界の気象測候所の観測データを大気大循環モデルや領域気候モデルに同化させて作られおり、この二つ目の目的を具体化した最たる例である。

2.2 データ同化手法発達背景

データ同化は、数値天気予報に高精度の初期値を与える必要性から発展してきた(露木・川畑編, 2008)。1950年代ごろの初期の数値予報実験では、手書きの実況天気図から、数値モデルの格子点上の値を人間が読み取って、それをコンピューターに入力していた。しかし、それでは、コンピューターによる予測の時間よりも、はるかに多くの時間をデータの入力に費やしてしまうことになる。そこで、何らかの数値計算手法を利用して、空間的に不規則に分布した観測データを、数値モデルの規則的な格子上にすばやく書きだす必要があった。これは、客観解析と呼ばれ、人間が行う主観解析と区別された。

これら、初期のデータ同化においては、空間的に離散化した(まばらに存在した)観測データを、より尤もらしく空間内挿し、より尤もらしいモデルの初期値を与えることが目的とされていた。たとえば、線形関数を複数結合した組み合わせ関数に対して、最小二乗法でパラメータを決定し、各気象場を空間内挿していく関数当てはめ法(Panofsky, 1949; Gilchrist and Cressman, 1954)、各格子周辺の複数観測点のデータを、さまざまな形で重み付けして第一推定値(データ同化前のモデル推定値)を修正していく逐次修正法(Bergthorsson and Döös, 1955)や、格子周辺観測点の観測値と第一推定値との差である解析

誤差について、その分散が最小になるように重み付けを行い、解析値を得る最適内挿法(Eliassen, 1954; Gandin, 1963)などの、現在からすると比較的簡易な方法がとられていた。

1980年代に入ると、4次元変分法や、カルマンフィルターなどの4次元場(空間3次元+時間1次元)に適用可能な、高度なデータ同化方法の開発が始まった。これらの方法では、最適内挿法などとは異なり、データの時間軸方向の流れに依存する背景誤差共分散(モデルの第一推定値の誤差の共分散)を用いるので、より高精度の解析値を得ることができる。さらに、数値モデルで計算可能な物理量に対するものならば、原理的にどんな観測データでも同化できることが圧倒的な利点である。変分法では、オイラー・ラグランジュ方程式を大規模空間で解くことは困難だったが、アジョイント法を利用することで、4次元変分法へと拡張がなされていった。その後1990年代になって、ECMWF(ヨーロッパ中期気象予報センター; Rabier *et al.*, 2000 など)や、日本の気象庁(Ishikawa and Koizumi, 2002)をはじめ、おおくの気象予報現業センターで主要な4次元データ同化手法として、利用されている。一方、カルマンフィルターは、離散時間の線形システムに対するフィルタリング理論であるが(Kalman, 1960)、それを連続時間の線形システムに拡張利用し(Kalman and Bucy, 1961)、アポロ計画等の宇宙工学をはじめ、現在もさまざまな分野で幅広く利用されている。特に、気象学分野で盛んに利用されているが(Ghil *et al.*, 1981; Parish and Cohn, 1985)、そのままでは、予報誤差の共分散行列の時間発展を陽に計算するために、計算資源を多く必要とし、現実的な適用が難しい。そこで、計算時間が少なく済む、近似的なカルマンフィルターが開発されてきている。そのなかでも、アンサンブル・カルマンフィルター(Evensen, 1994)は、最も現実的な方法とされ、さまざまな気象関連の数値予報センターにおいて利用の検討がなされており、4次元変分法との性能の比較もたびたび行われている(Lorenc, 2003; Kalnay *et al.*, 2007a, b)。また、GCMに付随する水・熱循環などの主に物理過程を扱う陸面過程モデルについては、NASAのゴッダード宇宙飛行センターが推進するGLDAS(全球陸域データ同化システム)で、衛星データを同化するためのデータの整備や、同化技術の開発が始められている(Zaitchik *et al.*, 2010; Gottschalck *et al.*, 2005 など)。

一方で、炭素・窒素循環や、植生動態などの主に生物過程を扱う陸域生物圏モデルへの本格的データ同化手法の適用は、陸域生物圏モデルの開発の歴史が比較的浅いことと、考慮すべきパラメータの多

さ、プログラムコードに条件文等の非線形な取り扱いが多数存在することから、パラメータ最適化時の挙動が不安定であることが多く、いまだ発展途上である。しかし、陸域生物圏モデルシミュレーションのほとんどは、時間方向の広がりを持つが、鉛直・水平方向のグリッド間の空間的つながりを持たないか、もしくははじめから 1 地点のみを対象とする研究が多い。そのため、よりコード設定が簡単なマルコフ連鎖モンテカルロ (MCMC) 法を利用したデータ同化が、1990 年代後半から行われることが多くなってきた (Braswell *et al.*, 2005 など)。さらに、2000 年代より、前述の地球物理学の知見を応用し、アンサンブル・カルマンフィルター (Williams *et al.*, 2005 など) や、アジョイント法 (Rayner *et al.*, 2005) などの高度なデータ同化手法を取り入れた研究も進みつつある。これらの発展には、データ同化の対象となる、水・CO₂ フラックスなどの地上観測データや、LAI (葉面積指数)・FAPAR (光合成有効放射吸収率)、土壌水分などの衛星から得られる広域観測データが、急速に増えていることや、陸域生物圏モデルが組み込まれた大気大循環モデルが増加し、その精度の向上が気候システム研究においても大きな課題になっていることが影響していると思われる。

3. 陸域生物圏モデルにおけるデータ同化

3.1 データ同化手法の種類

簡便な手法

ナッジング

黄金分割法

MCMC 法

高度なデータ同化手法

アンサンブル・カルマンフィルター

アジョイント法 (4 次元変分法)

陸域生物圏モデルに利用されている主なデータ同化手法を、上に挙げた。下に行くほど、実装が難しくなっていく。ナッジング・黄金分割法・MCMC 法については、モデルの計算結果 (第一推定値もしくは予報値) と、真値との差である背景誤差の統計情報が時間変化しないため、静的なデータ同化手法と呼ばれ、逆に、アンサンブル・カルマンフィルターとアジョイント法は、背景誤差の統計情報が時間変化するため、動的なデータ同化手法と呼ばれる (淡路ら, 2009)。また、アンサンブル・カルマンフィルターは、計算ステップごとに、モデル推定値を修正するので、逐次型データ同化手法と呼ばれ、逆に、アジョイント法は、非逐次型データ同化手法と呼ばれる (淡路ら, 2009)。本項では、陸域生物圏モデルにおけるデータ

同化手法のうち、すでにさまざまなグループによって利用されている MCMC 法と、アンサンブル・カルマンフィルター、今後、大きな飛躍が期待されるアジョイント法 (4 次元変分法) を中心に紹介する。ナッジングと黄金分割法については、その他のデータ同化手法適用例として、最後に紹介をする。なお、ここで紹介する三つの手法は、すべてベイズ推定に基づいている。ベイズ推定とは、ある観測事象をもって、その原因となった事象を推定するための確率論的方法である。具体的には、事後の確率分布は、観測値が得られる確率を示す尤度と、事前の確率分布の積に比例するとする、ベイズの定理を用いる。

3.2 データ同化手法の実際

3.2.1 MCMC (マルコフ連鎖モンテカルロ) 法

a. 計算スキーム

ここでは、Braswell *et al.* (2005) を参考にして、説明を行う。まず、ベイズ推定を陸域生物圏モデルに応用し、データ同化の対象となる n 個データを持つ観測データのベクトル $\mathbf{y} (= (y_1, y_2, \dots, y_n))$ が得られた時に、パラメータのベクトルが $\mathbf{p}$ である確率密度分布 $q(\mathbf{p}|\mathbf{y})$ を以下のように表す。

$$q(\mathbf{p}|\mathbf{y}) = vL(\mathbf{y}|\mathbf{p})\rho(\mathbf{p}) \quad (1)$$

ここで、 v は正規化定数である。 $L(\mathbf{y}|\mathbf{p})$ は、パラメータ $\mathbf{p}$ を与えたときに観測値 $\mathbf{y}$ が得られる確率を表す尤度関数といい、 $\rho(\mathbf{p})$ は、 $\mathbf{p}$ の事前確率分布である。 $L(\mathbf{y}|\mathbf{p})$ は、モデル推定値と観測値の差である観測誤差が、ガウス分布 (正規分布) を持つことを仮定すると、次のように、観測誤差を正規化した形で記述される。

$$L(\mathbf{y}|\mathbf{p}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(y_i - \mu_i)^2}{2\sigma_i^2}\right) \quad (2)$$

ここで、 n はデータ数、 y_i と μ_i は時刻 i の観測値とモデル推定値、 σ_i は時刻 i の観測の標準偏差である。 σ_i は、CO₂・水フラックスなどの観測では、時間・季節・気象条件に応じて変動すると考えられるが、それを求めるのは容易ではない。そこで、 σ_i に、任意の固定した値 (たとえば、生態系 CO₂ 交換 (NEE) フラックスに対しては、常に 5 $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ など)、もしくは観測データの絶対値に対して一定の割合をかけた値 (たとえば、NEE フラックスの絶対値に対して日中は 20%、夜間は 50% など) を用いることも多い。 $\rho(\mathbf{p})$ についても、2 式と同様に計算することができる。本方法では、この $q(\mathbf{p}|\mathbf{y})$ を最大にするパラメータ $\mathbf{p}$ を得ることが目的である。しかし、陸域生物圏モデルの構造は、たいてい非線形で、簡単に

は逆問題を解くことができないため、最適な $\mathbf{p}$ を直ちに導くことはできない。そこで、乱数とあらかじめ設定したパラメータの移動範囲から作り出した、微小に異なるパラメータのセットによってモデルシミュレーションを繰り返し、より $q(\mathbf{p}|\mathbf{y})$ を大きくする $\mathbf{p}$ を探索する(パラメータの最適化)。

パラメータの最適化によく用いられるものとして、Metropolis-Hastings アルゴリズム (Metropolis *et al.*, 1953) がある(図 2)。まず、乱数によって生成された新パラメータ $\mathbf{p}' (= \mathbf{p}_t + \Delta \mathbf{p})$ を利用し、前時間ステップ t との確率密度関数の割合 $\alpha (= q(\mathbf{p}'|\mathbf{y})/q(\mathbf{p}_t|\mathbf{y}) = L(\mathbf{y}|\mathbf{p}')\rho(\mathbf{p}')/L(\mathbf{y}|\mathbf{p}_t)\rho(\mathbf{p}_t))$ を計算する。 α が 1 以上の場合、次パラメータ $\mathbf{p}_{t+1}$ に、新パラメータ $\mathbf{p}'$ を採用する。 α が 1 未満の場合、次パラメータ $\mathbf{p}_{t+1}$ は、 α の確率で $\mathbf{p}'$ 、 $1-\alpha$ の確率で $\mathbf{p}_t$ になるとされる。ここで、 α の採用閾値に、乱数によって一様密度で発生させた 0 から 1 の間の値 u を利用することが多い。つまり、 α が u 以上の場合、次パラメータ $\mathbf{p}_{t+1}$ に $\mathbf{p}'$ を、 α が u 未満の場合、 $\mathbf{p}_t$ を採用することである。このように、過去の全ステップの情報を残しつつ、パラメータ場に対してランダムに移動をさせることで、 $q(\mathbf{p}|\mathbf{y})$ を極大付近に早く到達させることを目指す。この最適化のための計算ステップは、数万から数十万回繰り返されることが多い。

b. 適用例

Braswell *et al.* (2005) は、生態系炭素フラックスモデル SIPNET に対して、同方法を適用し、米国

Harvard Forest の 10 年間にも亘る 30 分間隔の CO_2 フラックスデータを同化させた。その結果、モデルと観測の間の二乗平均平方根誤差 (Root-Mean-Squared error: RMSE) は、同化前の 1.37 gCm^{-2} から、同化後には 0.97 gCm^{-2} に減少し、年間平均値の推定誤差も、 $57.2 \text{ gCm}^{-2} \text{ yr}^{-1}$ から $49.2 \text{ gCm}^{-2} \text{ yr}^{-1}$ に減少した。このように、モデルの推定精度は最適化以前より向上はしたが、最適パラメータは、フラックスの日・季節変化に強く制御されてしまい、年々変化の再現性の向上については改良の余地があることを示した。

Knorr and Kattge (2005) は、陸面過程モデル BETHY を米国カンザス州の C4 長草草原とオランダ Loobos の Scots pine サイトにおいて、30 分間隔の潜熱と CO_2 フラックスデータに対して同化した。その結果、最適化の際に値の連動するパラメータのグループがあり、しかも、それはサイトごとに異なることを報告している。また、Loobos サイトにおいてはわずか 7 日間分のデータについてパラメータの最適化を行った後、2 年間の CO_2 フラックスの再現に成功している。その他、Yuan *et al.* (2012), Wu *et al.* (2009), Lasslop *et al.* (2008) など、渦相關フラックスデータに対する同化法の適用例は多い。

フラックス以外では、Iizumi *et al.* (2009) が、水稻生育モデル SIMRIW を、日本列島の北日本・西日本の各 7・9 県の過去 25 年間 (1979-2003) の出穂日と収量データに対して同化し、13 種類のパラメータを最適化した。彼らは、各連鎖の 3 万回の推定から、ランダムに選んだ 5 千の推定を利用して、事後のパラメータの確率分布や各県ごとの出穂日と収量の不確実性の評価を行っている。さらに、Iizumi *et al.* (2011) では、SIMRIW を発展させた、ベイズ推論に基づいた広域コメ収量予測モデル PRYSBI を開発した。そこでは、MCMC によって推定された 5 千個のパラメータセットと累積確率密度関数を利用して、9 つの気候モデルによる 17 の将来気候変化シナリオが水稻収量に与える影響の評価を試みている。その結果、1990 年代に対して、2050 年代と 2090 年代には、水稻収量の中央値は、北海道で +17.2% (90%信頼区間: -5.2% から +40.3%) と +26.9% (+6.3% から +51.2%)、愛知県では +2.2% (-15.0% から +14.9%) と -0.8% (-33.4% から +17.9%) の変化をした。

3.2.2 アンサンブル・カルマンフィルター

a. 計算スキーム

アンサンブル・カルマンフィルター (EnKF) は、カルマンフィルター (KF) を利用する数ある同化手法の中で、非線形モデルに適用可能な方法である。基礎となる KF は、予報ステップごとに、カルマンゲインと呼ばれる重み付け行列を利用して、モデルによ

図2 MCMC 法を用いた最大尤度推定によるパラメータ最適化スキーム

る予報値(第一推定値)を観測値に近づけた解析値を得て、その解析値を次予報ステップの初期値に用いるという繰り返しを行う(図 3, 4)。それによって、モデルの背景誤差共分散(パラメータ最適化後のモデル推定値の誤差の共分散)そのものを時間発展させる。しかし、海洋大循環モデルなどの大規模モデルの場合、たとえば後述の予報誤差共分散の計算では、仮に各格子点の予報変数の総計を $O(10^8)$ だとすると、記憶容量に $O(10^{16})$ が、時間発展の計算に $O(10^{24})$ の演算回数が必要であり、理論通りの計算を行うには、確保しなければならない計算資源が膨大で、現実的には利用できない(淡路ら, 2009)。そこで EnKF では、わずかに異なる初期値からスタートした各モデルシミュレーションの結果について、その誤差の時間発展におけるアンサンブル(集合)を利用し(図 5)、モデルの背景誤差共分散を統計的に近似する。これによって、パラメータ最適化の計算時間を大幅に短縮することができる。

ここから、EnKF の導出を紹介するが、式展開については、大幅に省略しているの、詳細は淡路ら(2009)、三好(2005)、Evensen(1994)を参照してい

図 3 カルマンフィルターによる予報値と解析値の時間発展。

図 4 カルマンフィルターによるデータ同化スキーム(淡路ら(2009)図 3.2 に加筆)。

ただきたい。なおここでは、淡路ら(2009)の説明を参考にした。

時刻 t において、モデル状態変数(予報変数) $\mathbf{x}^f$ の真値 $\mathbf{x}_t^{\text{true}}$ から、観測値 $\mathbf{y}_t$ を次のように表す。

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t^{\text{true}} + \mathbf{r}_t \quad (3)$$

ここで、 $\mathbf{H}$ は線形の観測行列であり、モデルの変数から観測量への変換や、格子点から離れた観測点への空間内挿を示す。 $\mathbf{r}$ は観測誤差ベクトルである。次に、同化後の解析値 $\mathbf{x}_t^a$ (解析値)は、モデルによる予報値である $\mathbf{x}_t^f$ と、観測値 $\mathbf{y}_t$ とモデルから得た最適な加重平均の行列 $\mathbf{K}_t$ によって次のように与えられる。

$$\mathbf{x}_t^a = \mathbf{x}_t^f + \mathbf{K}_t (\mathbf{y}_t - \mathbf{H}_t \mathbf{x}_t^f) \quad (4)$$

ここで、モデルと観測の差を示す $\mathbf{y}_t - \mathbf{H}_t \mathbf{x}_t^f$ はイノベーションと呼ばれている。 $\mathbf{K}_t$ はカルマンゲインと呼ばれ、次のように計算される。

$$\mathbf{K}_t = (\mathbf{P}_t^f \mathbf{H}_t^T) (\mathbf{R}_t + \mathbf{H}_t \mathbf{P}_t^f \mathbf{H}_t^T)^{-1} \quad (5)$$

ここで、 $\mathbf{P}_t^f$ は背景誤差共分散(モデルによる予報値の誤差の共分散)、 $\mathbf{R}$ は観測誤差共分散(観測データに含まれる誤差の共分散)、 T は行列の転置である。次に、解析誤差共分散(最適化後の解析値の誤差共分散)である $\mathbf{P}_t^a$ は、次のようにあらわされる。

$$\mathbf{P}_t^a = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \mathbf{P}_t^f \quad (6)$$

$\mathbf{I}$ は単位行列である。ここで、背景誤差共分散の時間発展を示すリヤプノフ方程式(導出を省略)と、5, 6式と組みあわせると、リカッチ方程式(7式)という背景誤差共分散の時間発展を求めることができる。

$$\mathbf{P}_t^f = \mathbf{M} \left\{ \mathbf{P}_{t-1}^f - \mathbf{P}_{t-1}^f \mathbf{H}_{t-1}^T (\mathbf{R}_{t-1} + \mathbf{H}_{t-1} \mathbf{P}_{t-1}^f \mathbf{H}_{t-1}^T)^{-1} \right. \\ \left. \mathbf{H}_{t-1} \mathbf{P}_{t-1}^f \right\} \mathbf{M}^T + \mathbf{\Gamma} \mathbf{Q}_{t-1} \mathbf{\Gamma}^T \quad (7)$$

$\mathbf{M}$ は遷移状態行列、 $\mathbf{\Gamma} \mathbf{Q} \mathbf{\Gamma}^T$ は時間発展に伴い混入したシステムノイズの共分散行列である。このリカッチ方程式を利用すれば、モデルの力学に応じた背景誤差の時間発展を、モデルの状態変数を利用せずに ($\mathbf{x}_t^f$ を用いずに) 記述することができる。この点がカルマンフィルターの最大の特徴である。ここまでが、KF に必要な式の非常に簡略化した展開であり、4, 5, 7式を利用すると、モデルの逐次的なデータ同化を行うことができる。

次に、EnKF では、真値や誤差の実現値を多数発生させ、実際は正規分布するかどうかかわからない確率分布を、KF 同様に正規分布に近似して同化

図5 カルマンフィルターとアンサンブル・カルマンフィルターの違い(淡路ら(2009)図 3.6 に加筆)。アンサンブル・カルマンフィルターは、複数のアンサンブルメンバーを利用して、カルマンフィルターの確率密度を正規分布として近似する。

を行う(図 5b)。ここで、観測値 $y_{1:t-1}$, $y_{1:t}$ のもとでの x_t^{true} の予報値と解析値を $x_t^{f(l)}$, $x_t^{a(l)}$ (ただし, $l = 1, \dots, L$) とすると、各値の L 個(任意の数)のまとまり $\{x_t^{f(1)}, \dots, x_t^{f(L)}\}$, $\{x_t^{a(1)}, \dots, x_t^{a(L)}\}$ をアンサンブルと呼び、個々の値をアンサンブルメンバーという。そして、 $x_t^{f(l)}$ の時間発展は力学モデルとして下記のように記述される。

$$x_t^{f(l)} = M \left(x_{t-1}^{a(l)} \right) + G w_{t-1} + \Gamma q_{t-1}^{(l)} \quad (l=1, \dots, L) \quad (8)$$

M は、非線形力学モデル、 G は外力行列、 w は強制力(境界条件など)、 Γq は時間発展に伴うシステムノイズである。つまり、図 5a の確率分布を近似するアンサンブル $\{x_t^{f(l)}\}$ を求めるためには、構成するアンサンブルメンバー $x_t^{a(l)}$ ごとにシステムノイズの実現値 $q_{t-1}^{(l)}$ を発生させ、式 8 を利用すればよい。次に、アンサンブルによって近似された正規分布の平均と共分散行列は、バーを点けて記述される。

$$\bar{x}_t^f = \frac{1}{L} \sum_{l=1}^L x_t^{f(l)} \quad (9)$$

$$\bar{P}_t^f = \frac{1}{L-1} \sum_{l=1}^L \left(x_t^{f(l)} - \bar{x}_t^f \right) \left(x_t^{f(l)} - \bar{x}_t^f \right)^T \quad (10)$$

前述のカルマンフィルターでの 7 式では、 P_t^f を計算するのに非常に労力のかかる行列の計算を必要とするが、10 式で求めた $\bar{P}_t^f$ はアンサンブル計算の結果から共分散を計算するだけでよく、計算が容易であることが分かる。また、10 式の $\bar{P}_t^f$ を用いて、カルマンゲイン式(5 式)を近似し、 $\bar{K}_t$ を以下のように求めることができる。

$$\bar{K}_t = \bar{P}_t^f H_t^T \left(R_t + H_t \bar{P}_t^f H_t^T \right)^{-1} \quad (11)$$

以上の 8 式や 11 式を逐次的に使用して同化を行うのが EnKF であり、同化後のアンサンブル $\{x_t^{a(l)}\}$ の平均値 $\bar{x}_t^a$ 、共分散行列 $\bar{P}_t^a$ に対して、以下の関係式が近似的に成り立つ。

図6 アンサンブル・カルマンフィルターによって最適化されたCO₂フラックス(Williams *et al.* 2005のFig. 2)。オレゴン州のPonderosa 松のサイトにおける、2000-2002年のNEE(上図)、総呼吸量(R_{tot} : 中図)、GPP(下図)の日平均の観測値(○)と解析値(実線)。グレー部分は、アンサンブル平均に対する標準偏差で、解析に対する信頼限界の指標になる。モデルは、シンプルなボックスモデル(Aggregated canopy model: ACM)。

$$\bar{\mathbf{x}}_t^a = \bar{\mathbf{x}}_t^f + \mathbf{K}_t \left(\mathbf{y}_t - \mathbf{H}_t \left(\bar{\mathbf{x}}_t^f \right) \right) \quad (12)$$

$$\bar{\mathbf{P}}_t^a = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \bar{\mathbf{P}}_t^f \quad (13)$$

なお、 L は10から1000の間に設定されることが多い。このように、EnKFでは、アンサンブルの予報値の時間発展を元に、近似的に尤も確からしい解析値の時間発展を得る。完全な式展開は、淡路ら(2009)を参照されたい。

b. 適用例

英国エジンバラ大のグループが、いくつかのEnKFによる陸域生態系のデータ同化の研究を発表している。Williams *et al.* (2005)は、シンプルなボックスモデル(Aggregated canopy model: ACM)を利用して、オレゴン州のPonderosa 松のサイトでの炭素循環のシミュレーションを行った。その際に、EnKFを適用して、9つの未知のパラメータと5つの炭素プールの初期値の最適化を行った。同化対象とするデータは、3年間の日平均の渦相関法とチャンバ法によって求められたNEE(生態系正味CO₂交換)、GPP(生態系総一次生産)、総生態系呼吸の三つのフ

ラックスと、数点しかない各種炭素プールの観測データで、アンサンブルメンバーの数は200に設定した。その結果、各フラックスは、観測に近い季節変化を示し、モデルアンサンブルの標準偏差の中にほとんどの観測データが分布するようになり(図6)、フラックスに対する不確実性を半分近くにまで減少させることに成功した。また、土壌有機物質、木質バイオマス、木質残渣などの炭素プールに関する長期的なデータは、同化には有効であったが、その時間的な点数を増やすことが精度向上に重要であることを示した。

次に、同グループから、Quaife *et al.* (2008)は、ACMを内包するデータ同化システムDALECを利用して、衛星MODISによる群落表面の二方向反射特性(BRF)のデータを同化した。彼らは、まず、キャノピー反射率の計算に採用した幾何光学放射伝達(GORT)モデル(Ni *et al.*, 1999)のパラメータの中から、キャノピー形態や土壌反射率にかかわる7つのパラメータに対して、乱数を利用して微少に異なるパラメータ値のセットを 3×10^5 個作成した。次に、それらと計算結果であるBRFとの関係にLook-Up Table(LUT; Weiss *et al.*, 2000)を作成し、簡易なインバージョンを行った。そして、上出のオレゴン州のPonderosa 松のサイトにおいて、2000-2002年の間に炭素循環シミュレーションを行った。その際、MODIS BRFをEnKFによって同化し、GORTの7つのパラメータを最適化した。アンサンブルメンバーの数は同じく200であった。その結果、GPP、NEP(生態系純生産)と生態系呼吸フラックスについて、モデルアンサンブルの標準偏差としての不確実性を、3年間の平均で、それぞれ212.2, 834.5, 684.8 gCm⁻²から、151.3, 96.9, 155.3 gCm⁻²に減少させることに成功した。しかしながら、葉バイオマスは過大評価された。

他グループでは、Mo *et al.* (2008)による生物圏モデルBEPSへのEnKFの適用や、Chen *et al.* (2008)の渦相関フラックスデータのGPPと呼吸の分割モデルへのSmoothed EnKFの適用がある。

またEnKF以外のKFの適用では、Barbu *et al.* (2011)が、衛星による土壌水分指数(Soil Wetness Index)とLAIデータをSimplified Extended KFによって陸面過程モデルISBA-A-gsへ同化し、フランスの草原での水・炭素循環の推定を行った。その結果、観測とモデルの間のRMSEは2001-2007年の根圏土壌水分で13%減少し、2005-2007年のCO₂フラックスで5%減少した。

3.2.3 アジョイント法 (4 次元変分法)

a. 計算スキーム

MCMC 法のところで説明した、パラメータのベクトル $\mathbf{p}$ を与えたときの確率密度関数 $q(\mathbf{p}|\mathbf{y})$ を最大にする $L(\mathbf{y}|\mathbf{p})$ や $\rho(\mathbf{p})$ を求めることを、本方法でも目指す。

まず、 $L(\mathbf{y}|\mathbf{p})$ を、時刻 t の観測値 $\mathbf{y}_t$ に対するモデル推定値の不一致を表すコスト関数 $J_y(\mathbf{p})$ の負の指数関数で示す。

$$L(\mathbf{y}|\mathbf{p}) = \exp(-J_y(\mathbf{p})) \quad (14)$$

ここで、事前パラメータ $\mathbf{p}$ と同化された観測値 $\mathbf{y}_t$ が、独立であり、ガウス分布を持つ場合、 $J_y(\mathbf{p})$ は以下のようにならわされる

$$J_y(\mathbf{p}) = \sum_{t=0}^n \frac{1}{2} (\mathbf{H}_t(\mathbf{p}) - \mathbf{y}_t)^T \mathbf{C}_{y_t}^{-1} (\mathbf{H}_t(\mathbf{p}) - \mathbf{y}_t) \quad (15)$$

$\mathbf{C}_{y_t}$ は、時刻 t の観測に対する誤差の共分散行列で、観測データの数だけの次元を持つ対角行列である。 $\mathbf{H}_t$ は観測演算子である。同様に、パラメータの事前確率分布 $\rho(\mathbf{p})$ に対しても、次のように記述される。

$$\rho(\mathbf{p}) = \exp(-J_p(\mathbf{p})) \quad (16)$$

$$J_p(\mathbf{p}) = \frac{1}{2} (\mathbf{p} - \mathbf{p}_0)^T \mathbf{C}_p^{-1} (\mathbf{p} - \mathbf{p}_0) \quad (17)$$

$\mathbf{p}_0$ は、事前パラメータのベクトル、 $\mathbf{C}_p$ は、事前の誤差共分散行列であり、パラメータの数だけの次元を持つ対角行列である。1 式の $q(\mathbf{p})$ を最大にする最適化を行う場合、 $L(\mathbf{p})$ と $\rho(\mathbf{p})$ 指数関数の引数が最大になればよい。すなわち、引数にマイナスをかけた値が最小になればよく、15, 17 式の $J_y(\mathbf{p})$ と $J_p(\mathbf{p})$ をあわせた次のコスト関数 $J(\mathbf{p})$ を、最小にすればよい。

$$J(\mathbf{p}) = J_y(\mathbf{p}) + J_p(\mathbf{p}) = \sum_{t=0}^n \frac{1}{2} (\mathbf{H}_t(\mathbf{p}) - \mathbf{y}_t)^T \mathbf{C}_{y_t}^{-1} (\mathbf{H}_t(\mathbf{p}) - \mathbf{y}_t) + \frac{1}{2} (\mathbf{p} - \mathbf{p}_0)^T \mathbf{C}_p^{-1} (\mathbf{p} - \mathbf{p}_0) \quad (18)$$

ここでは、繰り返し計算によって、コスト関数 $J(\mathbf{p})$ を、極小に近づける作業を行う。まず、初期パラメータに対して、コスト関数を計算する。次に、計算されたコスト関数の $\mathbf{p}$ に対する 1 階導関数 (勾配) である $\nabla J(\mathbf{p})$ を、次のように求める。

$$\nabla J(\mathbf{p}) = \sum_{t=0}^n \mathbf{M}_0^T \mathbf{M}_1^T \cdots \mathbf{M}_{t-1}^T \mathbf{H}_t^T \mathbf{C}_{y_t}^{-1} (\mathbf{H}_t(\mathbf{M}_t(\mathbf{p}) - \mathbf{y}_t)) + \mathbf{C}_p^{-1} (\mathbf{p} - \mathbf{p}_0) \quad (19)$$

$\mathbf{M}$ は、非線形モデル演算子 $M(\mathbf{p})$ のヤコビアン (ヤコビ行列) といい、パラメータごとに $M(\mathbf{p})$ を偏微分した値である $\partial M(\mathbf{p}) / \partial \mathbf{p}$ を持つ関数行列であり、ここでは接線形モデル演算子として扱われる。 $\mathbf{H}_t$ は、 H_t の接線形演算子である。次ステップでは、このヤコビアンが最大になる、すなわちコスト関数が最も大きく減少する方向に、微小に変化させた新しいパラメータを設定する。そして、再度ヤコビアンを利用して、各パラメータの微小変化に対する最も傾きのきつい方向を求める (図 7)。この反復を行うことで、コスト関数を徐々に減少させ、観測データに対して、最適なパラメータを得る。詳しい定式化は淡路ら (2009) を参照のこと。

ここで、行列 $M(\mathbf{p})$ 内の変数がすべて実数のとき、 $\mathbf{M}^T$ は $M(\mathbf{p})$ のアジョイント (随伴行列) といい、誤差の原因を、時間軸をさかのぼる方向に伝播させる役割を持つ。アジョイントを利用したヤコビアンの計算には、自動微分アルゴリズム (例、TAF; Giering and Kaminski, 2002) を利用していることが多い。

パラメータに対する不確実性を、コスト関数 $J(\mathbf{p})$ のパラメータ $\mathbf{p}$ に対する 2 階導関数としてヘッシアン (ヘッセ行列) で評価する。ヘッシアンは、コスト関数のパラメータ微小変化に対する曲率を示し、そ

図 7 パラメータ場におけるコスト関数の分布 (Tarantola, 1987 より改変)。ここでは、パラメータの次元 (種類) が 2 つの場合を想定している。通常の最適化パラメータは、多次元のため本図のような可視化はできない。正接接線は、現在地 $\mathbf{p}_n$ における最も傾斜のきつい方向 ($\mathbf{y}_n$) へのコスト関数の接線を示しており、次ステップのパラメータ決定に利用される。

図8 CCDAS によるアジョイントを利用したパラメータ最適化スキーム

の逆行列を計算することで、事後パラメータによる不確実性を表現している。この計算では、線形関数を前提とするガウス分布を想定しているが、実際はモデルに非線形の関数も多く含まれるために、あくまで近似値としての不確実性を表現している。なお、ヘッシアンが、不確実性をあらわす理由は、コスト関数の2階導関数が、解析誤差共分散の逆数にあたるためである(詳細は、淡路ら(2009) 4.3章を参照)。

b. 適用例

これまでのところ、陸域生物圏モデルにアジョイント法を適用することに成功したのは、英国ブリストル大の Wolfgang Knorr, Marko Scholze (現所属は、スウェーデン・ルンド大)らのグループの開発した炭素循環データ同化システム(Carbon Cycle Data Assimilation System: CCDAS)のみであるため、ここでは、CCDASの仕組みと適用例を紹介する。

CCDASは、 $2^\circ \times 2^\circ$ の水平解像度を持つ大気輸送モデル TM2 (Heimann, 1995)と陸域生態系水・炭素循環モデル BETHY (Knorr, 2000)を組み合わせ、陸域の炭素フラックスを forward と inverse の両方向に計算することのできる計算スキームである。BETHYでは、人工衛星観測によって得られた FAPAR と、あらかじめ用意された気候データより、各陸面格子の CO_2 フラックスを計算する。その地表面フラックスは TM2 に送られ、全球の大気 CO_2 濃度分布に反映される。次に、そのモデル推定値と地上 41 箇所の

大気 CO_2 濃度観測データに対して、モデル推定値と観測値との差をあらわすコスト関数 $J(\mathbf{p})$ を以下のよう、計算する。

$$J(\mathbf{p}) = \frac{1}{2}(\mathbf{M}(\mathbf{p}) - \mathbf{c})^T \mathbf{C}_c^{-1}(\mathbf{M}(\mathbf{p}) - \mathbf{c}) + \frac{1}{2}(\mathbf{p} - \mathbf{p}_0)^T \mathbf{C}_p^{-1}(\mathbf{p} - \mathbf{p}_0) \quad (20)$$

ここで、 $\mathbf{M}$ はモデル、 $\mathbf{c}$ は観測された CO_2 濃度、 $\mathbf{C}_c$ は $\mathbf{c}$ に対する不確実性である。CCDAS では、このコスト関数の値を最小化するように、繰り返し計算をおこない、観測された CO_2 濃度について、最適なパラメータを求める(図8: Scholze 2003; Rayner *et al.*, 2005; Scholze *et al.*, 2007; Ziehn *et al.*, 2011a and 2011b)。そして、Rayner *et al.* (2005)は、CCDASを全球の大気 CO_2 濃度観測データの同化に応用し、1979-1999年までの各観測ステーションにおける大気 CO_2 濃度の季節変化・年々変化の推定精度の向上に成功した。たとえば、図9では、アラスカ・Barrow とコロラド・Niwt Ridge における CO_2 濃度の平均季節変化を観測に近く表現できているのがわかる。生態系モデル BETHY のみのパラメータの最適化で、大気 CO_2 濃度の時間変化をここまで再現できることは特筆に値する。

CCDASは基本的に、全球の二酸化炭素濃度の分布を推定するためのデータ同化システムだが、一方で、サイトスケールの水循環についてのデータ同化も試みられている。Kato *et al.* (2013)は、ボツワナ Maun のサバンナサイトを対象とし、渦相関法による潜熱フラックスと衛星による FAPAR の2つの観測データの同化を試みた。その結果、1つの観測データを同化する時よりも、2つの観測データを同時に同化した方が、よりバランスの取れた再現ができることを示した。このことは、観測データに誤差が大きく含まれると判断されたときに、複数のデータを同化することで、シミュレーション結果の偏りを減らすことができることを意味している。また、生態系の潜熱フラックスと FAPAR の季節変化は、 $V_{\max}$ (Rubisco のカルボキシレーションの最大活性)、フェノロジー、葉内・外 CO_2 濃度割合、土壌水分保持力に関するパラメータが大きく制御していることが分かり、パラメータの最適化により、当該パラメータの不確実性を大きく減らすことができた。

3.2.4 その他のデータ同化手法についての取り組み

Hazarika *et al.* (2005)では、モデル予報値をある一定の割合で、観測値に近づけるナッジング(nudging)が試みられている。その研究では、陸域生態系炭素循環モデル Sim-CYCLE (Ito and Oikawa, 2002)で推定された LAI を、衛星 MODIS から得られた LAI

図9 CCDAS によって最適化された (a) アラスカ・Barrow と (b) コロラド・Niwot Ridge の CO_2 濃度の平均季節変化 (Rayner *et al.* 2005 の Fig. 5 と 6 に加筆)。CCDAS (実線), 観測値 ($\diamond$)。点線は、全期間にわたって平均された観測の不確実性 (標準偏差) を示す。

データに対して近づけ、次ステップの計算の初期値として利用し、グローバルな NPP の推定を行った。

根の深度を最適化する試みが、いくつかなされている。Kleidon (2004) は、黄金分割法 (Golden Section Search Algorithm; Press *et al.*, 1992) という方法を利用して、光合成有効放射吸収量 (APAR) に対してのコスト関数を最小にする、根の深度のグローバルデータセットを作成した。彼らは、有効土壌水分を 0.005 m から 1.0 m まで徐々に変化させ、衛星から得られた APAR とモデルから得られた APAR の間のコスト関数を最小にする有効土壌水分量から、最適な根の深度を求めた。また、Ichii *et al.* (2009) は、同じく黄金分割法で、カリフォルニア州の蒸発散の分布についての推定精度の向上を試みた。彼らは、根の深度を 0.1 m から 10.0 m まで徐々に変化させ、衛星データベースの値と、TOPS という陸域水・炭素循環モデル (White and Nemani, 2004; Nemani *et al.*, 2009) によって得られた値の RMSE が、最小になる根の深度を探索した。この黄金分割法は、一次元のパラメータで、極小、もしくは極大がひとつだけある単峰性の目的関数に対しては、実装が簡便で有効であると考えられる。

4. 陸域生物圏モデルにおけるデータ同化の課題と将来性

4.1 現状と課題

近年では、陸域生態系の水・炭素循環研究においては、MCMC を利用した最尤推定法が、もっともよく用いられているデータ同化法である。それは、計算コードが比較的シンプルであり実装が容易だからであるが、一方で、十分に大きな尤度関数を得るた

めには、わずかに変化させたパラメータによるモデル計算を、繰り返し数千から数十万回行う必要があり、計算の負荷が高く、地点ごとに独立した計算のみに適用されている。

EnKF については、既存の陸域生物圏モデル本体の計算コードに対して、追加しなければいけない計算コードは比較的シンプルであり、実装は容易である。そのため、EnKF を利用した研究は増えているが、現在のところは、1 点ごとのプロット (地点) でのシミュレーションが行われるにとどまっている。この手法では、計算負荷が同化対象とするデータの点数が増えるごとに飛躍的に上昇するため、最適化するパラメータの数をある程度絞ったり、対象データの平均時間を長く取るなどして、負荷を減らす工夫がされている場合が多い。また、モデルの構造も簡略化しているケースも多い。それでも最適化がうまくいかないときは、データ同化の対象期間を短くする、パラメータの事前不確実性の値を小さくするなどの調整に迫られる場合もある。今後、経験の積み重ねによって、効率的に EnKF を利用するスキームが構築されれば、さらに面的な広がりを持つデータ同化についても、利用が進むと考えられる。

アジョイント法 (4 次元変分法) は、複数地点の大気 CO_2 濃度データなどの空間分布を取り扱わなければいけないデータ同化において、現時点で最も有効な方法である。また、同方法は、時間軸の正方向にも逆方向にもパラメータにおける不確実性を推定できるという利点がある。しかし、一方で、陸域生物圏モデル本体のコードを変える必要があったり、新たに追加する計算スキームが多いなど、実装は容易ではない。また、一般に陸域生物圏モデルは、条件

文や積算した物理量の多用などによって、非線形な構造になっているため、わずかなパラメータの変化でも、コスト関数がローカルな極小を複数持ったり、急激なコスト関数の変化が起こったりすることがよくある。そのため、最適なパラメータセットに到達していないのに、コスト関数のパラメータに対する傾きがゼロで、極小に達したと判断されることや、逆に非常に大きな傾きを取る場合に、計算が途中で止まってしまうという問題が頻発する。これについては、パラメータに与える微小変化の間隔を調節する必要がある。このようにコスト関数の最小化プロセスにおいて、アジョイント計算の導入は容易ではないため、フランスの LSCE (フランス気候環境科学研究所) では、アジョイントを利用しない変分法を試している。Santaren *et al.* (2007) は、シンプルで有限微分を用いて、フランスの Bray にあるフラックスタワーの観測データに対して陸域生態系モデル ORCHIDEE を同化した。彼らは、コスト関数を減少させるパラメータ場の探索方向の決定に、準ニュートン法である limited memory BFGS 法 (Broyden-Fletcher-Goldfarb-Shanno method; Bryd *et al.*, 1995) を利用した。まず 12 個のパラメータそれぞれの微小変化に対するコスト関数の変化が計算された (すなわち、1 個のパラメータだけ変化させた計算を、各パラメータに対して (計 12 回) 行う)。そして、コスト関数を減少させる方向の値を次パラメータに採用した。この作業を繰り返し、コスト関数の段階的な最小化を行った。パラメータの微小変化区間は固定された値 (10^{-3}) を採用した。そして、30 分平均の CO_2 ・顕熱・潜熱の各フラックスと、正味放射量に対して、12 種類のパラメータの最適化を行った。コスト関数は、いずれも 100 回程度の繰返し後に極小に達した。その結果、夏季の 3 週間のデータに対する同化では、日変化の再現性は高まったが、季節変化はそれほどではなかった。次に、1 年間のデータに対する同化は、季節変化よりも日変化の再現性をさらに高めた。このように、季節変化よりも日変化のほうが、再現性が高くなるのは、フラックスの日変化に関係の強いカルボキシン化速度などの光合成に関するパラメータは、 CO_2 と潜熱フラックスによって良く制御されたが、季節変化に関係の強い成長・維持・土壌呼吸などのパラメータは適切に制御されなかったことを示唆している。さらに、Kuppel *et al.* (2012) は、ヨーロッパ・日本・米国の 12 のフラックスサイトにおける CO_2 と潜熱フラックスデータに対して、ORCHIDEE の 21 のパラメータを最適化した。その際、フェノロジーなどの条件文等で閾値が設定され値が突然変化するパラメータには有限

微分を、ほかのパラメータには正接接線を計算する自動微分アルゴリズム (TAF; Giering and Kaminski, 2002) を利用して、コスト関数の傾きを求めた。その結果、日データに対する RMSE は、同化前よりサイト間の平均で 22% も減少し、同化後の LAI は、衛星 NDVI との間に位相の一致を見せた。また、Verbeeck *et al.* (2011) は、同様のアルゴリズムで、ブラジル Tapajós の熱帯林の NEE と潜熱フラックスデータに対して、ORCHIDEE を同化した。彼らは、最適化の対象とするパラメータについて、5 個から 162 個までの、さまざま組み合わせを試した。その結果、よりよく観測データを再現するには、10 m にもなる深い土壌深度と、直線的な根系密度の鉛直分布が必要であり、それらによって乾季にもかかわらず、蒸散と光合成は、高く維持されていることを示した。なお、常緑の熱帯林ではフェノロジーの変化がスムーズであるため、Kuppel *et al.* (2012) とは異なり、すべてのパラメータに対して、TAF を利用してコスト関数の傾きを計算している。

データ同化手法以外にも重要な課題がある。それは、まず、「観測データの精度の向上」である。陸域生態系について同化に利用できるデータに、衛星データ (LAI, FAPAR) があるが、熱帯の曇天率が高い地域では、生産力の過小評価をしている可能性がある (Kato *et al.*, 2013; Knorr *et al.*, 2010)。また、大気-陸域生態系間の物質循環のデータを提供する過相関法では、熱収支インバランスに代表されるような観測における系統誤差をはじめ、データの品質管理、欠測データの補完の是非など、多くの課題を抱える。これらを解決し観測データにおける不確実性を減らすことは、非常に重要である。次に、重要なのが、データ同化のベースとなる「陸域生物圏モデル自体の精度の向上」である。つまり、パラメータの最適化を施す以前のモデルについて、(時にはモデル簡略化も含む) 積極的な構造の改良、パラメータ初期値・不確実性の正しい設定を行うことは、データ同化をスムーズに進めることにつながるだけでなく、対象とする生態系のしくみを正しく理解することにもつながる。また、同化後のモデル解析値を、次ステップの初期値に利用するような逐次推定値の修正を行うようなカルマンフィルターなどの方法では、物質収支が保存されない場合もあり、同化の対象になっている変数がそれを許しているかどうかを吟味する必要がある。

4.2 データ同化手法の比較

Trudinger *et al.* (2007) は、シンプルで一次元の陸域生物圏モデルを利用し、仮想的に作成された観測データの同化について手法間の比較を行っている。

比較には、MCMC 法、EnKF、アジョイント法のほかに、非線形曲線近似を扱うレーベンバーグ・マルカート法と、遺伝的アルゴリズムが採用され、10 の機関が参加し、21 のモデルシミュレーションが行われた。彼らは、仮想的に、植物、リター、土壤に見立てた 3 つのコンパートメントを持つシンプルなもの一つのモデルを作成し、各コンパートメントのバイオマスをコントロールする NPP、リターと土壤の各分解速度に関する 4 つのパラメータを設定した。そして、異なる間引き処理を施した観測データに見立てたものに対して、パラメータの最適化を行った。その結果、最適化後のパラメータの値については、データ同化手法間の違いは小さく、手法そのものの影響は小さいことがわかった。むしろ、コスト関数の設定の仕方や、観測データの間引き処理やノイズの加え方の違いによる影響が、大きいことを示した。

陸域生物圏モデルのデータ同化は、まだ歴史が浅く研究が深まっていないため、各手法の実践的な特徴を捉えるには、今しばらく時間がかかりそうである。究極的に言えば、Trundinger *et al.* (2007) のように、最適化後のパラメータは、手法に関係なく同じ値をとることが期待される。特に、EnKF 法とアジョイント法は、よく比較されるが、非線形性を持つような複雑系問題については、問題設定、誤差、拘束条件、近似の方法によって、両者の精度は大きく変化するために単純には優劣を決めることができない。この点について、歴史の長い気候・海洋学の分野では、両者の精度に差はないと考えるのが妥当であるとされている(淡路ら, 2009)。そのため、データ同化手法の選択は、実装にかかる労力(コード作成の難易度、計算資源など)に大きく影響されていくだろうと考える。今後、研究の発展とともに、陸域生物圏モデル分野においても、同化手法の比較がたびたび企画されることを期待する。

4.3 将来性

データ同化手法は、単に最適なパラメータを得て、モデルの推定精度を向上させるだけではない。すでに、気候分野においてなされているように、理論によく馴染まれた尤もらしい観測データである「客観解析データ」と呼ばれる、データセットの作成をも可能とすると思われる。すなわち、それらを解析することにより、たとえば、20 世紀後半の炭素動態と気候変化・大気 CO₂ 濃度上昇の関係を、より確からしく明らかにできるかもしれないということだ。さらに、長期の観測データを入手できれば、生態系の生理・生態プロセスの応答特性の長期時系列変化を調べることができる。たとえば、Xu *et al.* (2006) では、FACE (開放系大気 CO₂ 増加実験) の行われてい

る Duke Forest において、CENTURY モデル (Parton *et al.*, 1987) を、1996 年から 2000 年までの葉バイオマス、木質バイオマスの成長量、リターフォール、リターや鉱質土壌プール、土壌呼吸の各データに対して同化した。その結果、最適化された炭素プール間の移動速度に関するパラメータについて、高 CO₂ 濃度区と外気 (Ambient) 区との間に差が出ており、CO₂ 濃度の上昇が生態系の物質循環速度を著しく変化させることを明確に示した。このように、高温・高 CO₂ 濃度に対する光合成の馴化などの、「将来の地球の物質循環を考える上で重要であるにもかかわらず決定的な証拠が示されていない課題」に対しても答えることが可能であると思われる。

4.4 観測とモデルをつなぐもの

データ同化は、観測とモデルをつなぐ橋のようなものである(三好, 2010)。これまでは、その複雑な統計数理的な理論を勉強しなければならず、物理や数学のバックグラウンドのない生物系モデル研究者にとっては、導入がためらわれた。しかし、すでに気候・海洋学で実証されているように、データ同化によって効率よく推定精度を上げることができれば、単に理論を試しているだけの空想上のシミュレーション結果というレベルを超えて、より観測に近い生理・生態学的な現象の把握に大きく役に立つであろう。

データ同化はモデル研究者だけでなく、陸域生態系の観測研究者にとっても恩恵をもたらす。まず考えられるのは、時間的にまばらな観測データの間を埋めることである。それ以外にもたとえば、グローバルなモデルのパラメータを最適化していく過程で、情報が不足している生態系タイプや気候帯を把握することができれば、あらたな観測点を置くときの参考にもなると思われる。いずれにせよ、観測網の発展や精度向上にかけられる労力に限界がある以上、観測者側もモデルを多かれ少なかれ利用し、時空間的に連続的なデータを提供していくことがすでに求められていると思われ、データ同化はその一助になり得るであろう。渦相関フラックス観測の世界的コミュニティである FLUXNET では、Model-Data Fusion (MDF) が大きなテーマになっているが、渦相関フラックスなどの連続・多地点のデータが蓄積しつつある今、観測系研究者とモデル研究者が協力し、データ同化を利用した研究を進めることが期待されている (Williams *et al.*, 2009)。

本総説でとりあげた、陸域生物圏モデルにおけるデータ同化は、いまだ発展途上の分野であり、どのグループにおいても、手探り状態で研究をすすめている段階である。しかし一方で、欧州では、ドイ

ツ・マックスプランク研究所(MPI)が、気候(GCM)-生態系結合モデル JSBACH へのデータ同化手法の導入を始めており、将来の地球環境を予測するための地球システムモデルにおいての本格的な運用が望まれている。このように、地球環境研究分野において、この陸域生物圏モデルにおけるデータ同化手法の重要性が高まっていることは間違いない。

用 語

解析値：データ同化後の推定値

観測演算子(観測行列)：数値モデルの予報変数を観測されるデータの物理量に変換する演算子(行列)のこと。観測量とモデルの予報値が同じ物理量を示す場合、予報変数を格子上から、観測点への空間内挿することを示す場合も多い。

観測誤差：観測値と、真値との差。観測値には、観測時に生じる系統誤差や、ノイズなどを含む測定誤差を含む。観測値を y 、モデルの状態変数(入力変数)を x とすると、 $y = H(x)$ とあらわされる。ここで、 $H(\cdot)$ は観測演算子。観測誤差 Δy は、 $\Delta y = y - H(x^t)$ と表される。 o と t は、観測値と真値を示す。

観測値：観測データ。データ同化で取り込まれる値。観測値には、測定誤差が含まれるため真値とは異なる。

第一推定値：数値シミュレーションモデルによるデータ同化前の推定値。モデルによる予報値が与えられる。

背景誤差：第一推定値の誤差

背景誤差共分散：第一推定値の誤差期待値のばらつきを示す。行列で表される場合、対角成分は誤差分散になり、非対角成分は格子点間もしくは観測点間の誤差共分散になる。最適内挿法や 3 次元変分法では、背景誤差共分散を一定として扱う。4 次元変分法やカルマンフィルターなどの高度な手法では、背景誤差共分散の時間発展を考慮するため、解析の精度が高くなる。

変分法：解析値を導出する過程で、評価関数(コスト関数)の勾配を利用し、その最小への到達を図るのだが、その勾配の導出の際に、物理学でよく利用される変分法を利用するので、変分法と呼ばれる。通常の微分が数の関数を扱うのに対して、変分法では関数の関数(汎関数)を取り扱う。空間 3 次元方向のベクトルを扱うものを 3 次元変分法、それに時間 1 次元を足したベクトルを扱うものを、

4 次元変分法と呼ぶ。

摘 要

コンピュータシミュレーションにおいて、モデルと観測との差を、客観的に埋めていくための統計的方法に、データ同化がある。これは、数値モデルに観測データを挿入(同化)することで、より高精度な状態量の推定、データ補間、パラメータ・初期値の最適化を行う手法である。近年、これらデータ同化手法は、陸域生物圏モデルにも用いられ、より尤もらしいシミュレーションを行う努力がなされている。

マルコフ連鎖モンテカルロ法を利用したデータ同化は、計算コードが比較的シンプルであることが特徴であり、これまでよく利用されている。しかし、一方で、多数の繰り返し計算を行うために、計算の負荷が高く、地点ごとに独立した計算のみに適用されている。アンサンブル・カルマンフィルターは、実装が容易であり、同化対象のデータ点数やパラメータ数を絞ることで、計算負荷を小さくすることが可能なために、利用が増えてきている。また、アジョイント法(4 次元変分法)は、時間軸の正方向にも逆方向にも各パラメータに対する不確実性を推定でき、計算負荷もそれほど大きくはないという利点がある。しかし、一方で、陸域生物圏モデルの非線形性によって起こる問題を避けるために、最適化プロセスの実装について、工夫が必要である。

データ同化手法は、モデルの推定精度を向上させるだけではなく、理論によく馴らされた尤もらしい観測データである客観解析データを作成することができる。また、最適パラメータの時間変化から環境変化に対する生理・生態的応答の長期間の変化を調べることも将来的に可能である。

陸域生物圏モデルにおけるデータ同化手法の研究は、始まったばかりであり、乗り越えなければならない課題は多いが、生態学・気候学分野をはじめとする、地球環境研究において、その重要性が高まっていることは間違いない。

謝 辞

本総説執筆に当たり、W. Knorr 博士、M. Scholze 博士から大いなる助言をいただいた。また、海洋研究開発機構・地球環境変動領域の美山透博士には、本総説にたいして多大なご助言をいただいた。国立環境研究所・地球環境研究センターの伊藤昭彦博士と小田知宏博士(現 NCAR)、海洋研究開発機構・地球環境変動領域の立入郁博士にも、ご助言を頂いた。本研究を行うにあたり、日本学術振興会から海外特

別研究員として援助を受けた。心より感謝する。

引用文献

- 淡路敏之, 蒲地政文, 池田元美, 石川洋一(編著), 2009: データ同化—観測・実験とモデルを融合するイノベーション. 京都大学学術出版会, 284 pp.
- Barbu, A. L., Calvet, J. C., Mahfouf, J. F., Albergel, C., Lafont, S., 2011: Assimilation of Soil Wetness Index and Leaf Area Index into the ISBA-A-gs land surface model: grassland case study. *Biogeosci.* **8**, 1971-1986, doi:10.5194/bg-8-1971-2011.
- Bergthorsson, P. and Döös B. R., 1955: Numerical weather map analysis. *Tellus*, **7**, 329-340.
- Braswell, B. H., Sacks, W. J., Linder, E., Schimel, D. S., 2005: Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. *Global Change Biol.*, **11**, 335-355.
- Byrd, R. H., Lu, P., Nocedal, J., 1995: A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Stat. Comput.*, **16**, 1190-1208.
- Cao, M. and Woodward, F. I., 1998: Dynamic responses of terrestrial ecosystem carbon cycling to global climate change. *Nature*, **393**, 249-252.
- Chen, M., Liu, S., Tieszen, L. L., Hollinger, D. Y., 2008: An improved state-parameter analysis of ecosystem models using data assimilation. *Ecol. Model.*, **219** (3-4), 317-326.
- Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A., Totterdell, I. J., 2000: Acceleration of global warming due to carbon cycle feedbacks in a coupled climate model. *Nature*, **408**, 184-187.
- Dickinson, R. E., Jaeger, J., Washington, W. M., Wolski, R., 1981: Boundary subroutine for the NCAR global climate model. National Center for Atmospheric Research Tech. Note TN-173+1A, 75 pp.
- Dickinson, R. E., Shaikh, M., Bryant, R., Graumlich, L., 1998: Interactive canopies for a climate model. *J. Climate*, **11**, 2823-2836.
- Eliassen, A., 1954: Provisional report on calculation of spatial covariance and autocorrelation of the pressure field, Report 5. Institute of Weather and Climate Res, Academy of Science, Oslo, p 11.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10143-10162.
- Gandin, L. S., 1963: Objective Analysis of Meteorological Fields. Hydromet Press, 242 pp.
- Ghil, M., Cohn, S. E., Tavantzis, J., Bube, K., Isaacson, E., 1981: Application of estimation theory to numerical weather prediction. In *Dynamic Meteorology: Data Assimilation Methods* ed. by Bengtsson, L., Ghil, M., Källén, E., Springer-Verlag, Berlin-Heidelberg-New York, pp. 139-224.
- Ghil, M., Ide, K., Bennett, A., Courtier, P., Kimoto, M., Nagata, M., Saiki, M., Sato, N. (eds), 1997: Data Assimilation in Meteorology and Oceanography: Theory and Practice. *J. Meteorol. Soc. Japan*, **75**, 111-196.
- Giering, R., Kaminski, T., 2002: Recomputations in reverse mode AD. In: *Automatic Differentiation of Algorithms: From Simulation to Optimization* ed. by Corliss, G., Griewank, A., Fauré, C., Hascoet, L., Naumann, W., Springer Verlag, Heidelberg, chapter **33**, 283-291.
- Gilchrist, B. and Cressman, G. P., 1954: An experiment in objective analysis. *Tellus*, **6**, 309-318.
- Gottschalck, J., Meng, J., Rodell, M., Houser, P., 2005: Analysis of Multiple Precipitation Products and Preliminary Assessment of Their Impact on Global Land Data Assimilation System Land Surface States. *J. Hydrometeorol.* **6**, 573-598.
- Hazarika, M., Yasuoka, Y., Ito, A., Dye, D. G., 2005: Estimation of net primary productivity by integrating remote sensing with an ecosystem model. *Remote Sens. Environ.* **94**, 298-310.
- Heimann, M., 1995: The TM2 tracer model, model description and user manual. Technical Report No. 10, ISSN 0940-9327. Deutsches Klimarechenzentrum, Hamburg, 47 pp.
- 細川敬祐(編), 2009: データ同化の考え方とその方法(講演: 中野慎也). 平成 21 年度 MTI 研究会サイエンスセッション: MTI 研究領域における計算科学, MTI ハンドブック, http://www2.nict.go.jp/aeri/swe/MTI/MTI_symposium/mti-handbook/top/h21/, p. 9.
- Ichii, K., Wang, W., Hashimoto, H., Yang, F., Votava, P., Michaelis, A. R., Nemani, R. R., 2009: Refinement of rooting depths using satellite-based evapotranspiration seasonality for ecosystem modeling in California. *Agric. For. Meteorol.* **149**, 1907-1918.
- Iizumi, T., Yokozawa, M., Nishimori, M., 2009: Parameter estimation and uncertainty analysis of a large-scale crop model for paddy rice: Application of a Bayesian approach. *Agric. For. Meteorol.* **149**, 333-348.
- Iizumi, T., Yokozawa, M., Nishimori, M., 2011: Probabilistic evaluation of climate change impacts on paddy rice productivity in Japan. *Clim. Chang.* **107**, 391-415. doi: 10.1007/s10584-010-9990-7.
- Ishikawa, Y. and Koizumi, K., 2002: One month cycle experiments of the JMA mesoscale 4-dimensional variational data assimilation (4D-Var) system. *CAS/JSC WGNE Res. Activ. Atmos. Ocean. Model.*, **32**, 0126-

- 0127
- Ito, A. and Oikawa, T., 2002: A simulation model of the carbon cycle in land ecosystems (Sim-CYCLE): A description based on dry-matter production theory and plot-scale validation. *Ecol. Model.*, **151**, 147-179.
- Kalman, R. E., 1960: A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng. (Trans. ASME Ser. D)* **82**, 35-45.
- Kalman, R. E. and Bucy, R. S., 1961: New Results in Linear Filtering and Prediction Theory. *J. Basic Eng. (Trans. ASME Ser. D)* **83**, 95-108.
- Kalnay, E., Li, H., Miyoshi, T., Yang, S. C., Ballabrera-Poy, J., 2007a: 4D-Var or Ensemble Kalman Filter? *Tellus A*, **59**, 758-773.
- Kalnay, E., Li, H., Miyoshi, T., Yang, S. C., Ballabrera-Poy, J., 2007b: Response to the Discussion on "4D-Var or EnKF?" by Nils Gustaffson. *Tellus A*, **59**, 778-780.
- Kato, T., Scholze, M., Knorr, W., Veenendaal, E., Kaminski, T., Kattge, J., Gobron, N., 2013: Simultaneous *assimilation* of satellite and eddy covariance data for *improving* terrestrial water and carbon simulations at a semi-arid woodland site in Botswana. *Biogeosci.* **10**(2), 789-802.
- Kleidon, A., 2004: Global datasets of rooting zone depth inferred from inverse methods. *J. Clim.*, **17**(13), 2714-2722.
- Knorr, W., 2000: Annual and interannual CO₂ exchanges of the terrestrial biosphere: Processbased simulations and uncertainties. *Glob. Ecol. Biogeogr.*, **9**(3), 225-252.
- Knorr W., and Kattge, J., 2005: Inversion of terrestrial biosphere model parameter values against eddy covariance measurements using Monte Carlo sampling. *Glob. Chang. Biol.* **11**, 1333-1351.
- Knorr, W., Kaminski, T., Scholze, M., Gobron, N., Pinty, B., Giering, R., Mathieu, P.-P., 2010: Carbon Cycle Data Assimilation with a Generic Phenology Model. *J. Geophys. Res. Atmos.* **115**, G04017, doi:10.1029/2009JG001119.
- Kuppel, S., Peylin, P., Chevallier, F., Bacour, C., Maignan, F., and Richardson, A. D., 2012: Constraining a global ecosystem model with multi-site eddy-covariance data. *Biogeosciences*, **9**, 3757-3776, doi:10.5194/bg-9-3757-2012.
- Lasslop, G., Reichstein, M., Kattge, J., Papale, D., 2008: Influences of observation errors in eddy flux data on inverse model parameter estimation. *Biogeosci.* **5**, 1311-1324.
- Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP -a comparison with 4D-Var. *Quart. J. R. Meteorol. Soc.*, **129**, 3183-3203.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M. N., Teller, A. H., Teller, E., 1953: Equation of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1087-1092.
- 三好建正, 2005: アンサンブル・カルマンフィルタ—データ同化とアンサンブル予報の接点—. 天気, **52**(2), 3-14.
- 三好建正, 2010: データ同化への誘い. 天気, **57**, 208-211.
- Mo, X., Chen, J. M., Ju, W., Black, A. T., 2008: Optimization of ecosystem model parameters through assimilating eddy covariance flux data with an Ensemble Kalman Filter. *Ecol. Model.*, **217**, 157-173.
- Nemani, R. R., Hashimoto, H., Votava, P., Melton, F., Wang, W., Michaelis, A., Mutch, L., Milesi, C., Hiatt, S., White, M., 2009: Monitoring and forecasting ecosystem dynamics using the Terrestrial Observation and Prediction System (TOPS). *Remote Sens. Environ.*, **113**, 1497-1509.
- Ni, W., Li, X., Woodcock, C. E., Caetano, M. R., Strahler, A. H., 1999: An analytical hybrid GORT model for bi-directional reflectance over discontinuous plant canopies. *IEEE Transactions on Geosci. and Remote Sens.* **37**(2), 987-999.
- Panofsky, H. A., 1949: Objective weather map analysis. *J. Meteorol.* **6**, 386-392.
- Parish, D. F. and Cohn, S. E., 1985: A Kalman filter for a two-dimensional shallow water model: Formulation and preliminary experiments. Office Note 304, National Meteorological Center, Washington, DC., 64 pp.
- Parton, W. J., Schimel, D. S., Cole, C. V., Ojima, D. S., 1987: Analysis of factors controlling soil organic levels of grasslands in the Great Plains. *Soil Sci. Soc. Am. J.*, **51**, 1173-1179.
- Press, W. H., Flannery, B. P., Teukolsky, S. A., Vetterling, W. T., 1992: Golden section search in one dimension (2nd edition) In Numerical Recipes in C: The Art of Scientific Computing, Cambridge University Press, Cambridge, pp. 397-401.
- Quaife, T., Lewis, P., Kauwe, M., Williams, M., Law, B., Disney, M., Bowyer, P., 2008: Assimilating Canopy Reflectance data into an Ecosystem Model with an Ensemble Kalman Filter. *Remote Sens. Environ.* **112**(4), 1347-1364.
- Rabier, F., Jarvinen, H., Klinker, E., Mahfouf, J. F., Simmons, A., 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quart. J. R. Meteorol. Soc.*, **126**, 1143-1170.
- Rayner, P., Scholze, M., Knorr, W., Kaminski, T., Giering, R., Widmann, H., 2005: Two decades of terrestrial Car-

- bon fluxes from a Carbon Cycle Data Assimilation System (CCDAS). *Glob. Biogeochem. Cycles*, **19**, doi:10.1029/2004GB002254.
- Santaren, D., Peylin, P., Viovy, N., Ciais, P., 2007: Optimizing a process-based ecosystem model with eddy-covariance flux measurements: A pine forest in southern France. *Glob. Biogeochem. Cycles*, **21**, GB2013;doi:10.1029/2006GB002834.
- Scholze, M., 2003: Model studies on the response of the terrestrial carbon cycle on climate change and variability, Examensarbeit, Max-Planck-Institut für Meteorologie, Hamburg, Germany. 144 pp.
- Scholze, M., Kaminski, T., Rayner, P., Knorr, W., Giering, R., 2007: Propagating uncertainty through prognostic CCDAS simulations. *J. Geophys. Res.*, **112**, doi:10.1029/2007JD008642.
- Sellers, P. J., Mintz, Y., Sud, Y. C., Dalcher, A., 1986: A simple biosphere model (SiB) for use within general circulation models. *J. Atmos. Sci.*, **43**, 505–531.
- Sellers, P. J., Randall, D. A., Collatz, G. J., Berry, J. A., Field, C. B., Dazlich, D. A., Zhang, C., Collelo, G. D., Bounoua, L., 1996: A revised land surface parameterization (SiB2) for atmospheric GCMs. Part I: Model formulation. *J. Clim.*, **9**, 676–705.
- Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J., Levis, S., Lucht, W., Sykes, M., Thonicke, K., Venevski, S., 2003: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ Dynamic Vegetation Model. *Glob. Chang. Biol.*, **9**, 161–185.
- Tarantola, A., 1987: Inverse Problem Theory: Methods for Data Fitting and Parameter Estimation, Elsevier, Amsterdam 614 pp.
- Trudinger, C. M., Raupach, M. R., Rayner, P., Kattge, J., Liu, Q., Pak, B., Reichstein, M., Renzullo, L., Richardson, A., Stephen, H., Styles, J., Wang, Y. P., Briggs, P., Barrett, D., Nikolova, S., 2007: OptIC project: An intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models. *J. Geophys. Res.*, **112**, G02027, doi:10.1029/2006JG000367.
- 露木 義, 川畑拓矢(編), 2008: 気象学におけるデータ同化. 気象研究ノート第 217 号, 277 pp.
- Verbeeck, H., Peylin, P., Bacour, C., Bonal, D., Steppe, K., Ciais, P., 2011: Seasonal patterns of CO₂ fluxes in Amazon forests: Fusion of eddy covariance data and the ORCHIDEE model. *J. Geophys. Res.*, **116**, G02018, doi:10.1029/2010JG001544.
- Weiss, M., Baret, F., Myneni, R. B., Pragnere, A., Knyazikhin, Y., 2000: Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. *Agron.*, **20**, 3–22.
- White, M. A. and Nemani, R. R., 2004: Soil water forecasting in the continental United States: relative forcing by meteorology versus leaf area index and the effects of meteorological forecast errors. *Can. J. Remote Sens.*, **30**, 717–730.
- Williams, M., Schwarz, P. A., Law, B. E., Irvine, J., Kumpius, M. R., 2005: An improved analysis of forest carbon dynamics using data assimilation. *Glob. Chang. Biol.*, **11**, 89–105.
- Williams, M., Richardson, A. D., Reichstein, M., Stoy, P. C., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D. Y., Kattge, J., Leuning, R., Luo, Y., Tomelleri, E., Trudinger, C. M., Wang, Y. -P. 2009: Improving land surface models with FLUXNET data. *Biogeosci.*, **6**, 1341–1359, doi:10.5194/bg-6-1341-2009.
- Wu, X. W., Luo, Y. Q., Weng, E. S., White, L., Ma, Y., Zhou, X. H., 2009: Conditional inversion to estimate parameters from eddy-flux observations. *J. Plant Ecol.*, doi:10.1093/jpe/rtp005.
- Xu, T., White, L., Hui, D., Luo, Y., 2006: Probabilistic inversion of a terrestrial ecosystem model: Analysis of uncertainty in parameter estimation and model prediction. *Glob. Biogeochem. Cycles*, **20**, doi:10.1029/2005GB002468.
- Yuan, W., Liang, S., Liu, S., Weng, E., Luo, Y., Hollinger, D., Zhang, H., 2012. Improving model parameter estimation using coupling relationships between vegetation production and ecosystem respiration. *Ecol. Model.*, **240**, 29–40.
- Zaitchik, B. F., Rodell, M., Olivera F., 2010: Evaluation of the Global Land Data Assimilation System using global river discharge data and a source-to-sink routing scheme. *Water Resour. Res.*, **46**, W06507, doi:10.1029/2009WR007811.
- Ziehn, T., Kattge, J., Knorr, W., Scholze, M., 2011a: Improving the predictability of CO₂ assimilation rates at the global scale. *Geophys. Res. Lett.*, **38**, L10404.
- Ziehn, T., Scholze, M., Knorr, W., 2011b: Development of an ensemble-adjoint optimization approach to derive uncertainties in net carbon fluxes. *Geosci. Model Dev. Discuss.*, **4**, 1513–1531, doi:10.5194/gmdd-4-1513-2011.