

中央氣象局全球模式四維變分比對實驗

曾建翰 張忍成

中央氣象局氣象科技研究中心
中國文化大學大氣科學系

(中華民國九十五年六月五日收稿；中華民國九十五年十一月十日定稿)

摘 要

本研究的目的是在嘗試完成一個四維變分過程，暫時不考慮加入真實的觀測資料，選定一初始場的模式預報為假想觀測軌跡，作為控制組；而另一相較於此初始場 12~24 小時內不同初始場的模式預報為實驗組，取兩者差值定義目標函數，利用四維變分過程求算最佳的初始場，使其預報結果朝假想觀測軌跡靠近。使用的工具為氣象局全球波譜模式的伴隨模式，以其求算目標函數的梯度，配合極小化程序（minimization），L-BFGS 方法，進行四維變分實驗。在實驗中，嘗試加入不規則的網格分佈成為假想觀測場，也嘗試加入集中某一地區：如台灣東亞區的觀測資料進行變分工作。整個計算過程，可供氣象局未來四維資料同化工作所需。

從三個個案 3 個比對實驗結果來說，所得之最佳初始場朝假想的觀測初始場逼近，其差距明顯比未經四維變分之原初始場與假想觀測初始場的差距為小。其中，以全部分析網格所做的理想化比對實驗結果必然最佳，可作為理論參考值。對所使用的全球模式預報結果而言，加入全球分佈的不規則觀測場比集中台灣東亞地區的觀測資料的預報結果為佳。而全球分佈的不規則觀測場加上背景差誤差分析所得的新初始場和預報軌跡相較於未經變分分析的預報軌跡而言，取預報 120 小時結果比較，約使平均誤差下降 12%~29%。以變分極小化作業效率的角度來看，疊代運算 30 次可使目標函數下降 1 個數量級，而以其 8 天期間的預報結果來看，亦與疊代運算 100 次相仿，所以疊代運算 30 次可作為實際作業時採行的參考值。

關鍵詞：伴隨模式，極小化程序，L-BFGS

一、前言

從 50 年代末至 60 年代初，非傳統觀測資料如雷達、衛星等開始被大氣作業單位及學界使用。非傳統觀測資料的特性是，在時間上的分佈並不侷限在某些特定時刻，所得的資訊並不限於傳統的溫度、氣壓和風速；再者，不論是非傳統資料或是傳統資料，空間位置的分佈常不具規則性，是故如何在數值預報模式中將這些時間、空間都不連續也不規則的資料同化成模式可以使用，成為目前大氣作業中心和學界的研究重點之一。同化資料的方法由早期的內插法、牛頓緩和法（nudging），到現今的變分法（variational method）或 Kalman 濾波法，而普遍的認知是：變分法假設較少，較優於內插法和緩和法。

大氣數值預報模式引進資料同化的目的無非就是取得較佳的初始場，期望能得到較佳的預報結果。由於大氣的問題具有四維問題的特性（三度空間+時間），當所求的問題被定義成目標函數後，不論是新資料的加入以求算新初始場，或是經不穩定擾動分析求算一組擾動系集的過程，學界稱為是一種四維變分過程（4 dimensional variational assimilation; 4D-Var）。技術上，本文沿襲大氣學界目前採用的伴隨模式求算目標函數的梯度，配合極小化程序（minimization）進行四維變分實驗。

研究的目的是在嘗試完成一個四維變分過程，暫時不考慮真實觀測資料的加入，而是以人為化的假想狀態實驗出發，將焦點放在以模式預報的架構下求算最佳的初始場，也就是說，先以一組模式初始場製造出預報軌跡，當成是真實大氣演變軌跡，而選取鄰近時間不同的初始場預報軌跡

與假想真實大氣作四維變分分析，這個過程學界稱為比對實驗（twin experiment）。在此，本文設計了 3 個實驗：1. 針對所有分析網格所做的理想化實驗；2. 模擬全球觀測場分佈的實驗；3. 模擬台灣東亞地區附近有密集觀測的實驗。目的是：1. 第 1 個理想實驗應該有很好的結果；2. 模擬全球觀測場分佈的第 2 個實驗，將控制組的資料在空間中以不規則的方式取用並放入目標函數中，這個結果和第 1 個實驗差別多少，背景場誤差項加入後改進多少；3. 第 3 個實驗是為模擬當台灣東亞地區有密集觀測時，對全球模式有什麼影響，和第 1 和第 2 個實驗差別多少。最後，整個計算目標函數極值的過程，可在未來供中央氣象局真正的四維資料同化工作所需。

本文共分四節，第二節介紹所用的資料及研究方法，簡介伴隨模式、目標函數和所用極小化程序的過程。第三節中詳細解釋 3 個實驗的作法，對三個個案各自進行比對實驗，分析四維變分後的結果。第四節中綜合本研究的結果，並討論未來工作可能的實行步驟及可能遭遇的困難。

二、資料來源與研究方法

為簡化問題起見，採用 ECMWF 的全球網格點資料，研究的工具為中央氣象局簡化版全球波譜模式和四維變分求算目標函數極小化的程序。研究方法採行學界行之有年的比對實驗，其精神在於已知答案為何的情形下，測試四維變分的效果有無達到最低標準，實驗的作法將留到第三節中討論。

（一）ECMWF / TOGA BAISC 資料

所採用的資料為 ECMWF 的全球網格點資

料，所用的變數包括水平風場、重力位場和溫度場。資料水平解析度為 $2.5^\circ \times 2.5^\circ$ ，垂直方向有 14 層。分別是 1000、850、700、500、400、300、250、200、150、100、70、50、30、10 hPa。所用資料的時間為 1989 年 1 月 27 日 12UTC、1999 年 12 月 19 日 12UTC 和 1999 年 12 月 26 日 12UTC，而選用的理由是氣象局模式對這三個時間的 24 至 96 小時的展期預報失效，推究其原因可能是環流由高緯流急速轉成低緯流型態有關，雖然本文進行人為化的實驗，但是已經從一些前置工作中發現這三個個案基本場有助於擾動快速生成，所以選用差異不大的初始場將有明顯差異的預報軌跡，對於四維變分實驗是一個很好的出發點。

(二) 模式簡介

氣象局現行作業模式的解析度為水平三角截斷 180 個波，垂直分層 30 層 (T180L30)。由於四維變分過程中採用伴隨模式計算目標函數的極值，而伴隨模式的研發及計算過程相當費時，所以於現階段僅用低解析度 (T42L18)、簡化物理過程 (含簡化邊界層、Kuo 氏積雲參數化) 的全球模式做實驗 (為因應氣象局未來變分研究，現今測試模式已採用 T42L30)。切線性模式和伴隨模式的功用請見下一節的簡介，而推導和模式的實作過程請見曾 (1997) 和 Zou et al. (1998)。

(三) 研究方法

本節簡短說明伴隨模式和四維變分的理論架構，首先，如果我們將非線性的模式用矩陣代數符號寫為，

$$\begin{cases} \frac{\partial \mathbf{x}}{\partial t} = \mathbf{n}(\mathbf{x}(t)), \\ \mathbf{x}(t_0) = \mathbf{x}_0 \text{ 表初始值} \end{cases} \quad (1)$$

其中 $\mathbf{n}$ 是非線性算子， $\mathbf{x}$ 為模式變數所組成的行向量，為 $(u, v, \theta, q, Ps)^T$ ，上標 T 表矩陣轉置，其內容含意為 (東西向風，南北向風，虛位溫，比濕，地面氣壓)^T，又 t 為時間， t_0 為初始時間。由(1)式可推導出

$$\mathbf{x}(t) = \int_{t_0}^t \mathbf{n}(\mathbf{x}(t')) dt' + \mathbf{x}(t_0), \quad (2)$$

這當中暗示：某個時間的 $\mathbf{x}(t)$ 可透過一非線性算子 $\mathbf{N}$ 和初始值 $\mathbf{x}(t_0)$ 求得，簡寫成

$$\mathbf{x}(t) = \mathbf{N}(\mathbf{x}(t_0))$$

接下來，本文所求的目標函數可定義成

$$\begin{aligned} \mathcal{J} = & \frac{1}{2} \sum_{t=t_0}^{t_f} (\mathbf{x}(t) - \bar{\mathbf{x}}(t))^T \mathbf{W} (\mathbf{x}(t) - \bar{\mathbf{x}}(t)) \\ & + \frac{1}{2} (\mathbf{x}(t_0) - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x}(t_0) - \mathbf{x}_b) \end{aligned} \quad (3)$$

其中 $\bar{\mathbf{x}}(t)$ 表示另一種相對狀態的變數，可以是觀測場、或是大氣基本場等。 $\mathbf{W}$ 是一個權重函數，用來調整 $(\mathbf{x}(t) - \bar{\mathbf{x}}(t))$ 中變數彼此間的權重與相關性，最簡單的作用可想成是使不同因次的物理量變數調整成一個近似無因次量

(non-dimensional) 的大小。而 $\mathbf{x}_b$ 稱為背景場，是一個給定的常數場， $\mathbf{B}^{-1}$ 是背景場誤差協方差矩陣，其作用類似於 $\mathbf{W}$ 。(3)式中， $\mathcal{J}$ 由 $\mathbf{x}(t)$ 和 $\bar{\mathbf{x}}(t)$ 在時間 t_0 和 t_f 之間的差值組成，是一個四維空間 (三度空間+時間) 所組成的函數，所以求算 $\mathcal{J}$ 的極值過程就稱為四維變分。

由(2)式可知，每一個時刻的 $\mathbf{x}(t)$ 都可以用 $\mathbf{x}(t_0)$ 來表示，而 $\bar{\mathbf{x}}(t)$ 可視作已知的常數，於是(3)式可化簡成全是由 $\mathbf{x}(t_0)$ 所組成的函數，此時問題變成，當 $\mathbf{x}(t_0)$ 為多少時，能得到 $\mathcal{J}$ 的極小值。在利用數值方法求算極小值時，首先必須知道 $\mathcal{J}$ 對 $\mathbf{x}(t_0)$ 的梯度值，原則上沿負梯度方向搜

尋可以找出一個較小的 $\mathcal{J}$ 值，此過程可一再重複疊代運算得到可以接受的最小 $\mathcal{J}$ 值。所以，四維變分過程可簡單說成是：找出一個 $\mathbf{x}(t_0)$ ，可使 $\mathcal{J}$ 值最小。

常見數值疊代運算求極值的方法有共軛梯度法 (conjugate gradient method)、牛頓法 (Newton method)、準牛頓法 (quasi-Newton method)，利用目標函數的梯度 (共軛梯度法、準牛頓法) 或是搭配目標函數的 Hessian matrix (牛頓法) 經數值疊代運算找出極值，詳細的說明可見 Gill et al.

(1981) 和曾 (2006) 書中的討論，在下一節中將簡單介紹本文所用的一種準牛頓法 (L-BFGS method)。所以在此處的推導過程中，下一步工作就是如何決定 $\frac{\partial \mathcal{J}}{\partial \mathbf{x}(t_0)}$ 這個梯度值以供數值疊代法求算極值所須。

先，將(3)式對 $\mathbf{x}(t_0)$ 變分可得

$$\delta \mathcal{J} = \left[\sum_{t=t_0}^{t_f} \mathbf{W}(\mathbf{x}(t) - \bar{\mathbf{x}}(t)) \right]^T \delta \mathbf{x}(t) + [\mathbf{B}^{-1}(\mathbf{x}(t_0) - \mathbf{x}_b)]^T \delta \mathbf{x}(t_0) \quad (4)$$

但(4)式中 $\left[\sum_{t=t_0}^{t_f} \mathbf{W}(\mathbf{x}(t) - \bar{\mathbf{x}}(t)) \right]^T \delta \mathbf{x}(t)$ 的 $\delta \mathbf{x}(t)$ 與 $\delta \mathbf{x}(t_0)$ 關係尚未定義，無法決定 $\frac{\partial \mathcal{J}}{\partial \mathbf{x}(t_0)}$ 為何，所以令

$$\mathbf{x} = \hat{\mathbf{x}} + \delta \mathbf{x}$$

其中 $\delta \mathbf{x}$ 表一個相較於 $\hat{\mathbf{x}}$ 很小的擾動量，將(1)式中的 $\mathbf{n}(\mathbf{x}(t))$ 對 Taylor 級數展開，

$$\mathbf{n}(\mathbf{x}) = \mathbf{n}(\hat{\mathbf{x}}) + \frac{\partial \mathbf{n}(\hat{\mathbf{x}})}{\partial \mathbf{x}} \delta \mathbf{x} + \mathbf{O}(\delta \mathbf{x}^2)$$

取一階導數近似式、扣除擾動高次項和假設

$$\frac{\partial \hat{\mathbf{x}}}{\partial t} = \mathbf{n}(\hat{\mathbf{x}}), \text{ 可得}$$

$$\frac{\partial \delta \mathbf{x}}{\partial t} = \frac{\partial \mathbf{n}}{\partial \mathbf{x}} \delta \mathbf{x}$$

對 $\delta \mathbf{x}$ 而言，上式可視為線性的方程式，在給定一個初始值 $\delta \mathbf{x}(t_0)$ 後，如果寫成離散型態，可暗示存在一個線性傳遞算子 $\mathbf{L}(t, t_0)$ 與初始值作用得到

$$\begin{aligned} \delta \mathbf{x}(t) &= \mathbf{L}_1(t - \Delta t) \delta \mathbf{x}(t - \Delta t) \\ &= \mathbf{L}_1(t - \Delta t) \mathbf{L}_2(t - 2\Delta t) \delta \mathbf{x}(t - 2\Delta t) \\ &= \dots \\ &= \mathbf{L}(t, t_0) \delta \mathbf{x}(t_0), \end{aligned} \quad (5)$$

其中 Δt 表示數值時間積分的一小段時間步長。

如果用 $\frac{\partial \hat{\mathbf{x}}}{\partial t} = \mathbf{n}(\hat{\mathbf{x}})$ ，先行計算每個時間步驟的

$\hat{\mathbf{x}}$ ，再將 $\frac{\partial \mathbf{n}(\hat{\mathbf{x}})}{\partial \mathbf{x}}$ 算出，在一步的數值時間積分內視為常數，而 $\delta \mathbf{x}$ 視為跟者每一步上相對 $\hat{\mathbf{x}}$ 的擾動，則(5)式可稱為切線性模式。再將(5)式代回(4)式，得

$$\begin{aligned} \delta \mathcal{J} &= \left[\sum_{t=t_0}^{t_f} \mathbf{W}(\mathbf{x}(t) - \bar{\mathbf{x}}(t)) \right]^T \mathbf{L}(t, t_0) \delta \mathbf{x}(t_0) \\ &\quad + [\mathbf{B}^{-1}(\mathbf{x}(t_0) - \mathbf{x}_b)]^T \delta \mathbf{x}(t_0) \end{aligned} \quad (6)$$

因為 $\mathbf{L}$ 算子為一線性算子，可再整理成

$$\begin{aligned} \delta \mathcal{J} &= \left[\sum_{t=t_0}^{t_f} \mathbf{L}^T(t, t_0) \mathbf{W}(\mathbf{x}(t) - \bar{\mathbf{x}}(t)) \right. \\ &\quad \left. + \mathbf{B}^{-1}(\mathbf{x}(t_0) - \mathbf{x}_b) \right]^T \delta \mathbf{x}(t_0). \end{aligned} \quad (7)$$

由導數定義

$$\delta \mathcal{J} = (\nabla \mathcal{J})^T \delta \mathbf{x}(t_0), \quad (8)$$

較(7)、(8)兩式，可知

$$\nabla \mathcal{J} = \sum_{t=t_0}^{t_f} \mathbf{L}^T(t, t_0) \mathbf{W}(\mathbf{x}(t) - \bar{\mathbf{x}}(t)) + \mathbf{B}^{-1}(\mathbf{x}(t_0) - \mathbf{x}_b) \quad (9)$$

這裏的 $\mathbf{L}^T$ 就是伴隨算子，因為是 $\mathbf{L}$ 的轉置運算，所以可想成是一個由 t 到 t_0 的時間回溯運算。此外，由(8)式來看，我們發現目標函數對初始值的梯度 $\frac{\partial \mathcal{J}}{\partial \mathbf{x}(t_0)}$ ，可以由(9)式求得。如果定義一個變數 $\delta \mathbf{x}_a(t_0, t)$ 與梯度關係成為

$$\frac{\partial \mathcal{J}}{\partial \mathbf{x}(t_0)} = \sum_{t=t_0}^{t_f} \delta \mathbf{x}_a(t_0, t) + \mathbf{B}^{-1}(\mathbf{x}(t_0) - \mathbf{x}_b) \quad (10)$$

而其中 $\delta \mathbf{x}_a(t_0, t)$ 可定義成

$$\delta \mathbf{x}_a(t_0, t) \equiv \mathbf{L}^T(t_0, t) \delta \mathbf{x}_a(t) \quad (11)$$

$$\delta \mathbf{x}_a(t) = \mathbf{W}(\mathbf{x}(t) - \bar{\mathbf{x}}(t))$$

於上式中，可以想成一個變數 $\delta \mathbf{x}_a(t)$ 被設定後，代入 $\mathbf{L}^T$ 算子計算。(11)式就是伴隨模式（離散型態），而採用 $\mathbf{x}_a$ 變數的寫法就是視為相應的伴隨變數。從(9)，(10)，(11)式的關係中，可發現目標函數對初始值的梯度值在暫不考慮背景場誤差的作用下，可經由伴隨模式計算（(11)式）求得。由於(11)式是一個線性模式，所以於計算(9)式等號右手邊第一項時，不需重複計算伴隨模式，僅需積分一次伴隨模式，將每個特定時間步驟中的 $\delta \mathbf{x}_a(t) = \mathbf{W}(\mathbf{x}(t) - \bar{\mathbf{x}}(t))$ 視為伴隨模式的強制項（forcing term），在特定時間內加入伴隨模式中。

在做四維變分時，並不會在意切線性模式中的 $\delta \mathbf{x}(t)$ 為多少，它只是推導梯度過程時的一個關係式，當目標函數中的 $\mathbf{W}(\mathbf{x}(t) - \bar{\mathbf{x}}(t))$ 決定

後，可以用伴隨模式得到梯度，當考慮背景場誤差項時需再加入(9)式右手邊第二項 $\mathbf{B}^{-1}(\mathbf{x}(t_0) - \mathbf{x}_b)$ 的計算。當梯度決定後，如前述提到的作法，沿正梯度方向應用數值疊代法運算可得目標函數的極大值，沿負梯度方向可得出極小值。技術性來說，如果單純以伴隨模式所求之梯度來求算極值的效率並非最好的方式，最基本也應該合成兩個梯度向量變成一個共軛梯度來求解。在下一節中將介紹 L-BFGS 法，利用合成數個梯度向量求算極值。

（四）L-BFGS 極小化程序簡介

當考慮最陡下降法（steepest descent method），如共軛梯度法（conjugate gradient method）尋求目標函數的極值時，僅用到此函數的一階導數，如果能用二階導數的資訊將能得到更精確的函數極值。當考慮複雜方程組時，所定義的函數一階導數並不容易直接解析計算，遑論二階導數，這也正是為何氣象研究中為求得目標函數的一階導數而發展數值法伴隨模式。也許有人會說，一階導數、二階導數用解析方法如果困難，可再用數值法計算，但困難在何處呢？我們認為原因在於：1. 不論如何推導，線性代數中在內積空間的恆等式 $\langle \mathbf{x}, \mathbf{L}\mathbf{y} \rangle = \langle \mathbf{L}^T \mathbf{x}, \mathbf{y} \rangle$ 要能滿足，如此在從(6)式推導至(7)式時是一個精確的步驟，有助於決定較精確的目標函數梯度值。不像解析計算積分變換時，常要考慮邊界條件的決定，在換成數值模式時，又要很小心設定邊界，否則將無法滿足內積空間的恆等式，這些討論可見曾等（1997）和 Talagrand and Courtier（1991）用簡化模式所做的探討。2. 單就執行伴隨模式所花費的數值計算時間而言，已為原模式的 3 ~ 4 倍，如果再有一個二階導數的模式，時間花費可

能要更長。

所以於 80 年代初期，Broyden、Fletcher、Goldfarb、Shanno 等人提出用一階導數去近似二階導數的方法，縮短了求算二階導數的困難性，最佳化（optimization）的教科書中稱之為 BFGS 法（BFGS method，為紀念這四個人，以其姓氏的第一字母為名）。隨後 Gill and Murray (1979)、Shanno and Phua (1980)、Buckley and Lenir (1983)、Liu and Nocedal (1989) 等修正這個方法，減少數值運算時所用的記憶體，稱為 Limited-memory BFGS 法，縮寫簡稱為 L-BFGS 法。

BFGS 法的流程簡述如下，設有一目標函數 J ，由一自變數初始向量 $\mathbf{x}$ 所組成，可對泰勒級數展開至二階精度，

$$J(\mathbf{x} + \mathbf{x}') = J(\mathbf{x}) + \mathbf{c}^T \mathbf{x}' + 0.5 \times \mathbf{x}'^T \mathbf{G} \mathbf{x}', \quad (12)$$

其中 $\mathbf{G}$ 為 Hessian matrix，為函數的二階變量，而 $\mathbf{x}'$ 為擾動向量，當 J 對 $\mathbf{x}'$ 有極值存在時， $\partial J / \partial \mathbf{x}' = 0$ ，則

$$\mathbf{c} + \mathbf{G} \mathbf{x}' = 0 \quad (13)$$

如前段提及，大氣模式有一定的複雜程度，並不容易從所定義的目標函數求得二階導數，而 BFGS 法即針對此，從(13)式出發，利用一階導數求得進似的二階導數，並根據此二階導數求算目標函數極值。詳細的推導過程請參見 Gill et al. (1981)，於此僅介紹實際應用時的流程。

步驟 1. 假設已經有一初始場 $\mathbf{x}^{(0)}$ ，其中上標表最佳化計算時的疊代次數，現設定第一次的 $\mathbf{G}^{(0)}$ 為單位矩陣 $\mathbf{I}$ ，並自行決定一個收斂係數 ε 。

步驟 2. 利用伴隨模式計算目標函數梯度，即

$$\mathbf{c}^{(0)} = \nabla J(\mathbf{x}^{(0)}) \quad (14)$$

同時，於每個計算過程中，驗證 $\|\mathbf{c}^{(k)}\| < \varepsilon$ 是否滿足，如不成立則停止計算或放大 ε 的值。

步驟 3. 設(13)式成立，則可解出搜尋向量（search direction）， $\mathbf{d}$ ，即

$$\mathbf{G}^{(k)} \mathbf{d}^{(k)} = -\mathbf{c}^{(k)} \quad (15)$$

步驟 4. 計算一最佳步長值（step size）， α ，使能極小化目標函數，即

$$J(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}) < J(\mathbf{x}^{(k-1)}), \text{ 如此可決定下一次疊代的向量}$$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}$$

步驟 5. 重新計算 Hessian matrix，

$$\mathbf{G}^{(k+1)} = \mathbf{G}^{(k)} + \mathbf{D}^{(k)} + \mathbf{E}^{(k)}$$

此處

$$\mathbf{D}^{(k)} = \frac{\mathbf{y}^{(k)} \mathbf{y}^{(k)T}}{(\mathbf{y}^{(k)} \cdot \mathbf{s}^{(k)})}; \quad \mathbf{E}^{(k)} = \frac{\mathbf{c}^{(k)} \mathbf{c}^{(k)T}}{(\mathbf{c}^{(k)} \cdot \mathbf{d}^{(k)})}$$

而

$$\mathbf{s}^{(k)} = \alpha_k \mathbf{d}^{(k)}$$

$$\mathbf{y}^{(k)} = \mathbf{c}^{(k+1)} - \mathbf{c}^{(k)}$$

$$\mathbf{c}^{(k+1)} = \nabla J(\mathbf{x}^{(k+1)})$$

步驟 6. 回步驟 3 重複執行 3 ~ 6 的步驟，得出下一次的搜尋向量，並計算目標函數極值直到滿足所設定的標準，或是計算資源所能容許的範圍。

L-BFGS 和 BFGS 的不同點在於針對上述步之(15)式做改良，主要的原因是實作時 $\mathbf{G}^{(k)}$ 的維度可能很大，當求解(15)式時，需要計算反矩陣 $\mathbf{G}^{(k)-1}$ ，不但徒費龐大計算資源，亦常無法得出精確的 $\mathbf{G}^{(k)-1}$ ，其原因為 $\mathbf{G}^{(k)}$ 常屬於矩陣內元素

多為 0 之疏鬆矩陣 (coarse matrix)，故一些學者如 Gill and Murray (1979)、Shanno and Phua (1980)、Buckley and Lenir (1983)、Liu and

Nocedal (1989) 等人，建議不直接求解(15)式，而用兩次不同疊代次數的初始變數向量 $\mathbf{x}^{(k+1)}$, $\mathbf{x}^{(k)}$ 模擬取代。其數學式為，

$$\begin{aligned} \mathbf{G}_{new}^{(k)-1} = & (V^{(k-1)\text{T}} \dots V^{(k-m)\text{T}}) \mathbf{G}_{old}^{(k)-1} (V^{(k-m)} \dots V^{(k-1)}) \\ & + \rho^{(k-m)} (V^{(k-1)\text{T}} \dots V^{(k-m+1)\text{T}}) g^{(k-m)} g^{(k-m)\text{T}} (V^{(k-m+1)} \dots V^{(k-1)}) \\ & + \rho^{(k-m+1)} (V^{(k-1)\text{T}} \dots V^{(k-m+2)\text{T}}) g^{(k-m+1)} g^{(k-m+1)\text{T}} (V^{(k-m+2)} \dots V^{(k-1)}) \\ & + \dots \\ & + \rho^{(k-1)} g^{(k-1)} g^{(k-1)\text{T}} \end{aligned} \quad (16)$$

其中

$$\rho^{(k)} = \frac{1}{y^{(k)\text{T}} g^{(k)}}, \quad V^{(k)} = 1 - \rho^{(k)} y^{(k)} g^{(k)\text{T}}$$

$$g^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$$

於(15)式中發現，只要求算初始的 $\mathbf{G}^{(0)-1}$ ，其後的 $\mathbf{G}^{(k)-1}$ 可用(16)式疊代運算而得出，而上標中的字母 m 表選用的疊代初始向量有 m 個，在本文後續的工作中 $m = 5$ ，意謂用最簡單的共軛梯度法先得出暫時的 5 個疊代階段的初始向量，而後得出 $\mathbf{G}^{(k)-1}$ 後沿前述 6 個步驟求算目標函數極小值。

三、個案分析

1989 年 1 月 27 日 12UTC、1999 年 12 月 19 日 12UTC 和 1999 年 12 月 26 日 12UTC 三個時間，多數的全球大氣數值模式，包括 ECMWF、JMA、NCEP 和氣象局模式的展期預報都失效，以 1999 年 12 月 19 日和 1999 年 12 月 26 日為例 (圖 1)，就 72 小時預報的異常相關係數而言，比平常時約下降了 0.1。在選擇個案時間點前後

24 小時的分析場存有顯著差值，顯示大氣基本流場有顯著變化，可作為研究變分同化或不穩定度研究的課題。所以本研究從這三個個案的時間點出發，於現階段並不打算利用完整的全球模式作分析，而是利用簡化版模式作分析，對於預報準確度的提升將列為下一階段的目標，現階段只進行人為化的比對實驗，驗證伴隨模式與四維變分的可行性。

(一)、比對實驗 (twin experiment)

選取個案時間點附近 12 或 24 小時內的初始場，將其輸入模式中積分，其預報軌跡假定為觀測值的演變，視為實驗的控制組。同時利用選取個案時間點的初始場代入模式中，得到另一組預報結果，作為實驗的實驗組。這兩組結果顯而易見有差異存在，而透過四維變分的極小化程序所得出的新初始場應該和控制組的初始場相近，這種已經知道標準答案為何，而為驗證實驗方法與結果的正確性，就是比對實驗。當然，在這種比對實驗中於 t_0 初始時間的控制組初始場資訊是不能加到目標函數中計算的，因為會使計算的結果過於理想化。本文中對三個個案各自進行 3 個

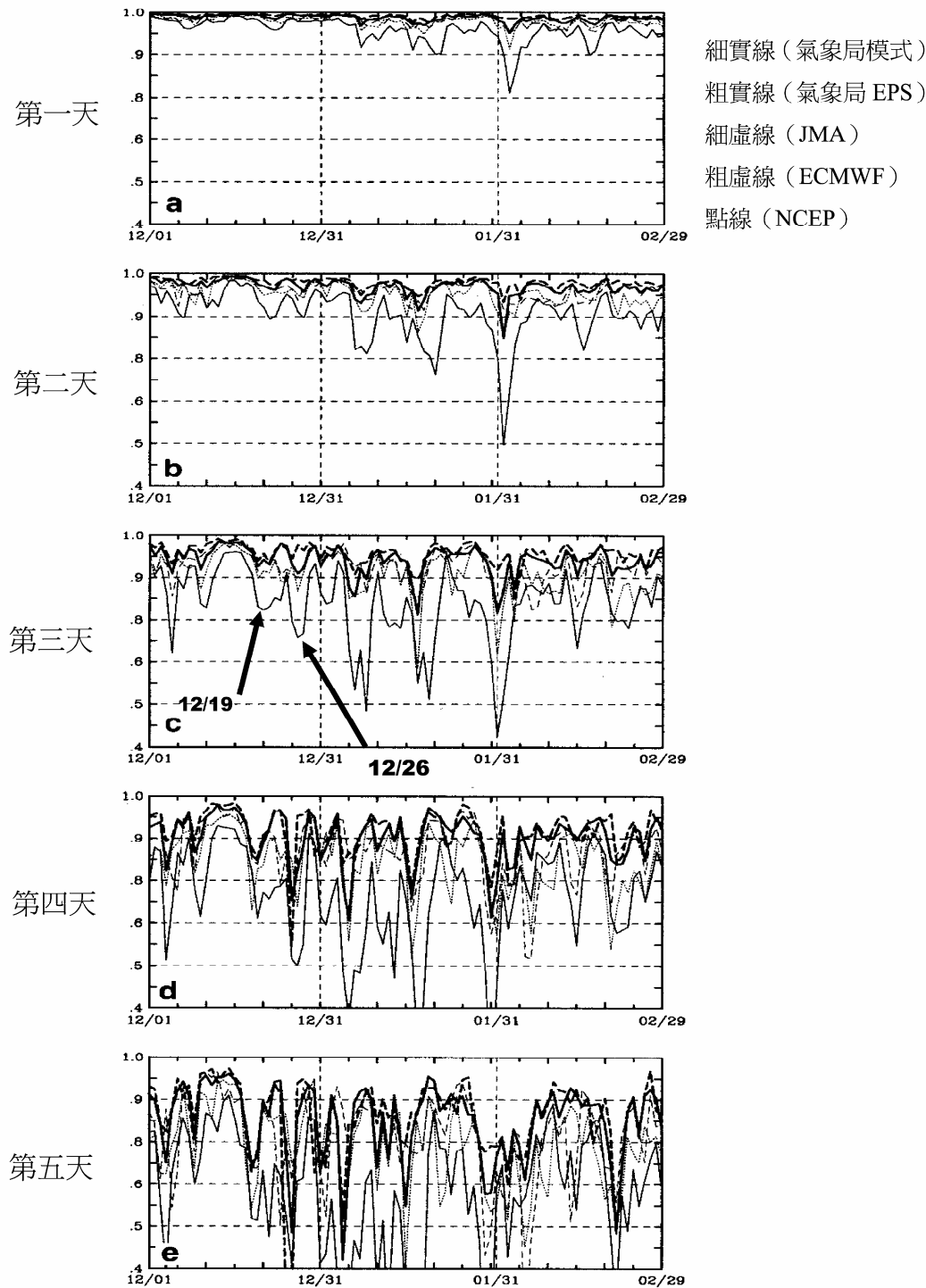


圖 1 1999/2000 冬季各預報中心預報場，以 ECMWF 格點資料為分析場，NCEP 氣候平均場為參考值，求得異常相關係數之時間序列圖，橫座標為日期，縱座標為相關係數值。所選個案以第三天為準，圖中以一粗實線箭頭標出。(取自張庭槐等，2001)

實驗：1.針對所有分析網格所做的理想化實驗；
2.模擬全球觀測場分佈的實驗；3.模擬台灣東亞地區附近有密集觀測的實驗，並就有無背景場誤差估計及疊代次數的效果做討論。3 個實驗的定義與作法留待於後續小節中詳細說明。

所用的目標函數定義按第二節的定義可寫為

$$J = \frac{1}{2} \sum_{n=1}^6 (\mathbf{x}(t) - \bar{\mathbf{x}}(t))^T \times \mathbf{W} (\text{權重矩陣}) \times (\mathbf{x}(t) - \bar{\mathbf{x}}(t)) + \frac{1}{2} e \times (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b)$$

此處 $t = 6\text{hr} + (n-1) \times 6\text{hr}$

其中 $\mathbf{x}$ 為實驗組變數向量， $\bar{\mathbf{x}}$ 為控制組變數向量，而 $\bar{\mathbf{x}}$ 可視為假想觀測分析資料。實驗中每隔

6 小時將 $\mathbf{x}$ 與 $\bar{\mathbf{x}}$ 做差值計算並定義成目標函數(如圖 2 說明)，當透過四維變分的極小化所得出之最佳初始場 (optimal initial fields) 的演化軌跡，假設是 $\mathbf{x}_{new}$ ，在變分的 36 小時窗區內將可以朝 $\bar{\mathbf{x}}$ 靠近。其中， $\mathbf{W}$ 是權重矩陣，其內容是：(1)用模式預報的 24 小時變數 $\mathbf{x}(t = 24\text{hrs})$ 減去 12 小時變數 $\mathbf{x}(t = 12\text{hrs})$ 並取平方值；(2)取模式各個變數、每一層最大值為代表，選定後將各個變數每一層的所有值都用各自的最大平方差值取代；(3)假設模式變數行向量的解析度是 m ，所以 $\mathbf{W}$ 是 $m \times m$ 的方陣，將(2)中的內容取倒數放入 $\mathbf{W}$ 的對角線位置。

上式中的 $\frac{1}{2} e \times (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b)$ 為背

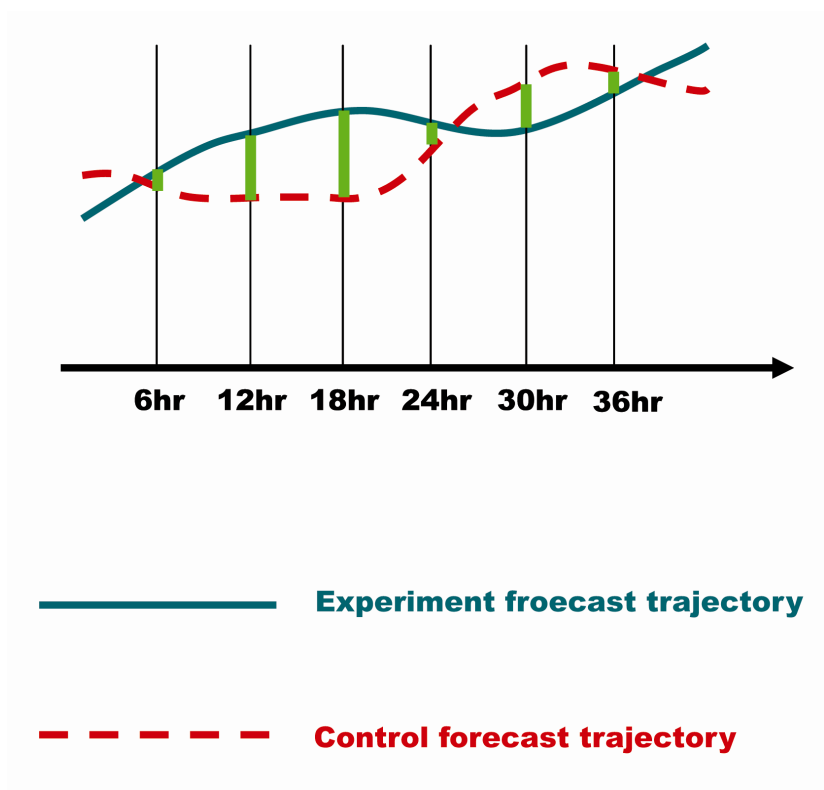


圖 2 本文中所定義目標函數之示意圖。

景場誤差估計項，此處利用 NMC 方法做簡單的 $\mathbf{B}^{-1}$ 計算，其估計方式為所選三個個案，針對特定目標時間，模式 24 小時預報減去 48 小時預報所得之差值平方後取倒數的三個個案平均值，以類似於權重矩陣 $\mathbf{W}$ 的作法放入 $\mathbf{B}^{-1}$ ，而 $\mathbf{B}^{-1}$ 為一對角線矩陣。式中 e 為經驗係數，設為 0.1，這是因為 $\mathbf{B}^{-1}$ 所用的是 24 小時的預報誤差，而在本文中的 $\mathbf{x}_b$ 是用第 1 次的 $\mathbf{x}(t_0)$ 代替，已經是代表初始時間的分析場，不像是實際作業時， $\mathbf{x}_b$ 是用 12 小時前模式的 12 小時預報值替代，所以設 $e = 0.1$ 。過程中， $\mathbf{x}_b$ 不隨變分極小化疊代而變。背景場誤差的加入對所有分析網格所做的理想化比對實驗也許不重要，但對其他二個實驗則有顯著影響，將在隨後詳細說明。

1. 所有分析網格所做的理想化比對實驗

以 1989 年 1 月 27 日的個案為例，選擇 00 UTC 時為控制組的初始場，為一假想正確的初始場 (hypothetical initial fields)，而同日 12 UTC 時為實驗組的初始場，二者的預報結果在非線性大氣模式的主導下，將隨預報時間增長而加大。如果假定 00 UTC 時為正確標準答案（從實際作業來說，00 UTC 的預報結果也比 12UTC 好），而 12 UTC 時為不完美的初始場，換言之，設定目標函數為兩者預報結果在 36 小時內每間隔 6 小時的總和。同理，本文亦針對 1999 年 12 月 19 日和 1999 年 12 月 26 日兩個個案做比對實驗，但假設的完美初始場為 12 月 18 日 12UTC 和 12 月 27 日 12UTC。在後續的文章中，對個案的稱呼將用實驗組的初始場時間的簡稱替代，如 1989 年 1 月 27 日 12UTC 稱為 89012712；1999 年 12 月 19 日 12UTC 稱為 99121912；1999 年 12 月 26 日 12UTC 稱為 99122612。

此處控制組變數 $\bar{\mathbf{x}}$ 的空間解析度與實驗組 $\mathbf{x}$ 一樣，也就是說在每個分析網格上都有一個觀測資料幫忙做變分同化，所以是一種對所有分析網格所做的理想化比對實驗。這個實驗主要的目的是測試四維變分的極小化程序、伴隨模式等配置是否正確，其結果可視為理想的參考值，可知道目標函數下降幾個數量級到達理論下限值，亦可決定疊代次數幾次會是一個好的選擇。

2. 模擬全球觀測場分佈的實驗

控制組與實驗組的日期設定同第 1 小節所述，但在此實驗中控制組變數 $\bar{\mathbf{x}}$ 的空間解析度不同於實驗組變數 $\mathbf{x}$ 的解析度，而且空間上分佈不規則，也就是模擬全球觀測場分佈的情形。此處的作法是隨機的在每一個模式預報層拿走一些資料，垂直層只剩 $\sigma = 0.2, 0.25, 0.3, 0.5, 0.7, 0.83, 0.85, 0.95$ 有取用，大約去除了相對於模式原有解析度一半的資料。設計這個實驗的目的是：模擬真實資料不規則的分佈情形；並測試背景差誤差項對變分過程的作用。所以，目標函數中的前一項已經變成

$$\mathcal{J} = \frac{1}{2} (\mathbf{H}\mathbf{x} - \bar{\mathbf{x}})^T \mathbf{W} (\mathbf{H}\mathbf{x} - \bar{\mathbf{x}})$$

而 $\mathbf{H}$ 就是我們改變取用資料的算子，此處為簡化問題，是一個由 0, 1 元素構成的對角線矩陣，而後續的伴隨模式給定類似於第二節所述，於此不再贅述。

3. 模擬台灣東亞地區附近有密集觀測的實驗

控制組與實驗組的日期設定同第 1 小節所述，和第 2 個實驗相似的是：在此實驗中控制組變數 $\bar{\mathbf{x}}$ 的空間解析度也不同於實驗組變數 $\mathbf{x}$ 的解析度，此處選擇經緯度範圍

(100° E, 2° N) ~ (140° E, 57° N) 的區域加入控制組變數 $\bar{\mathbf{x}}$ ，但在垂直層上則保留與原實驗組同樣的解析度。定義目標函數的方式類似於第 2 個實驗，也用一個 $\mathbf{H}$ 算子將所要計算的區域取出。這個實驗的目的是：模擬台灣東亞地區附近有密集觀測，將其結果與第 2 個節模擬全球觀測場分佈的實驗相比較，同時，測試背景差誤差項對變分過程的作用。

首先，就三個個案 3 個實驗的目標函數與梯度大小變化作探討。不考慮背景場的誤差項作用下，圖 3 為目標函數的對數值 (y 軸) 隨疊代次數 (x 軸) 的變化。為了將 3 種模擬觀測場作用的實驗表示在一張圖上，y 軸的座標有兩種，左側的量級較大 (約 $10^5 \sim 10$)，用來標示對所有分析網格所做的理想化比對實驗；而右側的量級較小 (約 $10^4 \sim 10^{-3}$)，用來標示模擬全球觀測及模擬台灣地區密集觀測的實驗。經過 100 次的疊代後，三個個案、3 種比對實驗的線型分別是：1. 所有分析網格的理想化比對實驗目標函數 (棕色細實線)、梯度 (藍色粗實線)；2. 模擬全球觀測場分佈的目標函數 (綠色細虛線)、梯度 (黑色粗虛線)；3. 模擬台灣東亞地區附近密集觀測的目標函數 (粉紅色細虛線)、梯度 (紫色粗虛線)。所有目標函數全都大約下降了 2 個數量級以上，而梯度也下降了約 1 個數量級以上，其中以模擬全球觀測場分佈的實驗下降最多，目標函數約 4 個數量級以上，而梯度也有 3 個數量級。推究模擬全球觀測場分佈的目標函數會下降較其他 2 種實驗為多的原因是：目標函數的變化來自全球區域且小於分析網格數目，在分析網格解析度上模式可以有比理想化實驗較大的自由度來做調整，同時有比模擬台灣東亞地區更多的資訊朝假想的

控制組觀測場接近。

考慮背景場的誤差項作用下，圖 4 為目標函數的對數值 (y 軸) 隨疊代次數 (x 軸) 的變化。同圖 3 的表現方式也用了 2 個左右側不同的數量級來表現不同實驗的目標函數變化。圖 4 中的線型定義如同圖 3，從線型走勢來看，目標函下降的數量級沒有圖 3 來的大，約 1 個數量級多，這是因為背景場誤差估計項

$\frac{1}{2} \mathbf{e} \times (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b)$ 的值是隨疊代次數而增加，所以目標函數值無法以像沒有背景場誤差的目標函數值之下降方式呈現。但是，我們不能直接比較沒有背景場誤差的目標函數與有背景場誤差的目標函數，並以其大小而判斷極小化程序的好壞，真正的比較應該是驗證隨後預報的誤差值，這將留於本節中的第 (三) 小節中討論。

圖 3, 4 中，很明顯的 1989 年的個案與 1999 年相差很大，很可能是 1989 年是個大氣阻塞快速生成的例子，雖然差異 12 小時的大氣基本場差異不大 (事實上反而是三個個案中最小)，但是變分程序歷經梯度值劇烈變動才找到方向使目標函數值下降，以前 10 次的疊代運算而言，這個個案目標函數下降率是最慢的。有關這個個案的資料詳細診斷分析和大氣基本流場不穩定的分析大家可以再參考 Tibaldi and Molteni (1990) 的文章。

另一方面，圖 3, 4 的意義除顯示極小化過程的進行外，亦可做為往後計算資源限制下應做多少次疊代運算的參考。如實際作業時，不可能執行到 100 次的疊代運算，從其他作業中心的經驗及變分極小化的學理來看，目標函數應該下降 1 個數量級以上就可以得到不錯的新初始場。

(二)、極小化程序的結果分析

除目標函數值下降外，經過極小化程序後所得之 $\mathbf{x}_m^{(100)}(t_0)$ 和 $\bar{\mathbf{x}}(t_0)$ 之差異應遠小於 $\mathbf{x}(t_0)$ 和 $\bar{\mathbf{x}}(t_0)$ 之差異。所設計的 3 個實驗中當然以所有分析網格加入的理想化比對實驗所得結果最佳，意即最接近控制組初始場 $\bar{\mathbf{x}}(t_0)$ 。於此，可以定義一個比值

$$ratio = \frac{(4dvarini \quad \mathbf{x}_m^{(100)}(t_0) - \bar{\mathbf{x}}(t_0))^2}{(nonvarini \quad \mathbf{x}(t_0) - \bar{\mathbf{x}}(t_0))^2}$$

驗證經變分分析後初始場與假想觀測場 $\bar{\mathbf{x}}(t_0)$ 差

值平方與未經變分分析初始場與假想觀測場 $\bar{\mathbf{x}}(t_0)$ 差值平方的比值，其結果如表 1 所示。從表 1 中發現，對於三個個案，4 個模式預報變數 u, v, θ, Ps （東西向風，南北向風，虛位溫，地面氣壓）而言，變分分析的作用約修正掉 80% 的原初始場與假想觀測場的差距（比值 0.2 表示 80% 的修正量）。其中以 99121912 這個個案修正最多，所示的變數值都修正掉 90%。以變數間修正量的大小而言，發現東西向風場修正的值比其他變數場來的少，原因為何尚須進一步研究。此處，未顯示比濕的結果，原因是所用的模式在水氣作

表 1 經變分分析後初始場與假想觀測場差值平方與未經變分分析初始場與假想觀測場差值平方的比值。

初始場差值平方比值	89012712	99121912	99122612
u	0.246	0.094	0.128
v	0.173	0.072	0.069
θ	0.168	0.022	0.049
Ps	0.163	0.053	0.055

表 2 加入背景場誤差估計，經變分分析後初始場與假想觀測場差值平方與未經變分分析初始場與假想觀測場差值平方的比值。

初始場差值平方比值	89012712	99121912	99122612
u	0.253	0.100	0.138
v	0.179	0.075	0.074
θ	0.176	0.024	0.054
Ps	0.165	0.059	0.044

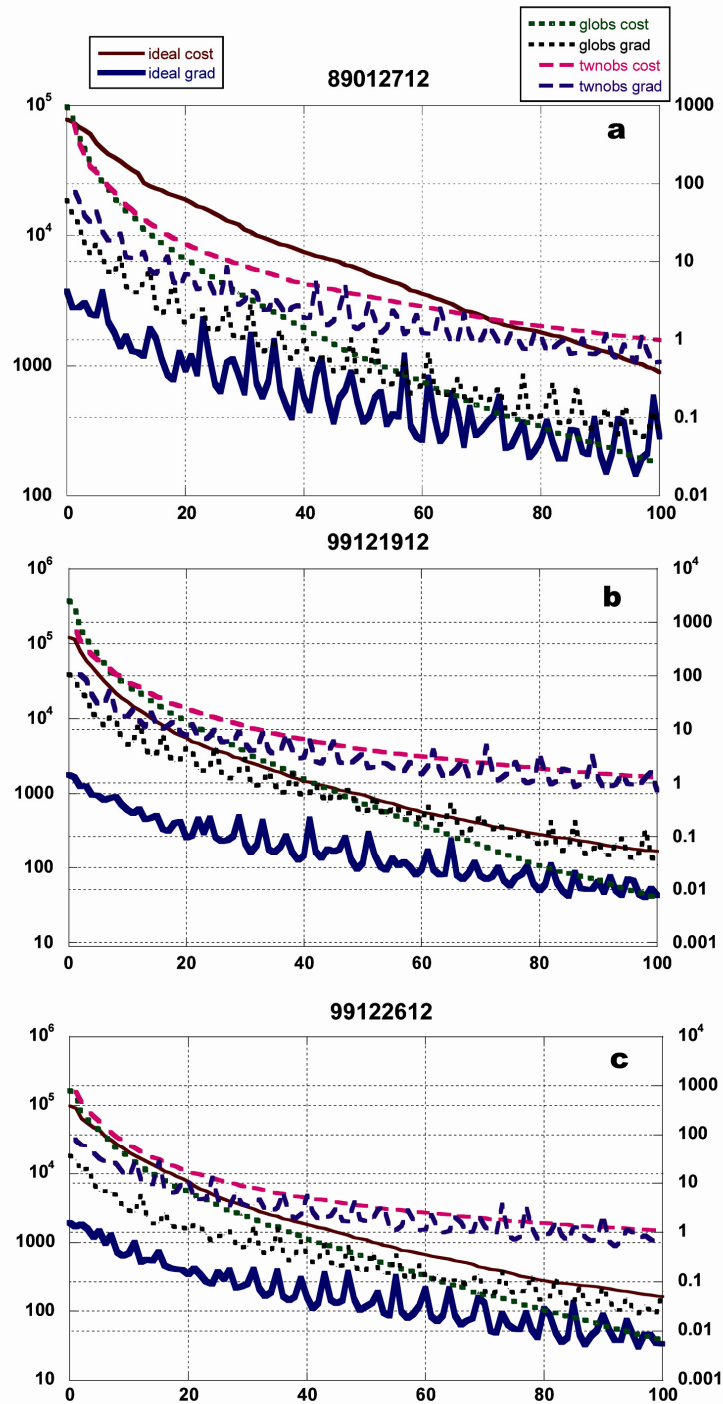


圖 3 三個個案 36 小時預報比對實驗目標函數對數值和目標函數梯度對數值隨疊代次數的演變圖。棕色線與粗藍線分別代表全部分析網格變分實驗的目標函數對數與梯度對數，座標讀數參考左邊。綠色線與黑色線是模擬全球觀測分佈變分實驗的目標函數對數與梯度對數；粉紅色與紫色線是模擬台灣東亞區密集觀測分佈變分實驗的目標函數對數與梯度對數，這 2 個實驗座標讀數參考右邊。

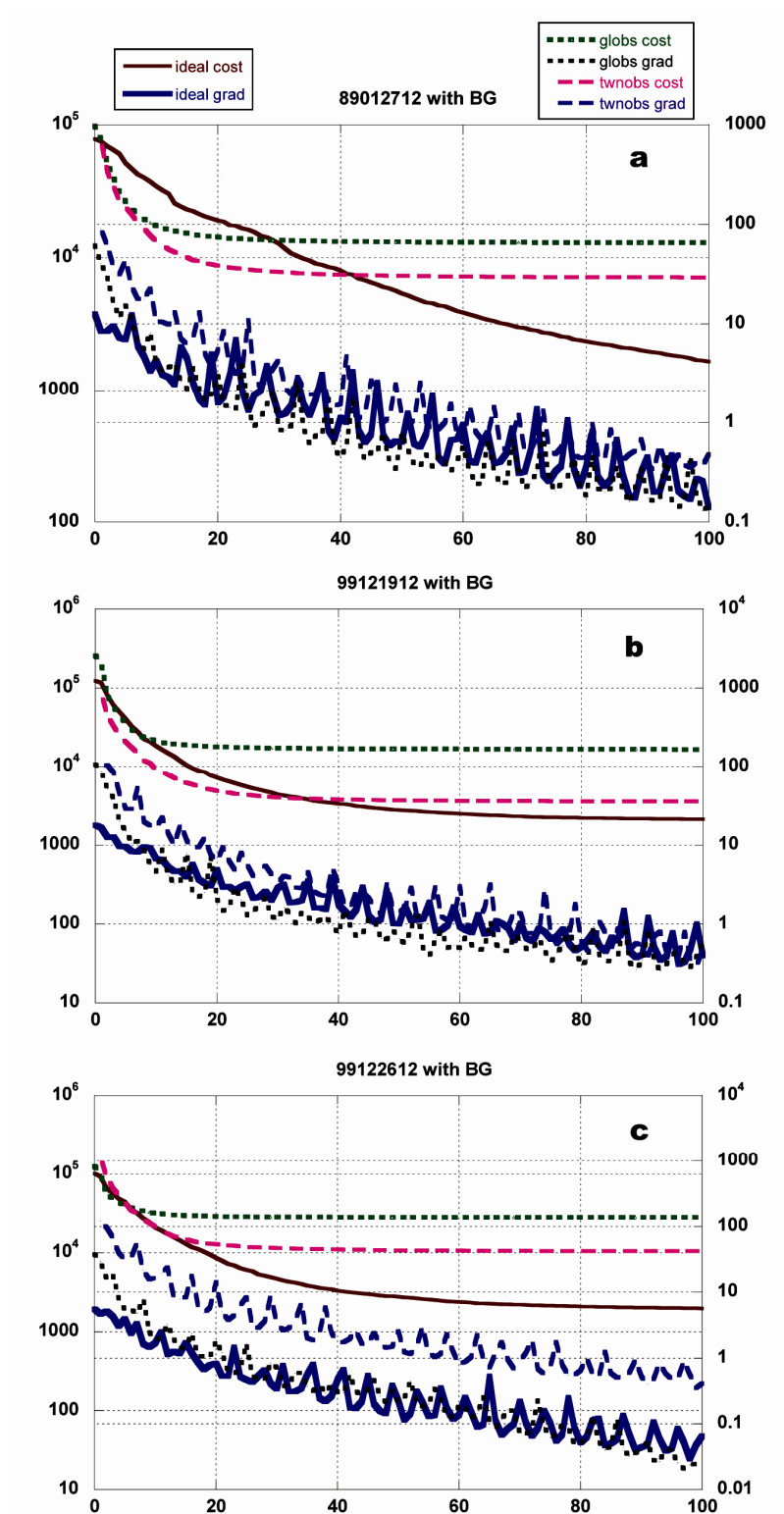


圖 4 同圖 3，但為加入背景場誤差項後的比對實驗目標函數對數值和目標函數梯度對數值隨疊代次數的演變圖。

用上並不完善，結果較不具代表性。同樣的，加入背景場誤差後的結果如表 2 所示。

另一方面，從目標函數的變化，也暗示假想觀測場新初始場朝接近，但是並不是說初始場愈接近假想觀測場就一定會有類似假想觀測場的預

報軌跡，這是由於大氣模式也是一種非線性系統，差異不大的初始場其預報軌跡不必然一樣。此處，僅是驗證極小化後初始場的結果，當然，在下一小節的討論中，假想觀測資料給定恰當的話，會發現有經過變分調整的初始場是比未經過

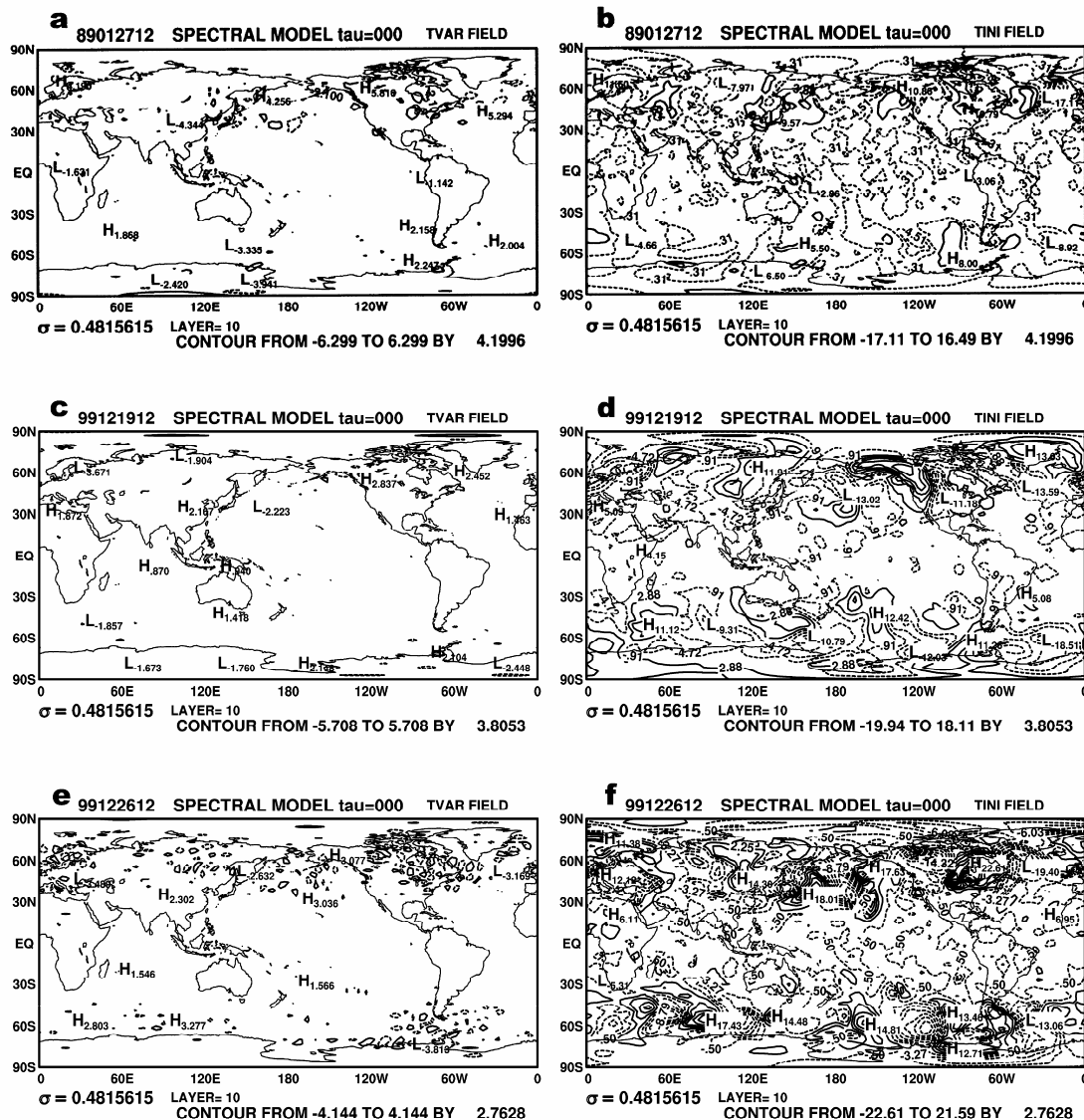


圖 5 約於 500 hPa 上, 經極小化變分分析後的初始場與原控制組初始場位溫 θ 的差值(圖 a: 89012712, 圖 c: 99121912, 圖 e: 99122612) 和未做變分分析之實驗組初始場與原控制組初始場位溫 θ 的差值(圖 b: 89012712, 圖 d: 99121912, 圖 f: 99122612) 之相互比較。

變分調整的好。

以驗證極小化程序的目的而言，於此僅圖示理想化實驗後的初始場，而理想化實驗中背景場誤差項的有無，由表 1 和表 2 的表現可知，對所得的初始場影響不大，故選擇圖示無背景場作用的結果，顯示圖層為 $\sigma = 0.48$ (約 500 hPa)，選擇顯示的變數為 θ 。圖 5 中是三個個案針對模式變數 θ 經極小化變分與否後的差異，為比較方便差異值選用相同的等值線間距作圖。顯然，調整後的初始場與原先初始場的差值遠小於未經變分分析的實驗組初始場與控制組初始場的差值，變分極小化的過程的確可以逼近已知答案的假想初始場。

由三組個案的 θ 調整後的分佈位置來看(u, v 也有類似情形，但未圖示)，可發現都處於噴流區附近，這些地區正好是大氣系統反應最敏感的地區，經過 100 次的疊代運算仍不能去除與假想正確初始場的差距，間接表示變分極小化程序已經不能再去去除這些地方的訊號，而這些地區正是影響往後預報的重點地區，有必要從學理的角度分析（如低頻震盪等理論）而非單純用變分極小化程序解決問題。再者，調整後的初始場顯示南、北極區都有一些訊息，代表模式對極區的處理也有改進空間。

(三)、新初始場的預報誤差分析

本節中將討論經變分極小化程序所得出的新初始場加回原預報模式後，所得的預報軌跡與未經過變分極小化的初始場預報軌跡有何差異。本文使用平均誤差 (total mean errors) 來定義彼此差異的大小，其為

$$\text{total mean errors} \equiv \frac{\sum_{i=1}^N \mathbf{M}_i \times (\mathbf{F}_i - \mathbf{A}_i)^2}{\sum_{i=1}^N \mathbf{M}_i}$$

其中， i 為網格指標， N 為所有網格數，所有變數的總和， $\mathbf{M}_i = \cos(\phi_i)$ 為面積因子， ϕ_i 為網格指標所在緯度， $\mathbf{F}$ 為實驗組預報值， $\mathbf{A}$ 為控制組預報值（假想成是實際大氣於分析網格上的觀測軌跡）， $\mathbf{A}$ 值僅隨個案不同而變，在每一個個案中的不同實驗中是相同的。

不考慮背景場誤差項的作用下，三個個案，3 個實驗的預報誤差分析如圖 6 所示。圖中橫座標表示預報時間，從 48 小時至 192 小時，縱座標表示平均誤差值，縱座標值愈大，表示平均誤差愈大，與控制組的預報軌跡相差愈大。圖中紅色線表示未經變分作用的實驗組預報軌跡與控制組預報軌跡的平均誤差值（圖中說明以 ctrlerrs 代表），可想成是一個標的值，經變分分析後的預報軌跡必須小於這條紅線值，才能說變分分析的工作成功。藍色線為經全部分析網格所做的理想化變分實驗後所得實驗組預報軌跡與控制組預報軌跡的平均誤差值（圖中說明以 idealerrs 代表）。從藍色線的平均誤差值來看，可說是幾乎接近控制組的預報軌跡，當然這是理想的比對實驗，所以結果特別好。綠色線為模擬全球觀測場分佈的變分實驗預報軌跡與控制組預報軌跡的平均誤差值（圖中說明以 globerrs 代表），可發現以 99121912 這個個案表現較佳，其他二個個案在到達 120 小時後，平均誤差值變大，對 89122712 這個個案而言，甚至在 120 小時後比未經變分分析的預報誤差值還大。黑色線為模擬台灣東亞地區附近有密集觀測的實驗預報軌跡與控制組預報軌跡的平均誤差值（圖中說明以 twobserrs 代表），圖中僅有 99121912 這個個案誤差值小於紅色線，其餘二個個案都不是很好，對於 89122712 這個個案而言，甚至有做變分分析所得預報誤差比沒有做變分分

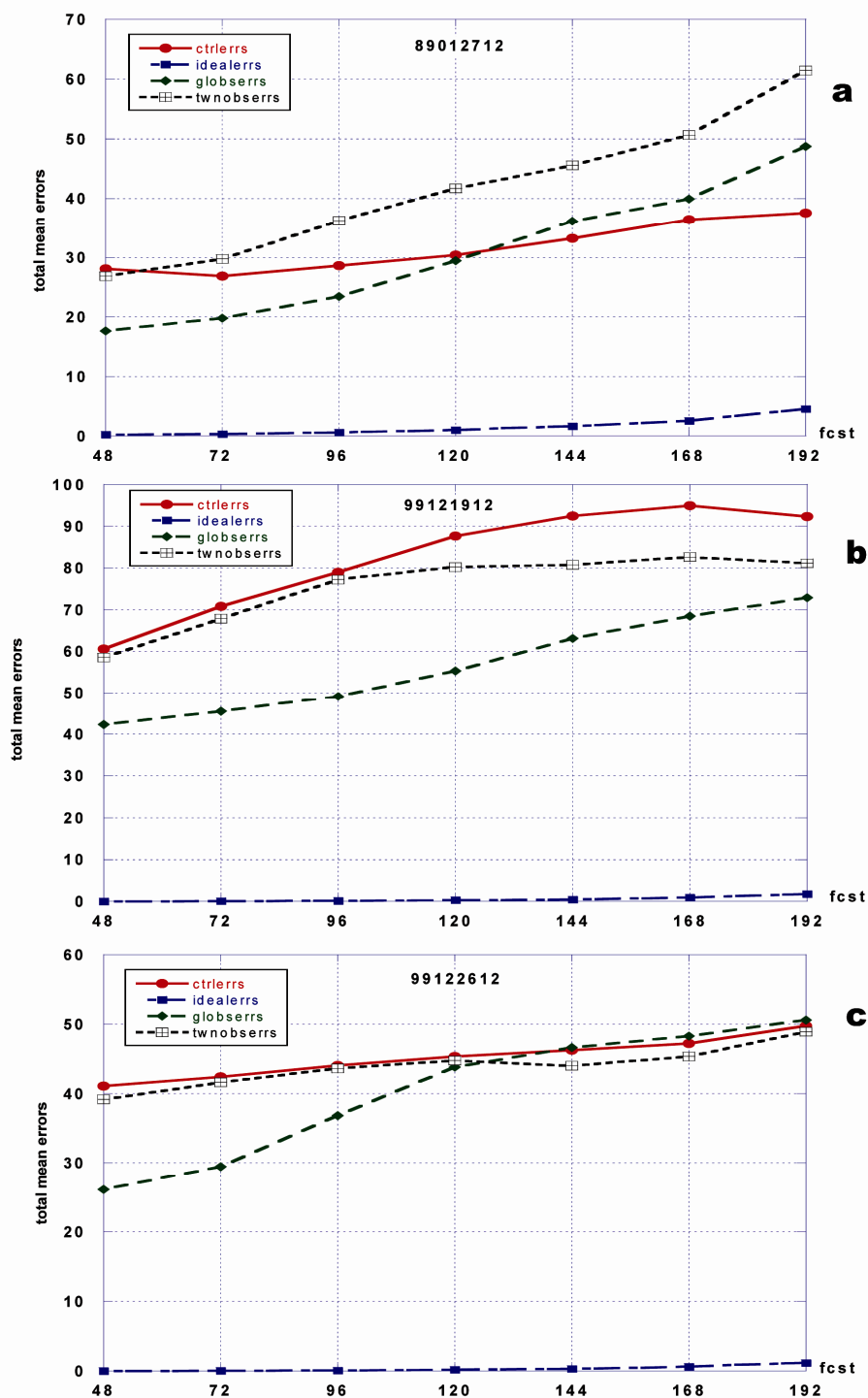


圖 6 沒有加入背景場誤差項，三個個案（圖 a: 89012712; 圖 b: 99121912; 圖 c: 99122612）經 3 個比對實驗的預報平均誤差，分別是：紅色線表未經變分分析的預報平均誤差；藍色線表經全部分析網格變分分析後的預報平均誤差；綠色線表經模擬全球觀測變分分析後的預報平均誤差；黑色線是模擬台灣東亞地區密集觀測後變分分析的預報平均誤差。

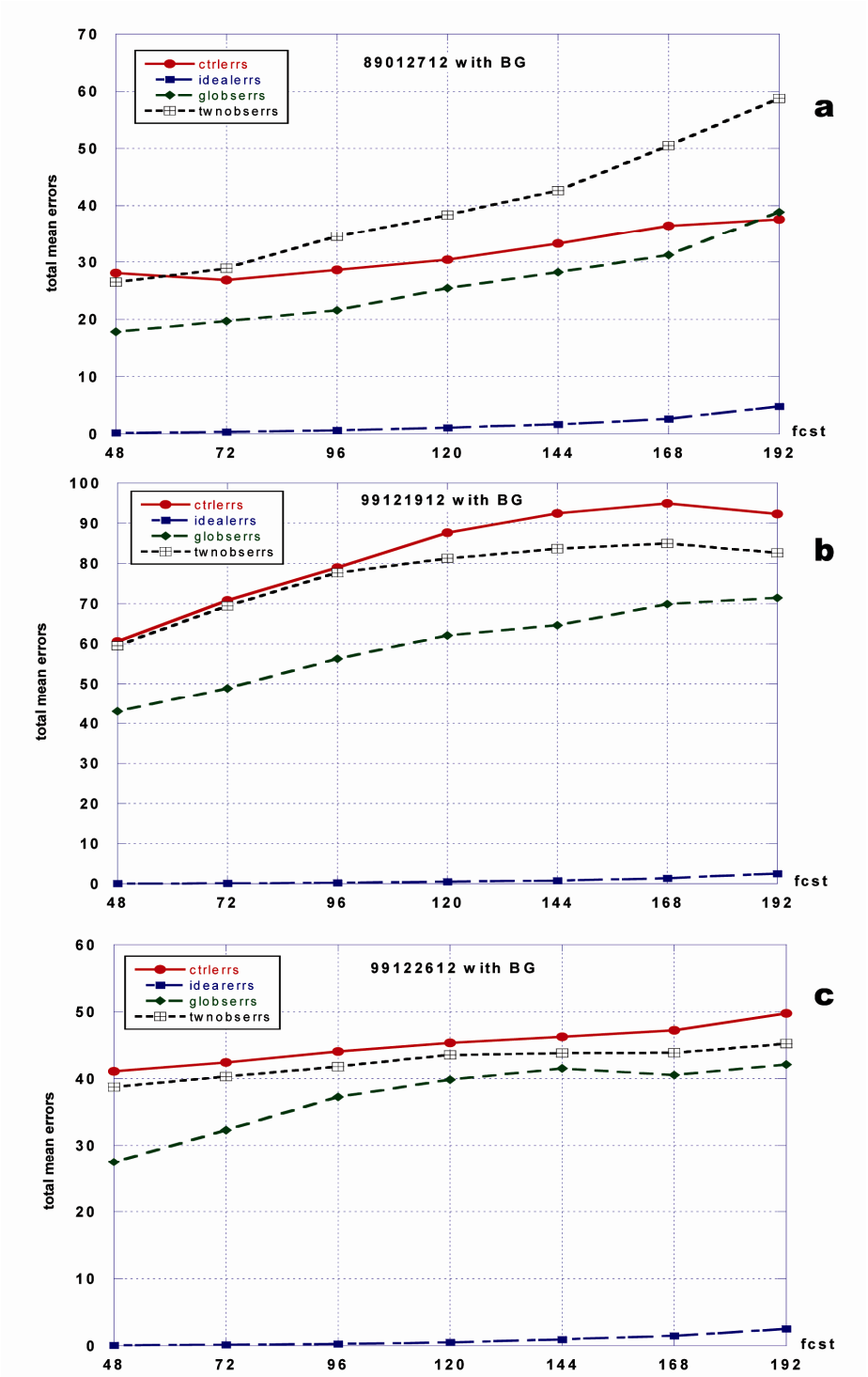


圖 7 加入背景場誤差項後，三個個案（圖 a: 89012712; 圖 b: 99121912; 圖 c: 99122612）經 3 個比對實驗的預報平均誤差，分別是：紅色線表未經變分分析的預報平均誤差；藍色線表經全部分析網格變分分析後的預報平均誤差；綠色線表經模擬全球觀測變分分析後的預報平均誤差；黑色線是模擬台灣東亞地區密集觀測後變分分析的預報平均誤差。

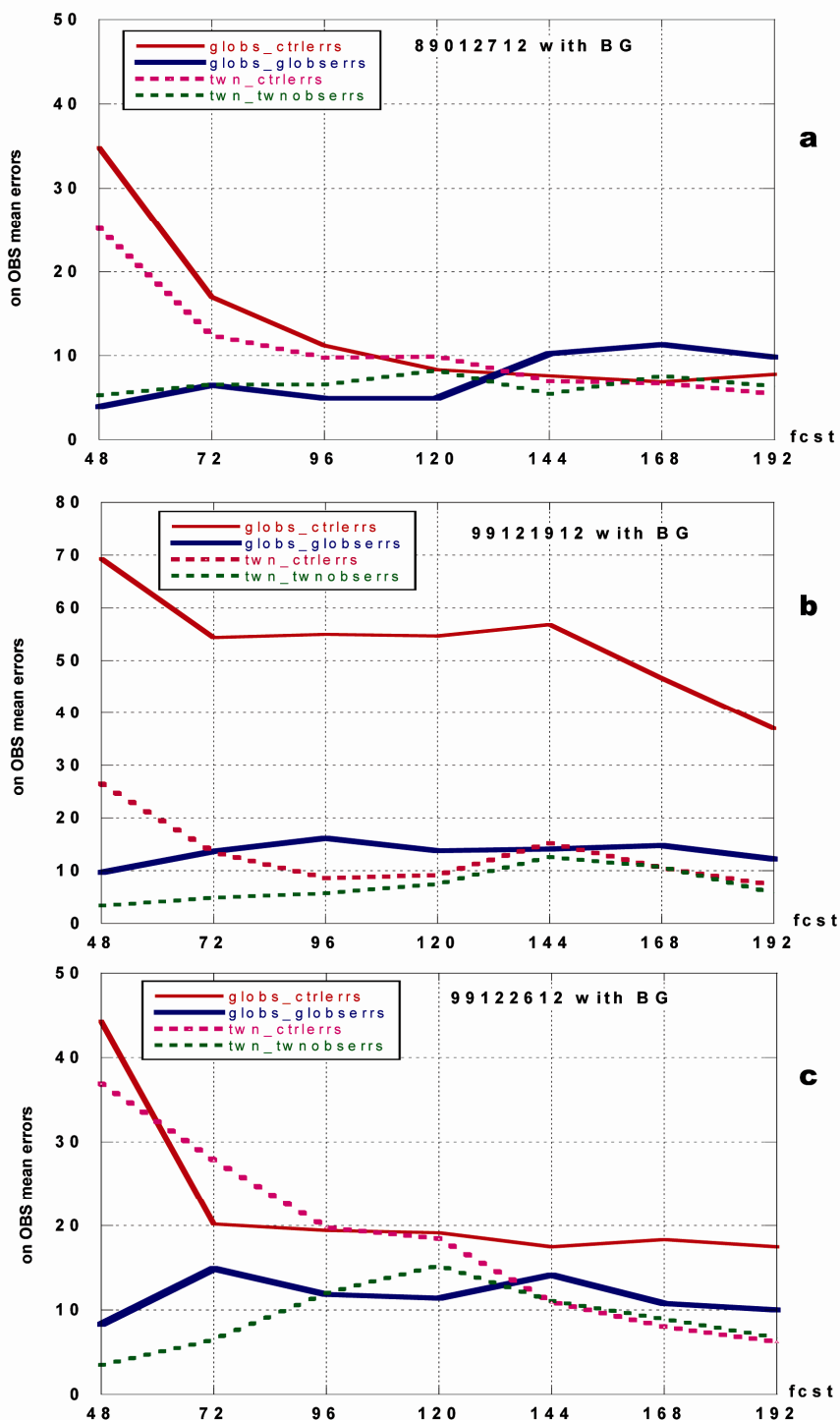


圖 8 加入背景場误差項後，在比對實驗的假想觀測網格上所計算的預報平均误差。紅色線是未經變分分析，在模擬全球觀測網格上所計算的預報平均误差；藍色線表經變分分析，在模擬全球觀測網格上所計算的預報平均误差；粉紅色虛線是未經變分分析，在模擬台灣東亞地區觀測網格上所計算的預報平均误差；綠色虛線表經變分分析，在模擬台灣東亞地區觀測網格上所計算的預報平均误差。

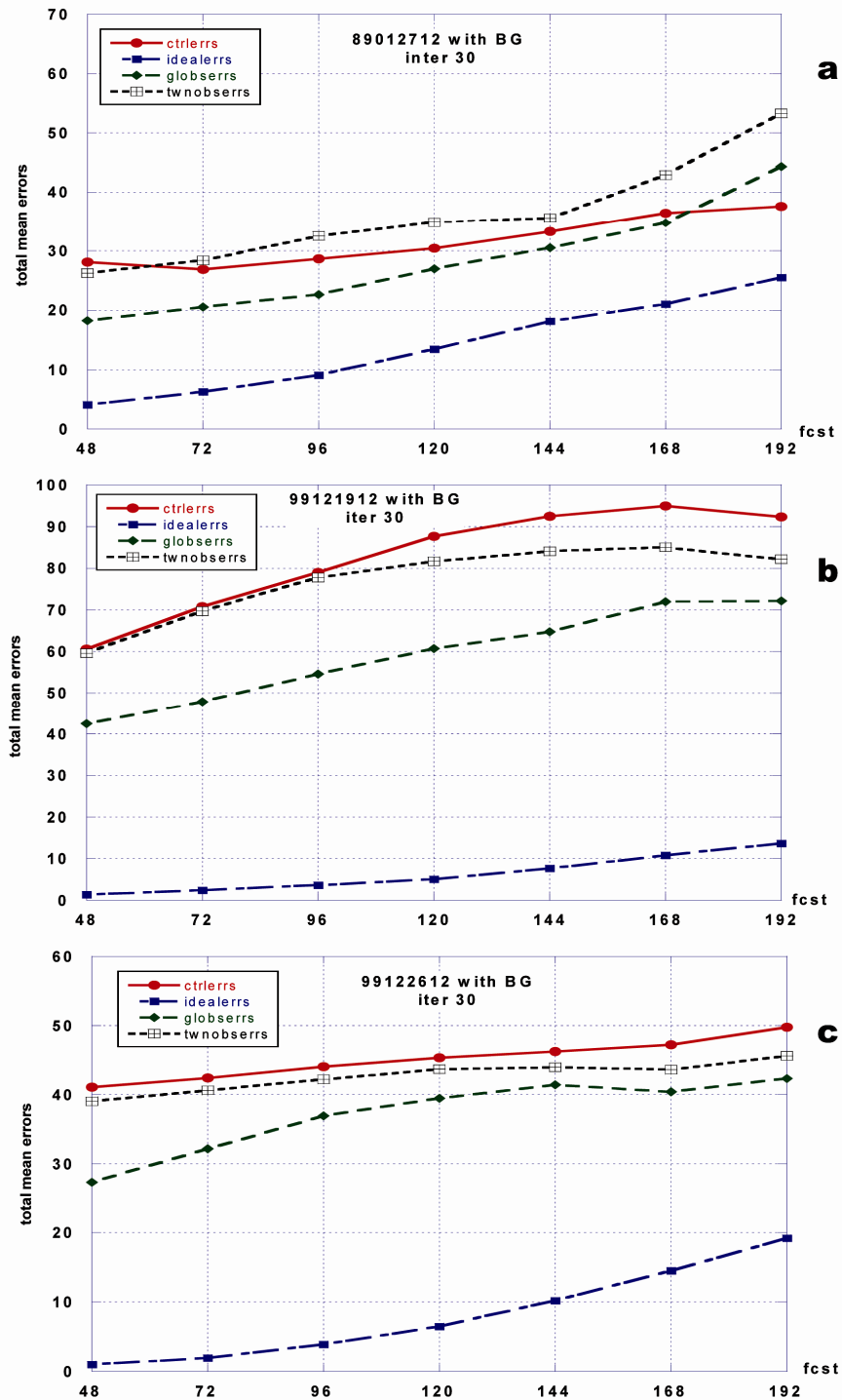


圖 9 加入背景場誤差項，經過疊代 30 次的極小化運算，三個個案（圖 a: 89012712; 圖 b: 99121912; 圖 c: 99122612）經 3 個比對實驗的預報平均誤差，分別是：紅色線表未經變分分析的預報平均誤差；藍色線表經全部分析網格變分分析後的預報平均誤差；綠色線表經模擬全球觀測變分分析後的預報平均誤差；黑色線是模擬台灣東亞地區密集觀測後變分分析的預報平均誤差。

析的誤差值還大。對於黑色線，台灣東亞地區有密集觀測的實驗來說，得到這樣的結果並不意外，因為所加入的模擬觀測點只有在台灣東亞地區，所加入的資訊太偏重於某一地區，對於用全球的平均誤差來比較，當然不見得會比較好。

圖 6 中的綠色線，即模擬全球觀測分佈的實驗，會是將來進行四維變分同化工作時最相似的狀況，但是其改進結果不若預期，分析其原因發現，雖然加入了全球分佈的模擬觀測資料，但是變分極小化的過程只負責把觀測點上的目標函數值變小，儘管所得新初始場和全部分析網格所做的理想比對實驗的初始場很接近（未圖示），但不儘然使所得預報軌跡與控制組接近，達到降低平均誤差的結果。所以，必須將背景場誤差項加入目標函數中，使得變分調整過程中，有一個如同模式變數的分析網格解析度的背景場約束條件，產生一個新的初始場，避免造成在觀測點附近也許有很好的調整，但在沒有觀測點的分析網格附近做出偏離背景場太多的調整。

圖 7 是在目標函數中加入背景場誤差後，經變分極小化程序所得初始場的預報軌跡。圖中的線型定義如同圖 6，如果觀察綠色線，模擬全球觀測分佈實驗的平均誤差要比沒有背景場的實驗小（圖 6 中的綠色線），僅在 89012712 這個個案預報的 192 小時處與未做變分分析的預報誤差相重合外，幾乎全部落在紅色線的下方，對於 89012712 與 99122612 這兩個個案而言，顯著的改進了 120 小時以後的預報誤差。對於圖 7 的藍色線，全部分析網格的理想化實驗而言，與圖 6 的藍色線是幾乎相同的。這點合乎直觀，當資料的提供充分，不需要背景場的作用。對於黑色線，台灣東亞地區有密集觀測的實驗也有一些改進，除 89012712 這個個案外，相較於紅色線而言，平均誤差下降了一些。

以模擬全球觀測分佈的變分實驗（圖 7 綠色線），在加入背景場誤差後的新初始場預報，以 120 小時預報平均誤差而言，大約的降低比率（（未做變分分析平均誤差－經變分分析平均誤差）/未做變分分析平均誤差）是：

89012712	99121912	99122612
$(30.44-25.44)/30.44=0.16$	$(87.67-62.08)/87.67=0.29$	$(45.37-39.87)/45.37=0.12$

根據張等（2001）的估算，這個縮減比率可以是氣象局與 NCEP 預報中 120 小時平均誤差的差距。附帶一提的是：張的文獻中所用的平均誤差估算只用到 500 hPa 的高度場，而本文所用的是所有模式變數，所有層數的平均誤差，所做的標準較嚴格；但是張的文章中統計了 2~5 年的平均誤差表現，所得結果較為客觀。此處，僅就平均誤差下降的比率數量級提供一個參考的比較，真

正的變分分析結果仍須於未來用真實資料，統計一段長時間後的表現為準。

如果只看假想觀測場網格上的平均誤差值（實驗 2, 3 中所取區域的定義），將發現變分極小化程序已經使得平均誤差值降低。於圖 8 中是針對假想觀測場網格上的平均誤差值所做的統計。在圖中紅色線是未經變分作用的實驗組預報軌跡與控制組預報軌跡在模擬全球觀測場網格上的平

均誤差值，藍色粗線為按實驗 2，加入模擬全球觀測資料，經變分分析所得實驗組預報軌跡與控制組預報軌跡在模擬全球觀測場網格上的平均誤差值，也就是圖 7 中的綠色線在假想觀測網格上的值。藍色線的值，大都小於紅色線的值，僅有 89012712 在 120 小時後的藍色線值略大於紅色線。同樣的情形，對於台灣東亞地區的平均誤差值計算，未經變分分析的誤差值（粉紅色虛線）在預報 120 小時內大於經變分分析的誤差值（綠色虛線），即使在預報 120 小時後，綠色線也沒有超過紅色線太多。這顯示變分極小化程序已經發揮作用，將假想觀測場網格上實驗組的預報軌跡朝假想觀測場網格上的控制組接近。

四、結論與討論

比對實驗為學界用來驗證四維變分系統的前置工作，實驗的過程為在已知某特定初始場的模式預報軌跡為假定觀測軌跡的前提下，必須能將一不同初始場的模式預報軌跡導向此特定初始場的演變軌跡。比對實驗的精神在於：以特定初始場演變的軌跡充當觀測軌跡，先去除模式預報誤差與觀測場分析誤差的部分，而僅就四維變分過程的正確與否來做驗證，如果實驗結果不能朝已知的控制組演變軌跡逼近，則此四維變分系統存在有錯誤，不能繼續進行四維變分資料同化的工作，必須先將其中的錯誤清除；反之，如果實驗結果朝控制組的演變逼近，這可以說此四維變分系統已經具有雛形，可應付隨後四維變分資料同化的工作。

雖然是進行比對實驗，但可將加入的假想觀測資料做一些變化，模擬真實同化工作時常遇到的情形，所以進行了 3 種不同的比對實驗。1.所

有分析網格的理想化比對實驗，意即將所有的分析網格上的假想觀測資料全部加入；2.模擬全球觀測場分佈的實驗，將假想的觀測資料在水平和垂直上去掉若干，以不規則分佈的方式呈現觀測資料；3.模擬台灣東亞地區有密集觀測的實驗，將假想的觀測資料只放在台灣東亞地區附近。

所選的三個個案都是全球作業中心和氣象局模式預報失敗的例子，1989 年的個案，事後分析因為是大氣基本流場中的內部動力不穩定作用，造成大氣流場迅速轉變。這個個案刺激了歐洲中心運用伴隨模式求算奇異向量執行系集預報（Tibaldi and Molteni, 1990），更帶動隨後變分法（1D-VAR, 3D-VAR, 4D-VAR）求算最佳化初始場的發展。至於本研究所選的其餘二個個案，由於也有類似大氣基本環流迅速改變的特性，初步分析顯示對擾動生成有很大幫助，相當值得一併研究。

從所有分析網格的理想化比對實驗結果來說，所得之最佳初始場朝假想正確的初始場逼近，其差距明顯比原初始場與假想正確初始場的差距為小（表 1, 2 和圖 5）。如果由所有模式預報變數經變數正則化分析所得之目標函數來看，整個差距較原先的誤差小了 1~2 個數量級（圖 3 和 4）。這顯示氣象局已經可以朝下一階段的工作進行，即將此方法代入作業模式做平行測試，檢驗其影響及效能。可以預期的是作業模式的過程遠較簡化模式複雜，由於伴隨模式由簡化模式導出，儘管簡化模式的控制變數與作業模式相同，但對於調整複雜模式的初始場勢必較為困難。再者，選定多長的時間做變分調整也待深入研究，以歐洲中心為例是選擇 12 小時（-6 ~ +6 小時）的窗區做調整。而下一階段的工作就會是結合氣

象局的 1D-VAR 和 3D-VAR (Chang et al., 2000; 2003) 做真正的四維變分資料同化。

背景場誤差項的作用在模擬全球觀測資料分佈或是台灣東亞地區密集觀測的預報誤差實驗中有很大的作用，減小了預報 120 小時後的平均誤差值。它給定了一個約束的條件，使得沒有觀測資料的分析網格上可以有調整的依據，不至在沒有觀測資料的地方做出不合理的調整。當然，對於所有分析網格上都有假想觀測資料的理想化比對實驗來說，背景場誤差項的作用就不明顯。至於背景場誤差的估計與計算一直是作業中心的研究課題，而且必須用自己作業中心數值模式的背景場誤差才算合理，常需要累積一段時間模式預報的統計性質才能估算，其相關學理的探討可見曾忠一（2006）書中的討論。

從全球模式的角度，除了理想化的比對實驗外，加入模擬全球觀測場分佈的資料將有效降低模式的預報平均誤差，而相對的，台灣東亞區的觀測資料在降低模式預報平均誤差的效果並不好。這也許是因為所計算的平均誤差是在全球分析網格上進行，而區域的預報改進不必然保證整體區域都變好。另外一方面，對於在實驗區域網格上所做的平均誤差而言，不論是模擬全球觀測場分佈的實驗或是台灣東亞區的觀測資料的實驗，其誤差值都明顯下降，這表示變分程序已經作用，縮小和控制組假想觀測場的差距。

以變分極小化作業效率的角度來看，到底應該執行多少次的疊代運算求得最佳初始場？目前來說每個作業中心都不相同，歐洲中心目前執行 50 次，而學界的運算如 NCAR 的 MM5 研究版是 100 次，而從本研究比對實驗的結果及氣象局電

腦資源來看，30 次似乎是不錯的選擇，這是因為目標函數已經下降了 1 個數量級，而預報誤差值的分析也支持 30 次的運算和 100 次的運算相同（圖 9）。在圖 9 中，三個個案、3 個實驗，加入背景場誤差，但是只疊代 30 次的比對實驗，以圖中的綠色線和黑色線的大小而言，幾乎等同圖 7 的形式，當然對理想化的實驗，圖中的藍色線部份，要比圖 7 有較大的平均誤差，但是其誤差值還是 3 個實驗中最小的。

如何定義目標函數，亦即那一個區域、那一些變數、或在那一個高度是另一個關注的焦點。因為不同的區域定義、不同變數，透過變分極小化程序得到的不同初始場，可以形成另外一組系集，提供系集預報另外的思考方向。另一方面，由作者初步的測試來說，氣象局模式上邊界的五層處理並不好，似乎在以後作變分分析或系集預報的工作時，應該將這五層摒棄在目標函數的計算範圍外。

在垂直的分層上，本研究為節省計算時間訂為 18 層，但隨後的實際測試與未來作業將完全承襲高解度模式的分層，意即氣象局目前的垂直解析度為 30 層，則伴隨模式變分系統的解析度也是 30 層。未來氣象局模式的垂直解析度分層為採用彈性的混合形式（如 sigma-theta, sigma-theta-pressure 等），而伴隨模式也會相應變動。至於四維變分同化的架構，未來將仿照歐洲中心朝向具有混合高、低解析度（高解析度模式做正向控制預報，低解析度進行同化迴圈計算）及增量分析（increment）調整的同化架構進行（Klinker et al., 2000; Mahfouf and Rabier, 2000; Rabier et al., 2000）。

致 謝

非常感謝兩位評審的寶貴意見，指出文中許多錯誤，特此致謝。研究期間謝謝周鑑本、張庭槐、陳雯美的討論與支持。謝謝張忍成老師的數值變分團隊陳御群、朱良斌的協助。謝謝已經不再做變分工作的顧欣怡在簡化版 Kuo 氏積雲參數化伴隨模式的幫助。

參考文獻

- 張庭槐、曾建翰、張忍成，2001：中央氣象局系集預報系統。氣象學報， **43**， 1-11。
- 曾忠一，2006：大氣科學中的反問題：反演、分析與同化。國立編譯館，1288 頁。
- 曾建翰、張忍成、林和，1997：伴隨理論在四維變分及其在氣象中之應用。大氣科學， **25**， 307-336。
- 曾建翰，2001：中央氣象局全球模式四維變分實驗，交通部中央氣象局研究及發展專題，29 頁。
- 曾建翰、張庭槐、張忍成，2005：中央氣象局四維變分預報系集系統。天氣分析與預報暨海象測報研討會論文彙編，142-146。
- Buckley, A. G. and A. Lenir, 1983: QN-like variable storage conjugate gradients. *Mathematical Programming*, **27**, 155-175.
- Chang, J.-C., C.-H. Tseng, and Y.-C. Chen, 2003: Some preliminary results on variational data assimilation of GPS/MET bending-angle profiles. In *2003 ROCSAT-3/COSMIC Science Workshop*, Sept. **8**, 2003, Taipei, Taiwan, 43-51.
- Chang, J.-C. J., C.-H. J. Tseng, H.-L. Liu, X. Zou, and M.-D. Cheng, 2000: Comparison of CWB's global analysis using a GPS ray-tracing operator with GPS/MET observation. *COSMIC International Workshop*, Sept. 27-29, Wanli, Taipei, Taiwan.
- Gill, P. E. and W. Murray, 1979: Conjugate-gradient methods for large-scale nonlinear optimization. *Tech. Report SOL 79-15*, System optimization laboratory, Department of operations research, Stanford University, Stanford, CA.
- Gill, P. E., W. Murray, and M. H. Wright, 1981: *Practical optimization*. Academic Press, 401pp
- Klinker, E., F. Rabier, G. Kelly and J.-F. Mahfouf, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. III: Experimental results and diagnostics with operational configuration. *Q. J. R. Meteorol. Soc.*, **126**, 1191-1215.
- Liu, D. C. and J. Nocedal, 1989: On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, **45**, 503-528.
- Mahfouf, J.-F. and F. Rabier, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. II: Experimental results with improved physics. *Q. J. R. Meteorol. Soc.*, **126**, 1171-1190.
- Rabier, F., H. Jarvinen, E. Klinker, J. F. Mahfouf and A. Simmons, 2000: The ECMWF operational implementation of

- four-dimensional variational assimilation. I: Experimental results with simplified physics. *Q. J. R. Meteorol. Soc.*, **126**, 649-667.
- Shanno, D. F. and K. H. Phua, 1980: Remark on algorithm 500 — a variable method subroutine for unconstrained nonlinear minimization. *ACM trans. Math. Software*, **6**, 618-622.
- Talagrand, O. and P. Courtier, 1991: The use of adjoint equations in numerical modeling of the atmospheric circulation. In: (A. Griewank, and G. Corlisse, editor) *Automatic differentiation of algorithms: theory, implementation and application*, pp169-180.
- Tibaldi, S. and F. Molteni, 1990: On the operational predictability of blocking. *Tellus*, **42A**, 343-365.
- Zou, X., F. Vandenberghe, M. Pondeca and Y. -H. Kuo, 1998: *Introduction to adjoint techniques and the MM5 adjoint modeling system*. NCAR/TN-435-STR, National Center for Atmospheric Research, 110pp.

The 4D-Var Twin Experiments using Central Weather Bureau Global Model

John Chien-Han Tseng¹ Jen-Chen Joseph Chang²

¹ **Research and Development Center, Central Weather Bureau**

² **Department of Atmospheric Science, Chinese Culture University**

(Manuscript received 5 June 2006; in final form 10 November 2006)

ABSTRACT

The purpose of this study is the implementation of the 4D-Var processes without adding true observation data. We use one initial condition integrated by numerical model to produce one forecast track to be hypothetical observation track, and use another initial condition within 12~24 h in comparison with previous initial field to produce other forecast track, so obviously the difference between these two forecast tracks can be defined as a cost function. The minimum of this cost function can be solved by the 4D-Var processes; i.e. we can get one optimal initial condition and its forecast will be close to the hypothetical observation track. We use the adjoint model to get the gradient of cost function and use minimization procedure, L-BFGS method, to get the minimum value of the cost function. In this study, we also try to simulate the irregular distributed observational data and somewhere like in Taiwan East Asia having intensive observation data for testing 4D-Var processes. The whole processes mainly follow the current Meteorological operation centers or foreign research units procedures, and they can be the foundation of the future 4D-Var operation at Central Weather Bureau.

From the 3 cases and 3 twin experiments, the optimal initial fields tend toward the hypothetical initial fields. More specifically, the differences between the optimal initial fields and the hypothetical initial fields are far less than the differences between the original initial fields and the hypothetical initial fields. The ideal twin experiment on entire analytical grids can get the smallest differences between optimal initial fields and the hypothetical initial fields, and this best results could be a reference theoretical value. Comparing global model forecasts, the results of twin experiment on global irregular observation data are better than the results using Taiwan East Asia intensive observation data. The 120 h forecast mean errors of the new initial fields by using global irregular observation data are smaller 12%~29% than the 120 h forecast mean errors of the initial fields which were not accommodated by 4D-Var processes. For the efficiency, the 30 iterations in minimization procedure can reduce the cost function about one order. Furthermore, if we check the 8 days forecasts, the 30 iterations results are similar to 100 iterations results. So, the 30 iterations can be the reference value for future Central Weather Bureau 4D-Var operation.

Key words: Adjoint model, Minimization, L-BFGS