

**DAO Office Note 99-02**

**Office Note Series on  
Global Modeling and Data Assimilation**

Robert Atlas, Head  
*Data Assimilation Office*  
*Goddard Space Flight Center*  
*Greenbelt, Maryland*

**Technical Remarks on Smoother Algorithms**

Yanqiu Zhu, Ricardo Todling,  
*General Sciences Corporation, Data Assimilation Office,*

Stephen E. Cohn  
*Data Assimilation Office, Goddard Laboratory for Atmospheres*



**Goddard Space Flight Center**  
Greenbelt, Maryland 20771  
**June 1999**

## Abstract

Among the currently existing data assimilation algorithms, 4D variational data assimilation (4D-VAR), 4D-PSAS, fixed-lag Kalman smoother (FLKS), and representer method as well as Kalman smoother belong to the smoother category. In this Office Note, the formulations of these smoothing algorithms are discussed from the Bayesian point of view. Their relationships are further explored for linear dynamics in the context of fixed-interval smoothing. The implementation approaches and computational aspects of the smoothing algorithms are also discussed and intercompared for the purpose of retrospective data assimilation. Finally, the extensions of the algorithms to nonlinear dynamics are presented.

# Contents

<b>Abstract</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Formulations of smoother algorithms in probabilistic framework</b>	<b>2</b>
2.1 FLKS formulation . . . . .	3
2.2 3D-PSAS-like formulation of FLKS . . . . .	7
2.2.1 Linear imperfect model . . . . .	8
2.2.2 Linear perfect model . . . . .	9
2.3 4D-VAR and 4D-PSAS . . . . .	10
2.3.1 Linear imperfect model . . . . .	10
2.3.2 Linear perfect model . . . . .	12
<b>3 The relationships among smoother algorithms</b>	<b>13</b>
3.1 4D-VAR . . . . .	14
3.2 FLKS . . . . .	15
3.3 4D-PSAS . . . . .	17
3.4 Representer method . . . . .	18
3.5 Kalman smoother . . . . .	19
3.6 Fixed-point smoother perspective . . . . .	22
<b>4 The computational aspects of 4D-PSAS and FLKS algorithms</b>	<b>24</b>
4.1 Linear perfect model . . . . .	25
4.2 Linear imperfect model . . . . .	27
<b>5 Extensions to nonlinear model</b>	<b>31</b>
5.1 FLKS algorithm . . . . .	32
5.2 4D-PSAS algorithm . . . . .	35
<b>Acknowledgments</b>	<b>38</b>
<b>Appendix A</b>	<b>39</b>
<b>Appendix B</b>	<b>39</b>
<b>References</b>	<b>45</b>

# 1 Introduction

Data assimilation has been applied successfully in providing the initial conditions for the numerical weather prediction via the assimilation of the current and past observation data along with the model dynamics. With the availability of the ever increasing observational data and the need of the research community, the long-term analysis of data assimilation is also much needed for medium and long-term weather forecast and climate studies. For this purpose, a delay in the production of the analysis is permitted, then one could conceive of more observations becoming available during the delay interval and being used in producing the analysis. Thus the so-called retrospective analysis proposed by Cohn et al. (1994), which incorporates future observation data, as well as current and past observation data, would serve the purpose. Because more observations are used in producing the retrospective analysis, it is expected to be more accurate and complete than the filter solution.

The problem using both future and current as well as past observations along with model dynamics is termed smoothing problem in estimation theory. Presently, several algorithms have been proposed and tested for the smoothing problems, among which the fixed-lag Kalman smoother (FLKS), 4D-VAR and 4D-PSAS have attracted more and more attention.

The FLKS algorithm was proposed by Cohn et al. (1994) as a means to perform the retrospective analysis, and numerical experiments using a two-dimensional linear shallow-water model were carried out to demonstrate the ability of the FLKS in improving the analysis quality. Further numerical experiments with suboptimal schemes were done by Todling et al. (1998) to reduce the computational burden. Their results indicated that retrospective data assimilation could be successful even when simple filtering schemes were used.

Both 4D-VAR and 4D-PSAS (Courtier, 1997) belong to fixed-interval smoother category, using future, current and past observations to generate the analysis inside a fixed interval (Cohn et al., 1994; Menard and Daley, 1996). However, both of them are currently mainly employed to provide the initial conditions for the numerical weather prediction, that is, they are practically used as a fixed-point smoother to produce the initial conditions by incorporating all of the observations within the fixed interval. They can also be considered as options for doing retrospective analysis. In order to obtain the retrospective analysis at each analysis point by incorporating the same amount of time levels of future observations, a moving 4D-VAR or 4D-PSAS has to be performed with fixed interval for each analysis point which acts as the initial point of the fixed interval.

Theoretically, all of these algorithms can be applied for the purpose of retrospective analysis. However, the relationships among these algorithms have not been fully explored yet, and the issue about the qualities of the analyses produced by different algorithms would necessitate further studies. In this study, our goal is to analyze different approaches (4D-VAR, 4D-PSAS, and FLKS) for doing retrospective data assimilation from a scientific and computational standpoint, and compare each in terms of its suitability to be the schemes used to do retrospective data assimilation within the GEOS DAS framework.

This note is organized as following. The formulations of FLKS, 4D-VAR and 4D-PSAS in the probabilistic framework are presented in section 2. The 3D-PSAS-like formulation of FLKS in

the GEOS DAS framework is also presented in this section. The relationships among FLKS, 4D-VAR and 4D-PSAS as well as their distinct characteristics are discussed in section 3. The representer method (Bennett, 1992; Bennett et al., 1996) and Kalman smoother (Evensen, 1997) are briefly described and compared with 4D-PSAS as well. Section 4 presents the comparisons of computational aspects of the the algorithms in the cases of linear perfect and imperfect model, respectively. Finally, the extension of the algorithms for linear case to nonlinear case is presented in section 5.

## 2 Formulations of smoother algorithms in probabilistic framework

In this section, the derivations of the FLKS, 4D-VAR and 4D-PSAS formula are presented in the probabilistic framework. Throughout this note, the notation of Cohn et al. (1994) is adopted.

Assuming the total numbers of analysis grids and observations at time  $t_k$  are  $n$  and  $p_k$ , respectively, and a forecast model is of the form

$$\mathbf{w}_{k|k-1}^f = \mathbf{A}_{k,k-1}(\mathbf{w}_{k-1|k-1}^a), \quad (1)$$

where the time indices refer to observation times, which also coincide with the filter analysis times. The  $n$ -vector  $\mathbf{w}_{k-1|k-1}^a$  is the filter analysis at previous observation time  $t_{k-1}$  using all observations up till the previous observation time  $t_{k-1}$ , the  $n$ -vector  $\mathbf{w}_{k|k-1}^f$  is the forecast at present observation time  $t_k$  and still uses only the observations up till the previous observation time  $t_{k-1}$ , the  $n \times n$  matrix  $\mathbf{A}_{k,k-1}$  denotes the discrete propagator between the two consecutive observation times  $t_{k-1}$  and  $t_k$ .

The discrete evolution equation for the unknown true state  $\mathbf{w}_k^t$  is given as

$$\mathbf{w}_k^t = \mathbf{A}_{k,k-1}(\mathbf{w}_{k-1}^t) + \mathbf{b}_k, \quad (2)$$

where  $\mathbf{b}_k$  represents the model error as a stochastic process, white in time with mean zero and covariance matrix  $\mathbf{Q}_k$ :

$$\mathcal{E}\{\mathbf{b}_k\} = \mathbf{0}, \quad (3a)$$

$$\mathcal{E}\{\mathbf{b}_k(\mathbf{b}_{k'})^T\} = \mathbf{Q}_k \delta_{kk'}, \quad (3b)$$

where  $\mathcal{E}\{\}$  represents expectation, the superscript  $T$  denotes the transpose, and  $\delta_{kk'}$  is the Kronecker delta. Since the state  $\mathbf{w}_k^t$  is given by the stochastic-dynamic model (2), it has a probability distribution function. All distribution functions are assumed to be differentiable, thus  $\mathbf{w}_k^t$  has a probability density function  $p(\mathbf{w}_k^t)$ . Here, we notice that  $\{\mathbf{w}_k^t\}$  is a Markov process. In other words, if  $k_1 < k_2 < \dots < k_m < k$ , the probability density of  $\mathbf{w}_k^t$  conditioned on  $\mathbf{w}_{k_1}^t, \mathbf{w}_{k_2}^t, \dots, \mathbf{w}_{k_m}^t$  is simply the probability density of  $\mathbf{w}_k^t$  conditioned on  $\mathbf{w}_{k_m}^t$ .

The discrete observation model for the  $p_k$ -vector  $\mathbf{w}_k^o$  is written as

$$\mathbf{w}_k^o = \mathbf{h}_k(\mathbf{w}_k^t) + \mathbf{b}_k^o. \quad (4)$$

For simplicity, we assume the observations are available at every time  $t_k$ , and the  $p_k \times n$  matrix  $\mathbf{h}_k$  denotes the observation operator. Also, we take the observation error  $\mathbf{b}_k^o$  to be white in time, with mean zero and covariance matrix  $\mathbf{R}_k$ :

$$\mathcal{E}\{\mathbf{b}_k^o\} = \mathbf{0}, \quad (5a)$$

$$\mathcal{E}\{\mathbf{b}_k^o(\mathbf{b}_{k'}^o)^T\} = \mathbf{R}_k \delta_{kk'}. \quad (5b)$$

As we know, for various data assimilation problems, the conditional probability density  $p(\mathbf{w}_k^t | \mathbf{W}_L^o)$  constitutes the complete solution. Here,  $\mathbf{W}_L^o$  represents the set of realizations of all observations available up till to some time  $t_L$ :

$$\mathbf{W}_L^o \equiv \{\mathbf{w}_0^o, \mathbf{w}_1^o, \dots, \mathbf{w}_L^o\}. \quad (6)$$

The probability density  $p(\mathbf{w}_k^t | \mathbf{W}_k^o)$  yields the filtering solutions at times  $t_k$ ,  $k = 1, 2, \dots$ ,  $p(\mathbf{w}_{k-l}^t, \mathbf{w}_{k-l+1}^t, \dots, \mathbf{w}_k^t | \mathbf{W}_k^o)$  with fixed  $l$  gives the solutions for the fixed-lag smoothing problem at times  $t_{k-l}, t_{k-l+1}, \dots, t_k$  (Cohn et al., 1994, 1997; Anderson and Moore, 1979), and  $p(\mathbf{w}_0^t, \mathbf{w}_1^t, \dots, \mathbf{w}_N^t | \mathbf{W}_N^o)$  with fixed  $N$  yields the solutions for fixed-interval smoothing problem at times  $t_0, t_1, \dots, t_N$ .

## 2.1 FLKS formulation

The conditional probability density for the fixed-lag Kalman smoother can be obtained as

$$\begin{aligned} & p(\mathbf{w}_k^t, \mathbf{w}_{k-1}^t, \mathbf{w}_{k-2}^t, \dots, \mathbf{w}_{k-l}^t | \mathbf{W}_k^o) \\ &= \frac{1}{p(\mathbf{W}_k^o)} p(\mathbf{w}_k^t, \mathbf{w}_{k-1}^t, \mathbf{w}_{k-2}^t, \dots, \mathbf{w}_{k-l}^t, \mathbf{W}_k^o) \\ &= \frac{1}{p(\mathbf{W}_k^o)} p(\mathbf{w}_k^t, \mathbf{w}_{k-1}^t, \mathbf{w}_{k-2}^t, \dots, \mathbf{w}_{k-l}^t, \mathbf{W}_{k-1}^o, \mathbf{w}_k^o) \\ &= \frac{1}{p(\mathbf{W}_k^o)} p(\mathbf{w}_k^o | \mathbf{w}_k^t, \mathbf{w}_{k-1}^t, \mathbf{w}_{k-2}^t, \dots, \mathbf{w}_{k-l}^t, \mathbf{W}_{k-1}^o) \\ &\quad \times p(\mathbf{w}_k^t, \mathbf{w}_{k-1}^t, \mathbf{w}_{k-2}^t, \dots, \mathbf{w}_{k-l}^t, \mathbf{W}_{k-1}^o) \\ &= \frac{p(\mathbf{W}_{k-1}^o)}{p(\mathbf{W}_k^o)} p(\mathbf{w}_k^o | \mathbf{w}_k^t, \mathbf{w}_{k-1}^t, \mathbf{w}_{k-2}^t, \dots, \mathbf{w}_{k-l}^t) p(\mathbf{W}^t | \mathbf{W}_{k-1}^o), \end{aligned} \quad (7)$$

where  $p(\mathbf{W}^t | \mathbf{W}_{k-1}^o)$  represents the *a priori* information on model state variable with

$$\mathbf{W}^t = \begin{pmatrix} \mathbf{w}_k^t \\ \mathbf{w}_{k-1}^t \\ \vdots \\ \mathbf{w}_{k-l}^t \end{pmatrix}. \quad (8)$$

It is shown here that the *a priori* estimation of the model state variable is the current best estimate, which is produced by using observations up till time  $t_{k-1}$ , and the observations are assimilated sequentially, i.e., one time level of observations at a time. In this way, the analysis is

updated once a new time level of observations are available, instead of waiting for the availability of the entire observations within the fixed lag like fixed interval smoother does.

We assume  $\mathbf{b}_k$ ,  $\mathbf{b}_k^o$ , and the probability density representing the *a priori* information on model state variable are Gaussian distributed, defining

$$\bar{\mathbf{W}} = \mathcal{E} \left\{ \mathbf{W}^t | \mathbf{W}_{k-1}^o \right\} = \begin{pmatrix} \mathcal{E} \left\{ \mathbf{w}_k^t | \mathbf{W}_{k-1}^o \right\} \\ \mathcal{E} \left\{ \mathbf{w}_{k-1}^t | \mathbf{W}_{k-1}^o \right\} \\ \vdots \\ \mathcal{E} \left\{ \mathbf{w}_{k-l}^t | \mathbf{W}_{k-1}^o \right\} \end{pmatrix} = \begin{pmatrix} \mathbf{w}_{k|k-1}^f \\ \mathbf{w}_{k-1|k-1}^a \\ \vdots \\ \mathbf{w}_{k-l|k-1}^a \end{pmatrix} \quad (9)$$

and error covariance matrix

$$\begin{aligned} \mathbf{P}^t &= \mathcal{E} \left\{ \begin{pmatrix} \mathbf{e}_{k|k-1}^f \\ \mathbf{e}_{k-1|k-1}^a \\ \vdots \\ \mathbf{e}_{k-l|k-1}^a \end{pmatrix} \begin{pmatrix} \mathbf{e}_{k|k-1}^f \\ \mathbf{e}_{k-1|k-1}^a \\ \vdots \\ \mathbf{e}_{k-l|k-1}^a \end{pmatrix}^T \middle| \mathbf{W}_{k-1}^o \right\} \\ &= \begin{pmatrix} \mathbf{P}_{k|k-1}^f & \mathbf{P}_{k,k-1|k-1}^{fa} & \mathbf{P}_{k,k-2|k-1}^{fa} & \cdots & \mathbf{P}_{k,k-l|k-1}^{fa} \\ \mathbf{P}_{k-1,k|k-1}^{af} & \mathbf{P}_{k-1|k-1}^a & \mathbf{P}_{k-1,k-2|k-1}^{aa} & \cdots & \mathbf{P}_{k-1,k-l|k-1}^{aa} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \mathbf{P}_{k-l,k|k-1}^{af} & \mathbf{P}_{k-l,k-1|k-1}^{aa} & \mathbf{P}_{k-l,k-2|k-1}^{aa} & \cdots & \mathbf{P}_{k-l|k-1}^a \end{pmatrix}, \end{aligned} \quad (10)$$

where

$$\mathbf{e}_{k|k-1}^f = \mathbf{w}_{k|k-1}^f - \mathbf{w}_k^t, \quad \mathbf{e}_{k-i|k-1}^a = \mathbf{w}_{k-i|k-1}^a - \mathbf{w}_{k-i}^t$$

for  $i = 1, 2, \dots, l$ , then we have

$$p(\mathbf{W}^t | \mathbf{W}_{k-1}^o) \propto \exp \left\{ -\frac{1}{2} (\mathbf{W}^t - \bar{\mathbf{W}})^T (\mathbf{P}^t)^{-1} (\mathbf{W}^t - \bar{\mathbf{W}}) \right\}. \quad (11)$$

Also since  $p(\mathbf{w}_k^o | \mathbf{w}_k^t, \mathbf{w}_{k-1}^t, \dots, \mathbf{w}_{k-l}^t)$  can be described as

$$p(\mathbf{w}_k^o | \mathbf{w}_k^t, \mathbf{w}_{k-1}^t, \dots, \mathbf{w}_{k-l}^t) \propto \exp \left\{ -\frac{1}{2} (\mathbf{w}_k^o - \mathbf{h}_k(\mathbf{w}_k^t))^T \mathbf{R}_k^{-1} (\mathbf{w}_k^o - \mathbf{h}_k(\mathbf{w}_k^t)) \right\}, \quad (12)$$

the conditional probability density function  $p(\mathbf{w}_k^t, \mathbf{w}_{k-1}^t, \mathbf{w}_{k-2}^t, \dots, \mathbf{w}_{k-l}^t | \mathbf{W}_k^o)$  can then be expressed as

$$p(\mathbf{w}_k^t, \mathbf{w}_{k-1}^t, \mathbf{w}_{k-2}^t, \dots, \mathbf{w}_{k-l}^t | \mathbf{W}_k^o) = \text{const} \cdot \exp(-\mathcal{J}_{FLKS}), \quad (13)$$

where the **cost function** of FLKS  $\mathcal{J}_{FLKS}$  is given as

$$\begin{aligned} \mathcal{J}_{FLKS} &= \frac{1}{2} (\mathbf{W}^t - \bar{\mathbf{W}})^T (\mathbf{P}^t)^{-1} (\mathbf{W}^t - \bar{\mathbf{W}}) \\ &\quad + \frac{1}{2} (\mathbf{w}_k^o - \mathbf{h}_k(\mathbf{w}_k^t))^T \mathbf{R}_k^{-1} (\mathbf{w}_k^o - \mathbf{h}_k(\mathbf{w}_k^t)). \end{aligned} \quad (14)$$

For the linear case (both linear model and linear observational operator), defining  $\mathcal{H} = (\mathbf{h}_k, \mathbf{0}, \dots, \mathbf{0})$ , the cost function  $\mathcal{J}_{FLKS}$  is of the form

$$\begin{aligned} \mathcal{J}_{FLKS} &= \frac{1}{2}(\mathbf{W}^t - \bar{\mathbf{W}})^T (\mathbf{P}^t)^{-1} (\mathbf{W}^t - \bar{\mathbf{W}}) \\ &\quad + \frac{1}{2}(\mathbf{w}_k^o - \mathcal{H}\mathbf{W}^t)^T (\mathbf{R}_k)^{-1} (\mathbf{w}_k^o - \mathcal{H}\mathbf{W}^t). \end{aligned} \quad (15)$$

If we define

$$\begin{aligned} \mathbf{W}^a &= [(\mathbf{P}^t)^{-1} + \mathcal{H}^T \mathbf{R}_k^{-1} \mathcal{H}]^{-1} [(\mathbf{P}^t)^{-1} \bar{\mathbf{W}} + \mathcal{H}^T \mathbf{R}_k^{-1} \mathbf{w}_k^o] \\ &= \bar{\mathbf{W}} + [(\mathbf{P}^t)^{-1} + \mathcal{H}^T \mathbf{R}_k^{-1} \mathcal{H}]^{-1} \mathcal{H}^T \mathbf{R}_k^{-1} (\mathbf{w}_k^o - \mathcal{H}\bar{\mathbf{W}}) \\ &= \bar{\mathbf{W}} + \mathbf{P}^t \mathcal{H}^T (\mathcal{H} \mathbf{P}^t \mathcal{H}^T + \mathbf{R}_k)^{-1} (\mathbf{w}_k^o - \mathcal{H}\bar{\mathbf{W}}) \\ &= \bar{\mathbf{W}} + \mathcal{K} (\mathbf{w}_k^o - \mathcal{H}\bar{\mathbf{W}}) \end{aligned} \quad (16)$$

and

$$\begin{aligned} \mathbf{P} &= [(\mathbf{P}^t)^{-1} + \mathcal{H}^T \mathbf{R}_k^{-1} \mathcal{H}]^{-1} \\ &= (\mathbf{I} - \mathcal{K} \mathcal{H}) \mathbf{P}^t, \end{aligned} \quad (17)$$

where

$$\mathcal{K} = \mathbf{P}^t \mathcal{H}^T (\mathcal{H} \mathbf{P}^t \mathcal{H}^T + \mathbf{R}_k)^{-1}, \quad (18)$$

then equation (15) becomes

$$\begin{aligned} \mathcal{J}_{FLKS} &= \frac{1}{2}(\mathbf{W}^t - \mathbf{W}^a)^T \mathbf{P}^{-1} (\mathbf{W}^t - \mathbf{W}^a) - (\mathbf{W}^a)^T \mathbf{P}^{-1} \mathbf{W}^a \\ &\quad + (\mathbf{w}_k^o)^T \mathbf{R}_k^{-1} \mathbf{w}_k^o + \bar{\mathbf{W}}^T (\mathbf{P}^t)^{-1} \bar{\mathbf{W}}. \end{aligned} \quad (19)$$

All the right-hand terms but the first are independent of  $\mathbf{W}^t$ , and can be absorbed in the constant factor of equation (13). This gives

$$p(\mathbf{w}_k^t, \mathbf{w}_{k-1}^t, \mathbf{w}_{k-2}^t, \dots, \mathbf{w}_{k-l}^t | \mathbf{W}_k^o) = \text{const} \cdot \exp(-\mathcal{J}') \quad (20)$$

where

$$\mathcal{J}' = \frac{1}{2}(\mathbf{W}^t - \mathbf{W}^a)^T \mathbf{P}^{-1} (\mathbf{W}^t - \mathbf{W}^a). \quad (21)$$

It is shown that, in the linear case, the a posteriori probability density is Gaussian. The center of this Gaussian is given by (16), and its covariance is given by (17). It is obvious that  $\mathbf{W}^a$  minimizes the cost function  $\mathcal{J}_{FLKS}$ , hence, it is also the maximum likelihood point.

Substituting equation (10) and the definition of  $\mathcal{H}$  into equation (18), we have

$$\mathcal{K} = \begin{pmatrix} \mathbf{P}_{k|k-1}^f \mathbf{h}_k^T \\ \mathbf{P}_{k-1,k|k-1}^{af} \mathbf{h}_k^T \\ \vdots \\ \mathbf{P}_{k-l,k|k-1}^{af} \mathbf{h}_k^T \end{pmatrix} (\mathbf{h}_k \mathbf{P}_{k|k-1}^f \mathbf{h}_k^T + \mathbf{R}_k)^{-1}. \quad (22)$$



Introducing

$$\mathbf{\Gamma}_k = \mathbf{h}_k \mathbf{P}_{k|k-1}^f \mathbf{h}_k^T + \mathbf{R}_k \quad (23)$$

and

$$\mathcal{K} = \begin{pmatrix} \mathcal{K}_{k|k} \\ \mathcal{K}_{k-1|k} \\ \vdots \\ \mathcal{K}_{k-l|k} \end{pmatrix}, \quad (24)$$

we can obtain the following gains of FLKS

$$\mathcal{K}_{k|k} = \mathbf{P}_{k|k-1}^f \mathbf{h}_k^T \mathbf{\Gamma}_k^{-1}, \quad (25a)$$

$$\mathcal{K}_{k-1|k} = \mathbf{P}_{k-1,k|k-1}^{af} \mathbf{h}_k^T \mathbf{\Gamma}_k^{-1}, \quad (25b)$$

$\vdots$

$$\mathcal{K}_{k-l|k} = \mathbf{P}_{k-l,k|k-1}^{af} \mathbf{h}_k^T \mathbf{\Gamma}_k^{-1}. \quad (25c)$$

Applying equations (24) and (9) into (16), and using

$$\mathbf{W}^a = \begin{pmatrix} \mathbf{w}_{k|k}^a \\ \mathbf{w}_{k-1|k}^a \\ \vdots \\ \mathbf{w}_{k-l|k}^a \end{pmatrix}, \quad (26)$$

we can see that the FLKS analysis equations are given as:

$$\mathbf{w}_{k|k}^a = \mathbf{w}_{k|k-1}^f + \mathcal{K}_{k|k} \mathbf{v}_k, \quad (27a)$$

$$\mathbf{w}_{k-1|k}^a = \mathbf{w}_{k-1|k-1}^a + \mathcal{K}_{k-1|k} \mathbf{v}_k, \quad (27b)$$

$\vdots$

$$\mathbf{w}_{k-l|k}^a = \mathbf{w}_{k-l|k-1}^a + \mathcal{K}_{k-l|k} \mathbf{v}_k, \quad (27c)$$

where  $\mathbf{v}_k$  is the innovation vector defined as

$$\mathbf{v}_k = \mathbf{w}_k^o - \mathbf{h}_k \mathbf{w}_{k|k-1}^f. \quad (28)$$

Applying the forecast step of the Kalman filter  $\mathbf{w}_{k|k-1}^f = \mathbf{A}_{k,k-1} \mathbf{w}_{k-1|k-1}^a$ ,  $\mathbf{v}_k$  can also be written as

$$\mathbf{v}_k = \mathbf{w}_k^o - \mathbf{h}_k \mathbf{A}_{k,k-1} \mathbf{w}_{k-1|k-1}^a. \quad (29)$$

Since the FLKS analysis error covariance is defined as

$$\mathbf{P} = \mathcal{E} \left\{ \begin{pmatrix} \mathbf{e}_{k|k}^a \\ \mathbf{e}_{k-1|k}^a \\ \vdots \\ \mathbf{e}_{k-l|k}^a \end{pmatrix} \begin{pmatrix} \mathbf{e}_{k|k}^a \\ \mathbf{e}_{k-1|k}^a \\ \vdots \\ \mathbf{e}_{k-l|k}^a \end{pmatrix}^T \middle| \mathbf{W}_k^o \right\}$$

$$= \begin{pmatrix} \mathbf{P}_{k|k}^a & \mathbf{P}_{k,k-1|k}^{aa} & \mathbf{P}_{k,k-2|k}^{aa} & \cdots & \mathbf{P}_{k,k-l|k}^{aa} \\ \mathbf{P}_{k-1,k|k}^{aa} & \mathbf{P}_{k-1|k}^a & \mathbf{P}_{k-1,k-2|k}^{aa} & \cdots & \mathbf{P}_{k-1,k-l|k}^{aa} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \mathbf{P}_{k-l,k|k}^{aa} & \mathbf{P}_{k-l,k-1|k}^{aa} & \mathbf{P}_{k-l,k-2|k}^{aa} & \cdots & \mathbf{P}_{k-l|k}^a \end{pmatrix}, \quad (30)$$

and also the analysis error covariance equation (17) can be expressed as

$$\mathbf{P} = \begin{pmatrix} \mathbf{I} - \mathcal{K}_{k|k} \mathbf{h}_k & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathcal{K}_{k-1|k} \mathbf{h}_k & \mathbf{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & \cdots & \vdots \\ -\mathcal{K}_{k-l|k} \mathbf{h}_k & \mathbf{0} & \cdots & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{P}_{k|k-1}^f & \mathbf{P}_{k,k-1|k-1}^{fa} & \mathbf{P}_{k,k-2|k-1}^{fa} & \cdots & \mathbf{P}_{k,k-l|k-1}^{fa} \\ \mathbf{P}_{k-1,k|k-1}^{af} & \mathbf{P}_{k-1|k-1}^a & \mathbf{P}_{k-1,k-2|k-1}^{aa} & \cdots & \mathbf{P}_{k-1,k-l|k-1}^{aa} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \mathbf{P}_{k-l,k|k-1}^{af} & \mathbf{P}_{k-l,k-1|k-1}^{aa} & \mathbf{P}_{k-l,k-2|k-1}^{aa} & \cdots & \mathbf{P}_{k-l|k-1}^a \end{pmatrix}, \quad (31)$$

equating equations (31) and (30), we get the equations for error covariances

$$\mathbf{P}_{k|k}^a = (\mathbf{I} - \mathcal{K}_{k|k} \mathbf{h}_k) \mathbf{P}_{k|k-1}^f, \quad (32a)$$

$$\mathbf{P}_{k-l|k}^a = \mathbf{P}_{k-l|k-1}^a - \mathcal{K}_{k-l|k} \mathbf{h}_k \mathbf{P}_{k,k-l|k-1}^{fa}, \quad (32b)$$

$$\mathbf{P}_{k,k-l|k}^{aa} = (\mathbf{I} - \mathcal{K}_{k|k} \mathbf{h}_k) \mathbf{P}_{k,k-l|k-1}^{fa}, \quad (32c)$$

where

$$\begin{aligned} \mathbf{P}_{k,k-l|k-1}^{fa} &= \mathcal{E} \left\{ \mathbf{e}_{k|k-1}^f (\mathbf{e}_{k-l|k-1}^a)^T \right\} \\ &= \mathcal{E} \left\{ \mathbf{A}_{k,k-1} \mathbf{e}_{k-1|k-1}^a (\mathbf{e}_{k-l|k-1}^a)^T \right\} \\ &= \mathbf{A}_{k,k-1} \mathbf{P}_{k-1,k-l|k-1}^{aa} \\ &= \left( \mathbf{P}_{k-l,k|k-1}^{af} \right)^T. \end{aligned} \quad (33)$$

Thus, we obtain the FLKS formula, equations (27a) to (27c) for the retrospective analyses, equations (32a) to (32c) as well as (33) for the error covariances.

## 2.2 3D-PSAS-like formulation of FLKS

In section 2.1, the FLKS formula are derived in a probabilistic approach. However, In practical implementation within the GEOS DAS framework, the evolution of the error covariances are not calculated explicitly due to the expensive computational cost, instead,  $\mathbf{P}_{k|k-1}^f$  is prescribed in advance. Assuming that the forecast error covariance  $\mathbf{P}_{k|k-1}^f$  is available at every analysis point, the 3D-PSAS-like formulation of FLKS will be derived in the linear perfect and imperfect model cases, respectively.

### 2.2.1 Linear imperfect model

Substituting the FLKS error covariance equations (32) and (33) into equations (25), we can rewrite  $\mathcal{K}_{k-i|k} \mathbf{v}_k$  in terms of the forecast error covariances  $\mathbf{P}_{k-i|k-i-1}^f$  ( $i = 0, 1, \dots, l$ ) which are specified at each analysis time:

$$\begin{aligned} \mathcal{K}_{k|k} \mathbf{v}_k &= \mathbf{P}_{k|k-1}^f \mathbf{h}_k^T \Gamma_k^{-1} \mathbf{v}_k \\ &= \mathbf{P}_{k|k-1}^f \mathbf{g}_0, \end{aligned} \quad (34a)$$

$$\begin{aligned} \mathcal{K}_{k-1|k} \mathbf{v}_k &= \left( \mathbf{P}_{k,k-1|k-1}^{fa} \right)^T \mathbf{h}_k^T \Gamma_k^{-1} \mathbf{v}_k \\ &= \mathbf{P}_{k-1|k-1}^a \mathbf{A}_{k,k-1}^T \mathbf{g}_0 \\ &= \mathbf{P}_{k-1|k-2}^f (\mathbf{I} - \mathcal{K}_{k-1|k-1} \mathbf{h}_{k-1})^T \mathbf{A}_{k,k-1}^T \mathbf{g}_0 \\ &= \mathbf{P}_{k-1|k-2}^f \mathbf{g}_1, \end{aligned} \quad (34b)$$

$$\begin{aligned} \mathcal{K}_{k-2|k} \mathbf{v}_k &= \left( \mathbf{P}_{k,k-2|k-1}^{fa} \right)^T \mathbf{h}_k^T \Gamma_k^{-1} \mathbf{v}_k \\ &= \left( \mathbf{P}_{k-1,k-2|k-1}^{aa} \right)^T \mathbf{A}_{k,k-1}^T \mathbf{g}_0 \\ &= \mathbf{P}_{k-2|k-2}^a \mathbf{A}_{k-1,k-2}^T (\mathbf{I} - \mathcal{K}_{k-1|k-1} \mathbf{h}_{k-1})^T \mathbf{A}_{k,k-1}^T \mathbf{g}_0 \\ &= \mathbf{P}_{k-2|k-2}^a \mathbf{A}_{k-1,k-2}^T \mathbf{g}_1 \\ &= \mathbf{P}_{k-2|k-3}^f (\mathbf{I} - \mathcal{K}_{k-2|k-2} \mathbf{h}_{k-2})^T \mathbf{A}_{k-1,k-2}^T \mathbf{g}_1 \\ &= \mathbf{P}_{k-2|k-3}^f \mathbf{g}_2, \end{aligned} \quad (34c)$$

$$\begin{aligned} &\vdots \\ \mathcal{K}_{k-l|k} \mathbf{v}_k &= \left( \mathbf{P}_{k,k-l|k-1}^{fa} \right)^T \mathbf{h}_k^T \Gamma_k^{-1} \mathbf{v}_k \\ &= \left( \mathbf{P}_{k-1,k-l|k-1}^{aa} \right)^T \mathbf{A}_{k,k-1}^T \mathbf{g}_0 \\ &= \left( \mathbf{P}_{k-2,k-l|k-2}^{aa} \right)^T \mathbf{A}_{k-1,k-2}^T \mathbf{g}_1 \\ &= \left( \mathbf{P}_{k-3,k-l|k-3}^{aa} \right)^T \mathbf{A}_{k-2,k-3}^T \mathbf{g}_2 \\ &\vdots \\ &= \mathbf{P}_{k-l|k-l}^a \mathbf{A}_{k-l+1,k-l}^T \mathbf{g}_{l-1} \\ &= \mathbf{P}_{k-l|k-l-1}^f (\mathbf{I} - \mathcal{K}_{k-l|k-l} \mathbf{h}_{k-l})^T \mathbf{A}_{k-l+1,k-l}^T \mathbf{g}_{l-1} \\ &= \mathbf{P}_{k-l|k-l-1}^f \mathbf{g}_l, \end{aligned} \quad (34d)$$

where

$$\mathbf{g}_0 = \mathbf{h}_k^T \Gamma_k^{-1} \mathbf{v}_k, \quad (35a)$$

$$\begin{aligned} \mathbf{g}_1 &= \left( \mathbf{I} - \mathcal{K}_{k-1|k-1} \mathbf{h}_{k-1} \right)^T \mathbf{A}_{k,k-1}^T \mathbf{g}_0 \\ &= \left[ \mathbf{I} - \mathbf{h}_{k-1}^T \Gamma_{k-1}^{-1} \mathbf{h}_{k-1} \mathbf{P}_{k-1|k-2}^f \right] \mathbf{A}_{k,k-1}^T \mathbf{g}_0, \end{aligned} \quad (35b)$$

$$\mathbf{g}_2 = \left( \mathbf{I} - \mathcal{K}_{k-2|k-2} \mathbf{h}_{k-2} \right)^T \mathbf{A}_{k-1,k-2}^T \mathbf{g}_1$$

$$= \left[ \mathbf{I} - \mathbf{h}_{k-2}^T \mathbf{\Gamma}_{k-2}^{-1} \mathbf{h}_{k-2} \mathbf{P}_{k-2|k-3}^f \right] \mathbf{A}_{k-1,k-2}^T \mathbf{g}_1 ; \quad (35c)$$

$$\begin{aligned} & \vdots \\ \mathbf{g}_l &= \left( \mathbf{I} - \mathcal{K}_{k-l|k-l} \mathbf{h}_{k-l} \right)^T \mathbf{A}_{k-l+1,k-l}^T \mathbf{g}_{l-1} \\ &= \left[ \mathbf{I} - \mathbf{h}_{k-l}^T \mathbf{\Gamma}_{k-l}^{-1} \mathbf{h}_{k-l} \mathbf{P}_{k-l|k-l-1}^f \right] \mathbf{A}_{k-l+1,k-l}^T \mathbf{g}_{l-1} . \end{aligned} \quad (35d)$$

**Remark:**

- It is clearly shown that  $\mathcal{K}_{k-i|k} \mathbf{v}_k$  is the linear combination of standard vectors with the coefficients given by the length projections of  $\mathbf{g}_i$  on the column vectors of forecast error covariance  $\mathbf{P}_{k-i|k-i-1}^f$ . The observation increment is spread out using the spatial structure of the forecast error covariance  $\mathbf{P}_{k-i|k-i-1}^f$ .

Therefore, the FLKS formula for  $l$  lags derived in section 2.1 can be rewritten as a recursive form in terms of  $\mathbf{P}_{k-i|k-i-1}^f$ :

$$\mathbf{w}_{k|k}^a = \mathbf{w}_{k|k-1}^f + \mathbf{P}_{k|k-1}^f \mathbf{g}_0 , \quad (36a)$$

$$\mathbf{w}_{k-1|k}^a = \mathbf{w}_{k-1|k-1}^a + \mathbf{P}_{k-1|k-2}^f \mathbf{g}_1 , \quad (36b)$$

$$\mathbf{w}_{k-2|k}^a = \mathbf{w}_{k-2|k-1}^a + \mathbf{P}_{k-2|k-3}^f \mathbf{g}_2 , \quad (36c)$$

$\vdots$

$$\mathbf{w}_{k-l|k}^a = \mathbf{w}_{k-l|k-1}^a + \mathbf{P}_{k-l|k-l-1}^f \mathbf{g}_l . \quad (36d)$$

The retrospective analysis increment is a linear combination of the column vectors of the forecast error covariance at that time.

## 2.2.2 Linear perfect model

Following Todling et al. (1998), for linear perfect model case where  $\mathbf{Q}_k = \mathbf{0}$  for  $k = 1, 2, \dots$ , we have a simple recursive formula for the retrospective analysis gains:

$$\begin{aligned} \mathcal{K}_{k|k} &= \mathbf{A}_{k,k-1} \mathcal{K}_{k-1|k} \\ &= \mathbf{A}_{k,k-2} \mathcal{K}_{k-2|k} \\ &\dots \end{aligned} \quad (37)$$

$$= \mathbf{A}_{k,k-l} \mathcal{K}_{k-l|k} . \quad (38)$$

Also using equation (25a), the FLKS analysis equations (27a) to (27c) can be further simplified as,

$$\mathbf{w}_{k|k}^a = \mathbf{w}_{k|k-1}^f + \mathbf{P}_{k|k-1}^f \mathbf{g}_0 , \quad (39a)$$

$$\mathbf{w}_{k-1|k}^a = \mathbf{w}_{k-1|k-1}^a + \mathbf{A}_{k,k-1}^{-1} \mathbf{P}_{k|k-1}^f \mathbf{g}_0 , \quad (39b)$$

$$\mathbf{w}_{k-2|k}^a = \mathbf{w}_{k-2|k-1}^a + \mathbf{A}_{k,k-2}^{-1} \mathbf{P}_{k|k-1}^f \mathbf{g}_0, \quad (39c)$$

⋮

$$\mathbf{w}_{k-l|k}^a = \mathbf{w}_{k-l|k-1}^a + \mathbf{A}_{k,k-l}^{-1} \mathbf{P}_{k|k-1}^f \mathbf{g}_0. \quad (39d)$$

The FLKS formula in the linear perfect model case is very simple. Once  $\mathbf{P}_{k|k-1}^f \mathbf{g}_0$  (i.e., the analysis increment) is evaluated by PSAS in the filter portion, the smoother solutions can be obtained by applying the quasi-inverse model to  $\mathbf{P}_{k|k-1}^f \mathbf{g}_0$  or by solving a linear system.

## 2.3 4D-VAR and 4D-PSAS

4D-VAR and 4D-PSAS are fixed-interval smoothers, aiming at producing the analyses at times  $t_k$  inside a fixed interval in which  $N + 1$  observations are available. The conditional probability density for 4D-VAR and 4D-PSAS can be written as

$$\begin{aligned} & p(\mathbf{w}_0^t, \mathbf{w}_1^t, \mathbf{w}_2^t, \dots, \mathbf{w}_N^t | \mathbf{W}_N^o) \\ &= \frac{1}{p(\mathbf{W}_N^o)} p(\mathbf{w}_0^t, \mathbf{w}_1^t, \mathbf{w}_2^t, \dots, \mathbf{w}_N^t, \mathbf{W}_N^o) \\ &= \frac{1}{p(\mathbf{W}_N^o)} p(\mathbf{W}_N^o | \mathbf{w}_0^t, \mathbf{w}_1^t, \mathbf{w}_2^t, \dots, \mathbf{w}_N^t) p(\mathbf{w}_0^t, \mathbf{w}_1^t, \mathbf{w}_2^t, \dots, \mathbf{w}_N^t) \\ &= \frac{1}{p(\mathbf{W}_N^o)} \prod_{k=0}^N p_{\mathbf{b}_k^o}(\mathbf{w}_k^o - \mathbf{h}_k(\mathbf{w}_k^t)) \prod_{k=1}^N p_{\mathbf{b}_k}(\mathbf{w}_k^t - \mathbf{A}_{k,k-1}(\mathbf{w}_{k-1}^t)) p(\mathbf{w}_0^t), \end{aligned} \quad (40)$$

where in the last line of the above equation, we use the assumptions that  $\{\mathbf{w}_0^o, \mathbf{w}_1^o, \dots, \mathbf{w}_N^o\}$  are independent, the observation error sequence  $\{\mathbf{b}_k^o\}$  and the model error sequence  $\{\mathbf{b}_k\}$  are white in time, respectively. It is obvious that the *a priori* information of model state variable is provided by the trajectory based on  $p(\mathbf{w}_0^t)$  which represents the *a priori* information of model initial state variable, and the  $N + 1$  observations are assimilated simultaneously.

If  $\{\mathbf{b}_k^o\}$ ,  $\{\mathbf{b}_k\}$ , and the *a priori* estimate of the model initial state variable  $\mathbf{w}_0^t$ , i.e.,  $\mathbf{w}^b$ , are Gaussian and independent from each other, also  $\mathbf{B}$  represents the error covariance of the *a priori* estimate, then the conditional probability density function  $p(\mathbf{w}_0^t, \mathbf{w}_1^t, \mathbf{w}_2^t, \dots, \mathbf{w}_N^t | \mathbf{W}_N^o)$  is proportional to  $\exp(-\mathcal{J}_N)$ , where the cost function  $\mathcal{J}_N$  is of the form

$$\begin{aligned} \mathcal{J}_N &= \frac{1}{2} (\mathbf{w}_0^t - \mathbf{w}^b)^T \mathbf{B}^{-1} (\mathbf{w}_0^t - \mathbf{w}^b) \\ &\quad + \frac{1}{2} \sum_{k=0}^N (\mathbf{w}_k^o - \mathbf{h}_k(\mathbf{w}_k^t))^T \mathbf{R}_k^{-1} (\mathbf{w}_k^o - \mathbf{h}_k(\mathbf{w}_k^t)) + \frac{1}{2} \sum_{k=1}^N \mathbf{b}_k^T \mathbf{Q}_k^{-1} \mathbf{b}_k. \end{aligned} \quad (41)$$

### 2.3.1 Linear imperfect model

For the linear case (both linear model and observation operator), the form of the cost function  $\mathcal{J}_N$  can be further rearranged. Let

$$\mathbf{A}_k = \mathbf{A}_{k,k-1} \mathbf{A}_{k-1,k-2} \dots \mathbf{A}_{2,1} \mathbf{A}_{1,0},$$

$$\mathbf{G}_k = \mathbf{h}_k \mathbf{A}_k, \quad \mathbf{d}_k = \mathbf{w}_k^o - \mathbf{G}_k \mathbf{w}^b, \\ \delta \mathbf{x} = \mathbf{w}_0^t - \mathbf{w}^b,$$

and

$$\delta \mathbf{x}(t_k) = \mathbf{w}_k^t - \mathbf{A}_k \mathbf{w}^b = \mathbf{A}_{k,k-1} \delta \mathbf{x}(t_{k-1}) + \mathbf{b}_k, \quad (42)$$

then the cost function can be rewritten as

$$\mathcal{J}_N = \frac{1}{2} \delta \mathbf{x}^T \mathbf{B}^{-1} \delta \mathbf{x} + \frac{1}{2} \sum_{k=0}^N (\mathbf{h}_k \delta \mathbf{x}(t_k) - \mathbf{d}_k)^T \mathbf{R}_k^{-1} (\mathbf{h}_k \delta \mathbf{x}(t_k) - \mathbf{d}_k) + \frac{1}{2} \sum_{k=1}^N \mathbf{b}_k^T \mathbf{Q}_k^{-1} \mathbf{b}_k. \quad (43)$$

Following Courtier (1997), introducing  $\mathbf{D}$  the block diagonal matrix consisting of  $\mathbf{B}$  for the first block and  $\mathbf{Q}_k$  for the others,  $\mathbf{R}$  with  $\mathbf{R}_k$  as the diagonal block elements, and

$$\mathbf{z} = \begin{pmatrix} \delta \mathbf{x} \\ \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_N \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} \mathbf{d}_0 \\ \mathbf{d}_1 \\ \vdots \\ \mathbf{d}_N \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} \mathbf{h}_0 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{h}_1 \mathbf{A}_1 & \mathbf{h}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{h}_2 \mathbf{A}_2 & \mathbf{h}_2 \mathbf{A}_{2,1} & \mathbf{h}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mathbf{h}_N \mathbf{A}_N & \mathbf{h}_N \mathbf{A}_{N,1} & \mathbf{h}_N \mathbf{A}_{N,2} & \dots & \mathbf{h}_N \end{pmatrix},$$

then the cost function is rewritten as

$$\mathcal{J}_N = \frac{1}{2} \mathbf{z}^T \mathbf{D}^{-1} \mathbf{z} + \frac{1}{2} (\mathbf{Gz} - \mathbf{d})^T \mathbf{R}^{-1} (\mathbf{Gz} - \mathbf{d}). \quad (44)$$

#### 4D-VAR:

4D-VAR incremental algorithm consists of minimizing the cost function  $\mathcal{J}_N$  directly using an iterative minimization procedure provided the information of the cost function and the gradient of the cost function with respect to the control variables are available at each iteration. The cost function value can be calculated following a forward integration of the equation (42), while the gradient of the cost function can be calculated by

$$\nabla_{\mathbf{z}} \mathcal{J}_N = \mathbf{D}^{-1} \mathbf{z} + \mathbf{G}^T \mathbf{R}^{-1} (\mathbf{Gz} - \mathbf{d}). \quad (45)$$

#### 4D-PSAS:

The solution of 4D-PSAS which minimizes the cost function  $\mathcal{J}_N$  is given as

$$\mathbf{z} = \mathbf{DG}^T (\mathbf{GDG}^T + \mathbf{R})^{-1} \mathbf{d}, \quad (46)$$

where  $(\mathbf{GDG}^T + \mathbf{R})^{-1} \mathbf{d}$  is obtained by minimizing a functional  $\mathcal{F}$  with respect to vector  $\mathbf{q}$ :

$$\mathcal{F} = \frac{1}{2} \mathbf{q}^T (\mathbf{GDG}^T + \mathbf{R}) \mathbf{q} - \mathbf{q}^T \mathbf{d}. \quad (47)$$

Appendix A shows that 4D-PSAS solution of the state analysis increment  $\delta \mathbf{x}$  is a linear combination of the column vectors of the error covariance  $\mathbf{B}$ .

### 2.3.2 Linear perfect model

If the linear model is assumed to be perfect, equation (42) becomes

$$\delta \mathbf{x}(t_k) = \mathbf{w}_k^t - \mathbf{A}_k \mathbf{w}^b = \mathbf{A}_{k,k-1} \delta \mathbf{x}(t_{k-1}), \quad (48)$$

then the cost function is of the form

$$\begin{aligned} \mathcal{J}_N &= \frac{1}{2} (\mathbf{w}_0^t - \mathbf{w}^b)^T \mathbf{B}^{-1} (\mathbf{w}_0^t - \mathbf{w}^b) \\ &\quad + \frac{1}{2} \sum_{k=0}^N (\mathbf{w}_k^o - \mathbf{h}_k \mathbf{w}_k^t)^T \mathbf{R}_k^{-1} (\mathbf{w}_k^o - \mathbf{h}_k \mathbf{w}_k^t) \end{aligned} \quad (49)$$

or

$$\mathcal{J}_N = \frac{1}{2} \delta \mathbf{x}^T \mathbf{B}^{-1} \delta \mathbf{x} + \frac{1}{2} \sum_{k=0}^N (\mathbf{h}_k \delta \mathbf{x}(t_k) - \mathbf{d}_k)^T \mathbf{R}_k^{-1} (\mathbf{h}_k \delta \mathbf{x}(t_k) - \mathbf{d}_k). \quad (50)$$

Introducing

$$\bar{\mathbf{G}} = \begin{pmatrix} \mathbf{G}_0 \\ \mathbf{G}_1 \\ \mathbf{G}_2 \\ \vdots \\ \mathbf{G}_N \end{pmatrix} = \begin{pmatrix} \mathbf{h}_0 \\ \mathbf{h}_1 \mathbf{A}_1 \\ \mathbf{h}_2 \mathbf{A}_2 \\ \vdots \\ \mathbf{h}_N \mathbf{A}_N \end{pmatrix}, \quad (51)$$

then  $\mathbf{d} = \mathbf{W}_N^o - \bar{\mathbf{G}} \mathbf{w}^b$ , and the cost function becomes

$$\mathcal{J}_N = \frac{1}{2} \delta \mathbf{x}^T \mathbf{B}^{-1} \delta \mathbf{x} + \frac{1}{2} (\bar{\mathbf{G}} \delta \mathbf{x} - \mathbf{d})^T \mathbf{R}^{-1} (\bar{\mathbf{G}} \delta \mathbf{x} - \mathbf{d}). \quad (52)$$

#### 4D-VAR:

The incremental 4D-VAR solution is obtained by minimizing the cost function  $\mathcal{J}_N$  directly using an iterative procedure with a forward integration of equation (48) and a backward integration of the adjoint model for the gradient of the cost function  $\nabla_{\delta \mathbf{x}} \mathcal{J}_N$ , which is given as

$$\nabla_{\delta \mathbf{x}} \mathcal{J}_N = \mathbf{B}^{-1} \delta \mathbf{x} + \bar{\mathbf{G}}^T \mathbf{R}^{-1} (\bar{\mathbf{G}} \delta \mathbf{x} - \mathbf{d}). \quad (53)$$

#### 4D-PSAS:

The solution of 4D-PSAS is achieved by minimizing the cost function  $\mathcal{J}_N$ :

$$\begin{aligned} \delta \mathbf{x} &= (\mathbf{B}^{-1} + \bar{\mathbf{G}}^T \mathbf{R}^{-1} \bar{\mathbf{G}})^{-1} \bar{\mathbf{G}}^T \mathbf{R}^{-1} \mathbf{d} \\ &= \mathbf{B} \bar{\mathbf{G}}^T (\bar{\mathbf{G}} \mathbf{B} \bar{\mathbf{G}}^T + \mathbf{R})^{-1} \mathbf{d}, \end{aligned} \quad (54)$$

and  $(\bar{\mathbf{G}} \mathbf{B} \bar{\mathbf{G}}^T + \mathbf{R})^{-1} \mathbf{d}$  is the solution of a minimization problem with respect to  $\mathbf{q}$ :

$$\mathcal{F}(\mathbf{q}) = \frac{1}{2} \mathbf{q}^T (\bar{\mathbf{G}} \mathbf{B} \bar{\mathbf{G}}^T + \mathbf{R}) \mathbf{q} - \mathbf{q}^T \mathbf{d}. \quad (55)$$

For linear perfect model case, it is very clear that 4D-PSAS solution of the state analysis increment  $\delta \mathbf{x}$  (equation (54)) is a linear combination of the column vectors of the error covariance  $\mathbf{B}$ .

### 3 The relationships among smoother algorithms

So far we have derived the FLKS and 4D-PSAS or 4D-VAR formula by applying Bayesian estimation theory (using the conditional probability density function). FLKS is a fixed-lag Kalman smoother, while 4D-PSAS and 4D-VAR can be considered as fixed-interval smoothers. However, we should notice that the definitions of fixed-lag and fixed-interval smoothers are mainly objective-oriented. In fact, they are different approaches or algorithms for solving the same problem. It is observed that fixed-lag and fixed-interval smoothers can be converted to the other. For instance, a fixed-interval smoother can be obtained from a fixed-lag smoother by setting

$$\mathbf{w}_{k-l|k-l-1}^f = \mathbf{w}^b, \mathbf{P}_{k-l|k-l-1}^f = \mathbf{B}, \text{ and lag } l = N,$$

while a fixed-lag smoother can be obtained by performing a moving fixed-interval smoother at every point  $k$  with

$$\mathbf{w}^b = \mathbf{w}_{k|k-1}^f, \mathbf{B} = \mathbf{P}_{k|k-1}^f, \text{ and } N = l.$$

Here, we would like to point out that, if the observation information at initial time are not included in the fixed-interval, instead are already embodied in  $\mathbf{w}^b$  and  $\mathbf{B}$ , then we should set

$$\mathbf{w}^b = \mathbf{w}_{k|k}^a, \mathbf{B} = \mathbf{P}_{k|k}^a, \text{ and } N = l.$$

In this section, we will mainly focus on analyzing the relationships among the three different algorithms – 4D-VAR, 4D-PSAS and FLKS, and comparing their analyses in terms of analysis qualities in order to locate the most suitable algorithm for doing retrospective data assimilation within the GEOS DAS framework.

We will start with the cost function of 4D-VAR and 4D-PSAS (see Section 2.3). Because  $p(\mathbf{w}_0^t, \mathbf{w}_1^t, \mathbf{w}_2^t, \dots, \mathbf{w}_N^t | \mathbf{W}_N^o)$  is a Gaussian density, the sequence  $\mathbf{w}_{0|N}^a, \mathbf{w}_{1|N}^a, \mathbf{w}_{2|N}^a, \dots, \mathbf{w}_{N|N}^a$  maximizes the density for fixed  $\mathbf{W}_N^o$ . Section 2.3 shows that for linear case this maximization is equivalent to the minimization of the cost function

$$\begin{aligned} \mathcal{J}_N &= \frac{1}{2}(\mathbf{w}_0^t - \mathbf{w}^b)^T \mathbf{B}^{-1}(\mathbf{w}_0^t - \mathbf{w}^b) \\ &+ \frac{1}{2} \sum_{k=0}^N (\mathbf{w}_k^o - \mathbf{h}_k \mathbf{w}_k^t)^T \mathbf{R}_k^{-1} (\mathbf{w}_k^o - \mathbf{h}_k \mathbf{w}_k^t) + \frac{1}{2} \sum_{k=1}^N \mathbf{b}_k^T \mathbf{Q}_k^{-1} \mathbf{b}_k, \end{aligned} \quad (56)$$

subject to  $\mathbf{w}_k^t = \mathbf{A}_{k,k-1} \mathbf{w}_{k-1}^t + \mathbf{b}_k$ . This problem can be solved by using Lagrangian multiplier  $\lambda$  to adjoining the constraint to the cost function  $\mathcal{J}_N$ :

$$\begin{aligned} \mathcal{J}'_N &= \frac{1}{2}(\mathbf{w}_0^t - \mathbf{w}^b)^T \mathbf{B}^{-1}(\mathbf{w}_0^t - \mathbf{w}^b) \\ &+ \frac{1}{2} \sum_{k=0}^N (\mathbf{w}_k^o - \mathbf{h}_k \mathbf{w}_k^t)^T \mathbf{R}_k^{-1} (\mathbf{w}_k^o - \mathbf{h}_k \mathbf{w}_k^t) + \frac{1}{2} \sum_{k=1}^N \mathbf{b}_k^T \mathbf{Q}_k^{-1} \mathbf{b}_k \\ &+ \sum_{k=1}^N \lambda_k^T (\mathbf{w}_k^t - \mathbf{A}_{k,k-1} \mathbf{w}_{k-1}^t - \mathbf{b}_k). \end{aligned} \quad (57)$$



Now consider the first variation of  $\mathcal{J}'_N$ ,

$$\begin{aligned} \delta \mathcal{J}'_N &= (\delta \mathbf{w}_0^t)^T \mathbf{B}^{-1} (\mathbf{w}_0^t - \mathbf{w}^b) + \sum_{k=0}^N (\delta \mathbf{w}_k^t)^T \mathbf{h}_k^T \mathbf{R}_k^{-1} (\mathbf{h}_k \mathbf{w}_k^t - \mathbf{w}_k^o) + \sum_{k=1}^N (\delta \mathbf{b}_k)^T \mathbf{Q}_k^{-1} \mathbf{b}_k + \\ &\quad \sum_{k=1}^N \left[ (\delta \lambda_k)^T (\mathbf{w}_k^t - \mathbf{A}_{k,k-1} \mathbf{w}_{k-1}^t - \mathbf{b}_k) + \lambda_k^T \delta \mathbf{w}_k^t - \lambda_k^T \mathbf{A}_{k,k-1} \delta \mathbf{w}_{k-1}^t - \lambda_k^T \delta \mathbf{b}_k \right], \end{aligned} \quad (58)$$

where  $\delta(\cdot)$  denotes the first variation of a variable. Rearranging the terms in equation (58), and setting the coefficients of  $\delta \mathbf{w}_k^t$ ,  $\delta \mathbf{b}_k$  and  $\delta \lambda$  to be zero, we obtain the discrete Euler-Lagrange equations:

$$\lambda_k - \mathbf{A}_{k+1,k}^T \lambda_{k+1} - \mathbf{h}_k^T \mathbf{R}_k^{-1} (\mathbf{w}_k^o - \mathbf{h}_k \mathbf{w}_k^t) = \mathbf{0}, \quad \text{for } k = 1, 2, \dots, N-1 \quad (59a)$$

$$\lambda_N = \mathbf{h}_N^T \mathbf{R}_N^{-1} (\mathbf{w}_N^o - \mathbf{h}_N \mathbf{w}_N^t), \quad (59b)$$

$$\mathbf{B}^{-1} (\mathbf{w}_0^t - \mathbf{w}^b) + \mathbf{h}_0^T \mathbf{R}_0^{-1} (\mathbf{h}_0 \mathbf{w}_0^t - \mathbf{w}_0^o) - \mathbf{A}_{1,0}^T \lambda_1 = \mathbf{0}, \quad (59c)$$

$$\mathbf{w}_k^t - \mathbf{A}_{k,k-1} \mathbf{w}_{k-1}^t - \mathbf{b}_k = \mathbf{0}, \quad (59d)$$

$$\mathbf{Q}_k^{-1} \mathbf{b}_k - \lambda_k = \mathbf{0}. \quad (59e)$$

Hence, the optimal estimates of  $\mathbf{w}_k^t$  and  $\mathbf{b}_k$  will be obtained by solving the two point boundary value problem,

$$\mathbf{w}_k^t = \mathbf{A}_{k,k-1} \mathbf{w}_{k-1}^t + \mathbf{Q}_k \lambda_k, \quad \text{for } k = 1, 2, \dots, N \quad (60a)$$

$$\lambda_k = \mathbf{A}_{k+1,k}^T \lambda_{k+1} + \mathbf{h}_k^T \mathbf{R}_k^{-1} (\mathbf{w}_k^o - \mathbf{h}_k \mathbf{w}_k^t), \quad \text{for } k = N-1, \dots, 2, 1 \quad (60b)$$

with boundary conditions

$$\lambda_N = \mathbf{h}_N^T \mathbf{R}_N^{-1} (\mathbf{w}_N^o - \mathbf{h}_N \mathbf{w}_N^t) \quad (61)$$

and the initial estimate of  $\mathbf{w}_0^t$ , i.e.  $\mathbf{w}^b$ , given.

### 3.1 4D-VAR

Equations (60b) and (61) are the actual adjoint equation for the 4D-VAR algorithm, which is integrated backward in time. The gradients of the cost function with respect to the control variables are given as following

$$\nabla_{\mathbf{w}_0^t} \mathcal{J}_N = \mathbf{B}^{-1} (\mathbf{w}_0^t - \mathbf{w}^b) + \mathbf{h}_0^T \mathbf{R}_0^{-1} (\mathbf{h}_0 \mathbf{w}_0^t - \mathbf{w}_0^o) - \mathbf{A}_{1,0}^T \lambda_1, \quad (62)$$

$$\nabla_{\mathbf{b}_k} \mathcal{J}_N = \mathbf{Q}_k^{-1} \mathbf{b}_k - \lambda_k. \quad (63)$$

4D-VAR is carried out using an iterative minimization procedure, with a forward integration of equation (59d) and a backward integration of equation (59a) at each iteration. If the procedure converges, then at the minimum the gradients of the cost function vanish, that is, equations (59c) and (59e) hold, thus providing the solution that minimizes the cost function  $\mathcal{J}_N$ .

### 3.2 FLKS

In the following the sweep method (Bryson and Ho, 1975) will be employed to solve this problem and to get the FLKS algorithm. We shall see that the FLKS algorithm is essentially equivalent to 4D-VAR, 4D-PSAS and Kalman smoother as well as Representer method for linear dynamics in the context of fixed-interval smoothing.

From equation (59c), we get

$$\mathbf{w}_0^t = \hat{\mathbf{w}}_0 + \mathbf{S}_0 \lambda_1, \quad (64)$$

where

$$\hat{\mathbf{w}}_0 = [\mathbf{B}^{-1} + \mathbf{h}_0^T \mathbf{R}_0^{-1} \mathbf{h}_0]^{-1} (\mathbf{B}^{-1} \mathbf{w}^b + \mathbf{h}_0^T \mathbf{R}_0^{-1} \mathbf{w}_0^o), \quad (65a)$$

$$\mathbf{S}_0 = [\mathbf{B}^{-1} + \mathbf{h}_0^T \mathbf{R}_0^{-1} \mathbf{h}_0]^{-1} \mathbf{A}_{1,0}^T. \quad (65b)$$

By superposition, we could write the solution at time  $t_k$  as

$$\mathbf{w}_k^t = \hat{\mathbf{w}}_k + \mathbf{S}_k \lambda_{k+1}, \quad (66)$$

where  $\hat{\mathbf{w}}_k$  and  $\mathbf{S}_k$  are still to be determined. Substituting equation (66) into equation (60b), this yields,

$$\lambda_k = (\mathbf{A}_{k+1,k}^T - \mathbf{h}_k^T \mathbf{R}_k^{-1} \mathbf{h}_k \mathbf{S}_k) \lambda_{k+1} + \mathbf{h}_k^T \mathbf{R}_k^{-1} (\mathbf{w}_k^o - \mathbf{h}_k \hat{\mathbf{w}}_k). \quad (67)$$

Also equation (66) yields  $\mathbf{w}_N^t = \hat{\mathbf{w}}_N$ , which leads to

$$\lambda_N = \mathbf{h}_N^T \mathbf{R}_N^{-1} (\mathbf{w}_N^o - \mathbf{h}_N \hat{\mathbf{w}}_N).$$

Now one forward sweep from time  $t_0$  to  $t_N$  is performed in order to determine  $\hat{\mathbf{w}}_k$  and  $\mathbf{S}_k$ . From equation (66) we have

$$\mathbf{w}_{k-1}^t = \hat{\mathbf{w}}_{k-1} + \mathbf{S}_{k-1} \lambda_k. \quad (68)$$

Multiplying equation (68) by  $\mathbf{A}_{k,k-1}$ , and subtracting it from equation (66), this gives,

$$\mathbf{w}_k^t - \mathbf{A}_{k,k-1} \mathbf{w}_{k-1}^t = \hat{\mathbf{w}}_k - \mathbf{A}_{k,k-1} \hat{\mathbf{w}}_{k-1} + \mathbf{S}_k \lambda_{k+1} - \mathbf{A}_{k,k-1} \mathbf{S}_{k-1} \lambda_k. \quad (69)$$

Substituting equation (60a) into equation (69), we have,

$$(\mathbf{Q}_k + \mathbf{A}_{k,k-1} \mathbf{S}_{k-1}) \lambda_k - \mathbf{S}_k \lambda_{k+1} = \hat{\mathbf{w}}_k - \mathbf{A}_{k,k-1} \hat{\mathbf{w}}_{k-1}. \quad (70)$$

Applying equation (67) to equation (70), and setting the coefficient of  $\lambda$  equal to zero, then we obtain the following two equations:

$$\hat{\mathbf{w}}_k = \mathbf{A}_{k,k-1} \hat{\mathbf{w}}_{k-1} + (\mathbf{Q}_k + \mathbf{A}_{k,k-1} \mathbf{S}_{k-1}) \mathbf{h}_k^T \mathbf{R}_k^{-1} (\mathbf{w}_k^o - \mathbf{h}_k \hat{\mathbf{w}}_k), \quad (71a)$$

$$\mathbf{S}_k = (\mathbf{A}_{k,k-1} \mathbf{S}_{k-1} + \mathbf{Q}_k) \mathbf{A}_{k+1,k}^T - (\mathbf{A}_{k,k-1} \mathbf{S}_{k-1} + \mathbf{Q}_k) \mathbf{h}_k^T \mathbf{R}_k^{-1} \mathbf{h}_k \mathbf{S}_k. \quad (71b)$$

Rearranging the above two equations, we have

$$\hat{\mathbf{w}}_k = \mathbf{A}_{k,k-1} \hat{\mathbf{w}}_{k-1} + (\mathbf{A}_{k,k-1} \mathbf{S}_{k-1} + \mathbf{Q}_k) \mathbf{h}_k^T \left[ \mathbf{h}_k (\mathbf{A}_{k,k-1} \mathbf{S}_{k-1} + \mathbf{Q}_k) \mathbf{h}_k^T + \mathbf{R}_k \right]^{-1} (\mathbf{w}_k^o - \mathbf{h}_k \mathbf{A}_{k,k-1} \hat{\mathbf{w}}_{k-1}), \quad (72a)$$

$$\mathbf{S}_k = \left[ (\mathbf{A}_{k,k-1} \mathbf{S}_{k-1} + \mathbf{Q}_k)^{-1} + \mathbf{h}_k^T \mathbf{R}_k^{-1} \mathbf{h}_k \right]^{-1} \mathbf{A}_{k+1,k}^T, \quad (72b)$$

with the initial conditions from equations (65a) and (65b)

$$\hat{\mathbf{w}}_0 = \mathbf{w}^b + \mathbf{B}\mathbf{h}_0^T(\mathbf{h}_0\mathbf{B}\mathbf{h}_0^T + \mathbf{R}_0)^{-1}(\mathbf{w}_0^o - \mathbf{h}_0\mathbf{w}^b), \quad (73a)$$

$$\mathbf{S}_0 = [\mathbf{B}^{-1} + \mathbf{h}_0^T\mathbf{R}_0^{-1}\mathbf{h}_0]^{-1} \mathbf{A}_{1,0}^T. \quad (73b)$$

It is obvious that this set of equations can be solved by one forward sweep of the boundary condition from  $t_0$  to  $t_N$ . If we define

$$P_k^f = \mathbf{A}_{k,k-1}\mathbf{S}_{k-1} + \mathbf{Q}_k, \quad (74a)$$

$$P_0^a = [\mathbf{B}^{-1} + \mathbf{h}_0^T\mathbf{R}_0^{-1}\mathbf{h}_0]^{-1}, \quad (74b)$$

$$P_k^a = [(P_k^f)^{-1} + \mathbf{h}_k^T\mathbf{R}_k^{-1}\mathbf{h}_k]^{-1}, \quad (74c)$$

$$\Gamma_k = \mathbf{h}_k P_k^f \mathbf{h}_k^T + \mathbf{R}_k, \quad (74d)$$

$$K_k = P_k^f \mathbf{h}_k^T \Gamma_k^{-1} \quad (74e)$$

$$= P_k^a \mathbf{h}_k^T \mathbf{R}_k^{-1}, \quad (74f)$$

then the equations (72a) and (72b) can be rewritten as

$$\hat{\mathbf{w}}_k = \mathbf{A}_{k,k-1}\hat{\mathbf{w}}_{k-1} + K_k(\mathbf{w}_k^o - \mathbf{h}_k\mathbf{A}_{k,k-1}\hat{\mathbf{w}}_{k-1}), \quad (75a)$$

$$\mathbf{S}_k = P_k^a \mathbf{A}_{k+1,k}^T. \quad (75b)$$

It is an interesting observation that, given  $\mathbf{B} = \mathbf{P}_{0|-1}^f$  and  $\mathbf{w}^b = \mathbf{w}_{0|-1}^f$ , then equations (74a) - (74f) and (75a) for  $P_k^a$ ,  $P_k^f$ ,  $\Gamma_k$ ,  $K_k$  and  $\hat{\mathbf{w}}_k$  are the same as the Kalman filter equations for  $\mathbf{P}_{k|k}^a$ ,  $\mathbf{P}_{k|k-1}^f$ ,  $\Gamma_k$ ,  $\mathcal{K}_{k|k}$  and  $\mathbf{w}_{k|k}^a$ , i.e.,  $P_k^a = \mathbf{P}_{k|k}^a$ ,  $P_k^f = \mathbf{P}_{k|k-1}^f$ ,  $K_k = \mathcal{K}_{k|k}$ , and  $\hat{\mathbf{w}}_k = \mathbf{w}_{k|k}^a$ . Therefore, substituting equation (75b) into equation (66), we get the smoother estimate of  $\mathbf{w}_k^t$ , i.e.,  $\mathbf{w}_{k|N}^a$ , as

$$\mathbf{w}_{k|N}^a = \mathbf{w}_{k|k}^a + \mathbf{P}_{k|k}^a \mathbf{A}_{k+1,k}^T \lambda_{k+1}. \quad (76)$$

A backward sweep from time  $t_N$  to  $t_0$  is then performed to produce the smoother estimate. Using equations (75a) and (75b), the  $\lambda$  equation (67) can be written as

$$\begin{aligned} \lambda_k &= (\mathbf{I} - \mathbf{h}_k^T \mathbf{R}_k^{-1} \mathbf{h}_k \mathbf{P}_{k|k}^a) \mathbf{A}_{k+1,k}^T \lambda_{k+1} \\ &\quad + \mathbf{h}_k^T \mathbf{R}_k^{-1} (\mathbf{I} - \mathbf{h}_k \mathcal{K}_{k|k}) (\mathbf{w}_k^o - \mathbf{h}_k \mathbf{A}_{k,k-1} \mathbf{w}_{k-1|k-1}^a) \\ &= (\mathbf{I} - \mathcal{K}_{k|k} \mathbf{h}_k)^T \mathbf{A}_{k+1,k}^T \lambda_{k+1} + \mathbf{h}_k^T \Gamma_k^{-1} (\mathbf{w}_k^o - \mathbf{h}_k \mathbf{A}_{k,k-1} \mathbf{w}_{k-1|k-1}^a), \end{aligned} \quad (77)$$

with the boundary condition

$$\begin{aligned} \lambda_N &= \mathbf{h}_N^T \mathbf{R}_N^{-1} (\mathbf{I} - \mathbf{h}_N \mathcal{K}_{N|N}) (\mathbf{w}_N^o - \mathbf{h}_N \mathbf{A}_{N,N-1} \mathbf{w}_{N-1|N-1}^a) \\ &= \mathbf{h}_N^T \Gamma_N^{-1} (\mathbf{w}_N^o - \mathbf{h}_N \mathbf{A}_{N,N-1} \mathbf{w}_{N-1|N-1}^a). \end{aligned} \quad (78)$$

Using equation (29), equations (77) and (78) can be written as

$$\lambda_k = (\mathbf{I} - \mathcal{K}_{k|k} \mathbf{h}_k)^T \mathbf{A}_{k+1,k}^T \lambda_{k+1} + \mathbf{h}_k^T \Gamma_k^{-1} \mathbf{v}_k, \quad (79a)$$

$$\lambda_N = \mathbf{h}_N^T \Gamma_N^{-1} \mathbf{v}_N. \quad (79b)$$

Substituting equations (79a) and (79b) into equation (76), and introducing

$$\mathbf{g}_0 = \mathbf{h}_N^T \mathbf{\Gamma}_N^{-1} \mathbf{v}_N, \quad (80a)$$

$$\mathbf{g}_1 = (\mathbf{I} - \mathcal{K}_{N-1|N-1} \mathbf{h}_{N-1})^T \mathbf{A}_{N,N-1}^T \mathbf{g}_0, \quad (80b)$$

⋮

$$\mathbf{g}_l = (\mathbf{I} - \mathcal{K}_{N-l|N-l} \mathbf{h}_{N-l})^T \mathbf{A}_{N-l+1,N-l}^T \mathbf{g}_{l-1}, \quad (80c)$$

then after some manipulation, we can obtain

$$\mathbf{w}_{N-1|N}^a = \mathbf{w}_{N-1|N-1}^a + \mathbf{P}_{N-1|N-1}^a \mathbf{A}_{N,N-1}^T \mathbf{g}_0, \quad (81a)$$

$$\mathbf{w}_{N-2|N}^a = \mathbf{w}_{N-2|N-1}^a + \mathbf{P}_{N-2|N-2}^a \mathbf{A}_{N-1,N-2}^T \mathbf{g}_1, \quad (81b)$$

⋮

$$\mathbf{w}_{N-l|N}^a = \mathbf{w}_{N-l|N-1}^a + \mathbf{P}_{N-l|N-l}^a \mathbf{A}_{N-l+1,N-l}^T \mathbf{g}_{l-1}. \quad (81c)$$

That is, the smoother estimate of  $\mathbf{w}_k^t$ , i.e.,  $\mathbf{w}_{k|N}^a$ , can be written as

$$\begin{aligned} \mathbf{w}_{k|N}^a &= \mathbf{w}_{k|N-1}^a + \mathbf{P}_{k|k}^a \mathbf{A}_{k+1,k}^T \mathbf{g}_{N-k-1} \\ &= \mathbf{w}_{k|N-1}^a + \mathbf{P}_{k|k-1}^f (\mathbf{I} - \mathcal{K}_{k|k} \mathbf{h}_k)^T \mathbf{A}_{k+1,k}^T \mathbf{g}_{N-k-1} \\ &= \mathbf{w}_{k|N-1}^a + \mathbf{P}_{k|k-1}^f \mathbf{g}_{N-k}, \end{aligned} \quad (82)$$

where equation (74c) is employed to write the smoother analysis formula in terms of  $\mathbf{P}_{k|k-1}^f$ . We can see that the equation (82) is the same as the FLKS formula derived in section 2.2 for lag  $l = N - k$ .

### 3.3 4D-PSAS

The 4D-PSAS solution can also be derived from equations (59a) to (59e). Using equations (59a) and (59b), we have

$$\lambda_1 = \sum_{k=1}^N \mathbf{A}_{k,1}^T \mathbf{h}_k^T \mathbf{R}_k^{-1} (\mathbf{w}_k^o - \mathbf{h}_k \mathbf{w}_k^t). \quad (83)$$

Applying equation (83) into equation (59c), it follows

$$\begin{aligned} \mathbf{w}_0^t &= \mathbf{w}^b + [\mathbf{B}^{-1} + \mathbf{h}_0^T \mathbf{R}_0^{-1} \mathbf{h}_0]^{-1} \mathbf{h}_0^T \mathbf{R}_0^{-1} (\mathbf{w}_0^o - \mathbf{h}_0 \mathbf{w}^b) \\ &\quad + [\mathbf{B}^{-1} + \mathbf{h}_0^T \mathbf{R}_0^{-1} \mathbf{h}_0]^{-1} \sum_{k=1}^N \mathbf{A}_k^T \mathbf{h}_k^T \mathbf{R}_k^{-1} (\mathbf{w}_k^o - \mathbf{h}_k \mathbf{w}_k^t), \end{aligned} \quad (84)$$

If setting  $\delta \mathbf{x} = \mathbf{w}_0^t - \mathbf{w}^b$  and  $\mathbf{d}_k = \mathbf{w}_k^o - \mathbf{h}_k \mathbf{A}_k \mathbf{w}^b$ , then this equation becomes

$$\mathbf{B}^{-1} \delta \mathbf{x} = \mathbf{h}_0^T \mathbf{R}_0^{-1} (\mathbf{d}_0 - \mathbf{h}_0 \delta \mathbf{x}) + \sum_{k=1}^N \mathbf{A}_k^T \mathbf{h}_k^T \mathbf{R}_k^{-1} (\mathbf{w}_k^o - \mathbf{h}_k \mathbf{w}_k^t). \quad (85)$$

Also from equation (59e) it is seen that

$$\begin{aligned}\mathbf{Q}_k^{-1}\mathbf{b}_k &= \lambda_k \\ &= \sum_{i=k}^N \mathbf{A}_{i,k}^T \mathbf{h}_i^T \mathbf{R}_i^{-1} (\mathbf{w}_i^o - \mathbf{h}_i \mathbf{w}_i^t) .\end{aligned}\quad (86)$$

Both equations (85) and (86) involve the true state  $\mathbf{w}_k^t$ , whose evolution is described as

$$\begin{aligned}\mathbf{w}_k^t &= \mathbf{A}_{k,k-1} \mathbf{w}_{k-1}^t + \mathbf{b}_k \\ &= \mathbf{A}_k \mathbf{w}_0^t + \mathbf{A}_{k,1} \mathbf{b}_1 + \mathbf{A}_{k,2} \mathbf{b}_2 + \dots + \mathbf{A}_{k,k-1} \mathbf{b}_{k-1} + \mathbf{b}_k ,\end{aligned}\quad (87)$$

hence we have

$$\mathbf{w}_k^o - \mathbf{h}_k \mathbf{w}_k^t = \mathbf{d}_k - \mathbf{h}_k (\mathbf{A}_k \delta \mathbf{x} + \mathbf{A}_{k,1} \mathbf{b}_1 + \mathbf{A}_{k,2} \mathbf{b}_2 + \dots + \mathbf{A}_{k,k-1} \mathbf{b}_{k-1} + \mathbf{b}_k) .\quad (88)$$

Using equation (88), we can rewrite equations (85) and (86) in a compact formula

$$\mathbf{D}^{-1} \mathbf{z} = \mathbf{G}^T \mathbf{R}^{-1} (\mathbf{d} - \mathbf{G} \mathbf{z}) .\quad (89)$$

Therefore, the solution of the 4D-PSAS is given as

$$\mathbf{z} = \mathbf{D} \mathbf{G}^T (\mathbf{R} + \mathbf{G} \mathbf{D} \mathbf{G}^T)^{-1} \mathbf{d} ,\quad (90)$$

where  $(\mathbf{R} + \mathbf{G} \mathbf{D} \mathbf{G}^T)^{-1} \mathbf{d}$  is solved as the solution of a minimization problem. It is shown, in the next section, that the evaluation of the multiplication of  $\mathbf{G} \mathbf{D} \mathbf{G}^T$  with a vector requires a forward model integration and a backward model integration.

### 3.4 Representer method

Here we give a brief description of representer method (Bennett, 1992; Bennett et al., 1996), which is another approach to solve the same problem. With the same cost function as 4D-VAR and 4D-PSAS, the representer method solves the Euler-Lagrange equations (59a) - (59e) by searching the  $(N + 1)$ -dimensional space of representer coefficients.

The estimates of  $\mathbf{w}_k^t$  is given by:

$$\mathbf{w}_{k|N}^a = \mathbf{w}_{k|0}^f + \sum_{m=0}^N \mathbf{r}_{k,m} \mathbf{t}_m ,\quad (91)$$

where  $\mathbf{w}_{k|0}^f$  is the solution of the following forward model integration

$$\mathbf{w}_{k|0}^f = \mathbf{A}_{k,k-1} \mathbf{w}_{k-1|0}^f ,\quad (92a)$$

$$\mathbf{w}_{0|0}^f = \mathbf{w}^b ,\quad (92b)$$

and  $\mathbf{r}_{k,m}$  is the representer function, satisfying

$$\mathbf{r}_{k,m} = \mathbf{A}_{k,k-1} \mathbf{r}_{k-1,m} + \mathbf{Q}_k \alpha_{k,m} ,\quad (93a)$$

$$\mathbf{r}_{0,m} = \mathbf{B} \alpha_{0,m} .\quad (93b)$$

The representer adjoint variable  $\alpha_{k,m}$  satisfies

$$\alpha_{k,m} - \mathbf{A}_{k+1,k}^T \alpha_{k+1,m} = \mathbf{h}_k^T \delta_{km} , \quad (94a)$$

$$\alpha_{N,m} = \mathbf{h}_N^T \delta_{Nm} . \quad (94b)$$

Substitution of equation (91) into the Euler-Lagrange equations (59a) - (59e) yields a linear system for the vector  $\mathbf{t}$  of representer coefficients:

$$(\mathbf{R} + \mathbf{H}\mathbf{r}^T) \mathbf{t} = \mathbf{d} , \quad (95)$$

where  $\mathbf{H}$  is a block diagonal matrix with  $\mathbf{h}_k$  as block diagonal elements.

This approach requires  $2(N + 1) + 1$  integrations (1 forward integration for  $\mathbf{w}_{k|0}^f$ , 1 forward integration and 1 backward integration for each of the  $N + 1$  representers and their adjoints) followed by solving the linear system (95).

### 3.5 Kalman smoother

Kalman smoother (Evensen, 1997) is similar to the analysis method used in the Kalman filter except that the smoother estimate is calculated over the whole space and time domain  $[t_0, t_N]$ :

$$\mathbf{W}^a = \mathbf{W}^f + (\mathbf{H}\mathbf{P}^f)^T \mathbf{t} \quad (96)$$

with

$$(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}) \mathbf{t} = \mathbf{d} , \quad (97)$$

where  $\mathbf{W}^a$ ,  $\mathbf{W}^f$  and the forecast error covariance  $\mathbf{P}^f$  over the whole space and time domain are defined as

$$\mathbf{W}^a = \begin{pmatrix} \mathbf{w}_{0|N}^a \\ \mathbf{w}_{1|N}^a \\ \vdots \\ \mathbf{w}_{N|N}^a \end{pmatrix} , \quad \mathbf{W}^f = \begin{pmatrix} \mathbf{w}_{1|0}^b \\ \mathbf{w}_{1|0}^f \\ \vdots \\ \mathbf{w}_{N|0}^f \end{pmatrix} , \quad \mathbf{P}^f = \begin{pmatrix} \mathbf{B} & \mathbf{P}_{0,1|0}^f & \cdots & \mathbf{P}_{0,N|0}^f \\ \mathbf{P}_{1,0|0} & \mathbf{P}_{1,1|0}^f & \cdots & \mathbf{P}_{1,N|0}^f \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{N,0|0} & \mathbf{P}_{N,1|0}^f & \cdots & \mathbf{P}_{N,N|0}^f \end{pmatrix} . \quad (98)$$

$\mathbf{w}_{k|0}^f$  is also the solution of the forward model integration (92a) - (92b). The posterior error covariance matrix  $\mathbf{P}^a$  can be calculated as

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{C}\mathbf{H})\mathbf{P}^f \quad (99)$$

with  $\mathbf{C}$  defined as

$$\mathbf{C} = (\mathbf{H}\mathbf{P}^f)^T (\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1} . \quad (100)$$

Evensen's original smoother is an ensemble smoother, i.e., the error covariances  $\mathbf{P}^f$  and  $\mathbf{P}^a$  are computed using a Monte Carlo method. This method can achieve more accurate evaluations of the error covariances for strongly nonlinear dynamical systems, but it is not feasible for high dimensional problems. If we expand the forecast error covariance  $\mathbf{P}^f$  in terms of  $\mathbf{B}$  and  $\bar{\mathbf{Q}}$ , it

can be shown that the analyses produced by Kalman smoother are identical to those produced by 4D-PSAS (see Appendix A) except that Kalman smoother yields the analyses over the whole time domain at the same time, while 4D-PSAS yields only the analysis at time  $t_0$  for the initial condition, the analyses for time  $t > t_0$  are produced by issuing a forecast of equation (59d) from the initial condition.

For Kalman smoother, the forecast error covariance  $\mathbf{P}^f$  over the whole space and time domain (equation (98)) can be computed as

$$\begin{aligned} \mathbf{P}^f &= \mathcal{E}\{\mathbf{e}^f \mathbf{e}^{fT}\} \\ &= \mathcal{E}\left\{ \begin{pmatrix} \mathbf{e}_{0|0} \\ \mathbf{A}_1 \mathbf{e}_{0|0} + \mathbf{b}_1 \\ \vdots \\ \mathbf{A}_N \mathbf{e}_{0|0} + \mathbf{A}_{N,1} \mathbf{b}_1 + \dots + \mathbf{b}_N \end{pmatrix} \begin{pmatrix} \mathbf{e}_{0|0} \\ \mathbf{A}_1 \mathbf{e}_{0|0} + \mathbf{b}_1 \\ \vdots \\ \mathbf{A}_N \mathbf{e}_{0|0} + \mathbf{A}_{N,1} \mathbf{b}_1 + \dots + \mathbf{b}_N \end{pmatrix}^T \right\} \\ &= \mathbf{U} \mathbf{B} \mathbf{U}^T + \mathbf{V} \bar{\mathbf{Q}} \mathbf{V}^T, \end{aligned} \quad (101)$$

where

$$\mathbf{U} = \begin{pmatrix} \mathbf{I} \\ \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_N \end{pmatrix}, \quad \mathbf{V} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{2,1} & \mathbf{I} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mathbf{0} & \mathbf{A}_{N,1} & \mathbf{A}_{N,2} & \dots & \mathbf{I} \end{pmatrix}.$$

Therefore, we can see that equation (100) can be rewritten as

$$\mathbf{C} = (\mathbf{U} \bar{\mathbf{B}} \bar{\mathbf{G}}^T + \mathbf{V} \bar{\mathbf{Q}} \mathcal{G}^T) (\bar{\mathbf{G}} \bar{\mathbf{B}} \bar{\mathbf{G}}^T + \mathcal{G} \bar{\mathbf{Q}} \mathcal{G}^T + \mathbf{R})^{-1}, \quad (102)$$

where  $\bar{\mathbf{Q}}$  and  $\mathcal{G}$  are defined in Appendix A. Kalman smoother (equation (96)) then is of the form

$$\mathbf{W}^a = \mathbf{W}^f + (\mathbf{U} \bar{\mathbf{B}} \bar{\mathbf{G}}^T + \mathbf{V} \bar{\mathbf{Q}} \mathcal{G}^T) (\bar{\mathbf{G}} \bar{\mathbf{B}} \bar{\mathbf{G}}^T + \mathcal{G} \bar{\mathbf{Q}} \mathcal{G}^T + \mathbf{R})^{-1} \mathbf{d}. \quad (103)$$

In 4D-PSAS, it is shown from equation (169) in Appendix A that the solution of initial condition for 4D-PSAS is given as

$$\mathbf{w}_{0|N}^a = \mathbf{w}^b + \mathbf{B} \bar{\mathbf{G}}^T (\bar{\mathbf{G}} \bar{\mathbf{B}} \bar{\mathbf{G}}^T + \mathcal{G} \bar{\mathbf{Q}} \mathcal{G}^T + \mathbf{R})^{-1} \mathbf{d}, \quad (104)$$

and from equations (168) and (166) in Appendix A, the solution for the model errors can be obtained as

$$\bar{\mathbf{b}} = \bar{\mathbf{Q}} \mathcal{G}^T (\bar{\mathbf{G}} \bar{\mathbf{B}} \bar{\mathbf{G}}^T + \mathcal{G} \bar{\mathbf{Q}} \mathcal{G}^T + \mathbf{R})^{-1} \mathbf{d}, \quad (105)$$

where  $\bar{\mathbf{b}}$  is defined as  $\bar{\mathbf{b}} = \begin{pmatrix} \mathbf{0} \\ \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_N \end{pmatrix}$ . Now a forecast integration of equation (59d) can be issued

from the initial condition  $\mathbf{w}_{0|N}^a$ , using equations (104) and (105), to produce the analyses  $\mathbf{w}_{k|N}^a$

for  $k = 0, 1, \dots, N$ . Written in a compact formula, it is easy to show that the analyses  $\mathbf{w}_{k|N}^a$  are given as the same form as equation (103) of Kalman smoother.

**Remarks:**

- From this section, we see that all 4D-VAR, 4D-PSAS, FLKS, Kalman smoother, and the representer algorithms can be derived from the same cost function, especially, 4D-PSAS, the representer, Kalman smoother and FLKS are the solutions of the same discrete Euler-Lagrange equations, which also provides the gradient of the cost function via adjoint model for the 4D-VAR algorithm. If the solution of the problem is unique, and the algorithms converge, then we may expect that the analyses produced by these algorithms are the same. These algorithms are essentially equivalent for linear dynamics in the context of fixed-interval smoothing.
- On the other hand, these algorithms also have their own distinct features. For fixed-interval smoothers, i.e., 4D-VAR, 4D-PSAS, representer algorithm and Kalman smoother, the  $N + 1$  time levels of observations are used simultaneously (equation (90)), and the *a priori* estimation are specified to be the pure model trajectory starting from  $\mathbf{w}^b$ , while for the FLKS the observations are used sequentially, i.e., one time level of observations at a time, and the *a priori* estimation is the current best estimation which incorporates all of the observational information up to and including time  $t_{k-1}$ . These differences determine their different characteristics of implementation in practice.
- If the observations are assumed to occur at every time step, then the minimum of 4D-VAR might be found in the analysis space of dimension  $n \times (N + 1)$ , the minima of 4D-PSAS, the representer and Kalman smoother are solved in the observational space of dimension  $\sum_{k=0}^N p_k$ , and FLKS is solved in the observational space of dimension  $p_k$ . Thus, it is reasonable to expect that FLKS algorithm would be more feasible than the other algorithms. Also, it is seen that 4D-PSAS and the representer algorithm are very similar, so we will focus on 4D-PSAS algorithm thereafter.

Suppose the assimilation time length is  $[t_0, t_m]$ , also we assume that the  $N + 1$  observation times coincide with the analysis time,  $m \geq N$ . It is shown from the results we obtained in this section that, the analyses  $\mathbf{w}_{k|N}^a$  (for  $k = 0, 1, \dots, N$ ) produced by these algorithms, which use all and only the  $N + 1$  time levels of observations, are identical. However, we should be aware that in this case, only one implementation of 4D-VAR or 4D-PSAS is performed over the entire assimilation period, and the FLKS algorithm is actually not fixed lag, i.e., one lag calculation is performed for the first time level of observations, two lag calculations are performed for the second time level of observations, and so on, until  $N$  lag calculations are performed for the last time level of observations. In the other words, the retrospective analysis at time  $t_0$  produced in this way incorporates  $N$  future time levels of observations, at time  $t_1$   $N - 1$  future time levels of observations, ..., at time  $t_{N-1}$  1 future time level of observations, and at time  $t_N$  actually the filter solution.



If a large amount of observations are available, then it is not feasible to implement one 4D-VAR or 4D-PSAS over the entire observational period. Generally, the total time levels of observations are divided into several subsets, each subset contains  $N + 1$  time levels of observations, then one implementation of 4D-VAR or 4D-PSAS will be performed for each subset, that is, the first implementation is over the period  $t_0$  to  $t_N$ , the second implementation is over the period  $t_N$  to  $t_{2N}$ , and so on. However, the analyses generated in this way are not as good as those generated by FLKS algorithm with fixed lag  $N$  except the analyses at times  $t_0$ ,  $t_N$ ,  $t_{2N}$ , etc.. The only practical way to produce the same quality analysis for each point from these algorithms, in which the same time levels  $N + 1$  of observations are used, is to perform moving 4D-VAR and 4D-PSAS at every point with fixed-interval  $N$  given  $\mathbf{w}^b = \mathbf{w}_{k|k-1}^f$  and  $\mathbf{B} = \mathbf{P}_{k|k-1}^f$  and to perform FLKS with fixed lag  $N$ .

### 3.6 Fixed-point smoother perspective

Furthermore, we would like to take a look of FLKS from the fixed-point smoother perspective – the reanalysis at time  $t_{k-l}$  with fixed  $l$  is produced by incorporating future observations at times  $t_{k-l+1}$ ,  $t_{k-l+2}$ ,  $\dots$ ,  $t_k$ . The conditional probability density function of FLKS for fixed lag  $l$ ,  $p(\mathbf{w}_{k-l}^t | \mathbf{W}_k^o)$ , can be written as

$$\begin{aligned} p(\mathbf{w}_{k-l}^t | \mathbf{W}_k^o) &= \frac{1}{p(\mathbf{W}_k^o)} p(\mathbf{w}_{k-l}^t, \mathbf{W}_{k-l-1}^o, \mathbf{w}_{k-l}^o, \mathbf{w}_{k-l+1}^o, \dots, \mathbf{w}_k^o) \\ &= \frac{p(\mathbf{W}_{k-l-1}^o)}{p(\mathbf{W}_k^o)} p(\mathbf{w}_{k-l}^o, \mathbf{w}_{k-l+1}^o, \dots, \mathbf{w}_k^o | \mathbf{w}_{k-l}^t, \mathbf{W}_{k-l-1}^o) p(\mathbf{w}_{k-l}^t | \mathbf{W}_{k-l-1}^o) \\ &= \frac{p(\mathbf{W}_{k-l-1}^o)}{p(\mathbf{W}_k^o)} p(\mathbf{w}_{k-l}^o, \mathbf{w}_{k-l+1}^o, \dots, \mathbf{w}_k^o | \mathbf{w}_{k-l}^t) p(\mathbf{w}_{k-l}^t | \mathbf{W}_{k-l-1}^o). \end{aligned} \quad (106)$$

#### 1. Linear perfect model

For the linear perfect model case, if the observations are independent, then the conditional probability density function can be further simplified as

$$p(\mathbf{w}_{k-l}^t | \mathbf{W}_k^o) = \frac{p(\mathbf{W}_{k-l-1}^o)}{p(\mathbf{W}_k^o)} p(\mathbf{w}_{k-l}^o | \mathbf{w}_{k-l}^t) p(\mathbf{w}_{k-l+1}^o | \mathbf{w}_{k-l}^t) \dots p(\mathbf{w}_k^o | \mathbf{w}_{k-l}^t) p(\mathbf{w}_{k-l}^t | \mathbf{W}_{k-l-1}^o), \quad (107)$$

thus the cost function  $\mathcal{J}_{FP1}$  is of the form

$$\begin{aligned} \mathcal{J}_{FP1} &= \frac{1}{2} (\mathbf{w}_{k-l}^t - \mathbf{w}_{k-l|k-l-1}^f)^T (\mathbf{P}_{k-l|k-l-1}^f)^{-1} (\mathbf{w}_{k-l}^t - \mathbf{w}_{k-l|k-l-1}^f) \\ &\quad + \frac{1}{2} \sum_{i=k-l}^k (\mathbf{w}_i^o - \mathbf{h}_i \mathbf{A}_{i,k-l} \mathbf{w}_{k-l}^t)^T \mathbf{R}_i^{-1} (\mathbf{w}_i^o - \mathbf{h}_i \mathbf{A}_{i,k-l} \mathbf{w}_{k-l}^t). \end{aligned} \quad (108)$$

#### Remark:

Compared with equation (49), equation (108) is identical with the 4D-PSAS cost function  $\mathcal{J}_N$  given  $k = l = N$ ,  $\mathbf{w}_{k-l|k-l-1}^f = \mathbf{w}^b$ , and  $\mathbf{P}_{k-l|k-l-1}^f = \mathbf{B}$ .

If the observations are assimilated sequentially, the conditional probability density function of FLKS for lag  $l$ ,  $p(\mathbf{w}_{k-l}^t | \mathbf{W}_k^o)$ , is given as

$$p(\mathbf{w}_{k-l}^t | \mathbf{W}_k^o) = \frac{p(\mathbf{W}_{k-1}^o)}{p(\mathbf{W}_k^o)} p(\mathbf{w}_k^o | \mathbf{w}_{k-1}^t) p(\mathbf{w}_{k-l}^t | \mathbf{W}_{k-1}^o) \quad (109)$$

with the cost function  $\mathcal{J}_{FP2}$  of the form

$$\begin{aligned} \mathcal{J}_{FP2} &= \frac{1}{2} (\mathbf{w}_{k-l}^t - \mathbf{w}_{k-l|k-1}^a)^T (\mathbf{P}_{k-l|k-1}^a)^{-1} (\mathbf{w}_{k-l}^t - \mathbf{w}_{k-l|k-1}^a)^T \\ &\quad + \frac{1}{2} (\mathbf{w}_k^o - \mathbf{h}_k \mathbf{A}_{k,k-l} \mathbf{w}_{k-l}^t)^T \mathbf{R}_k^{-1} (\mathbf{w}_k^o - \mathbf{h}_k \mathbf{A}_{k,k-l} \mathbf{w}_{k-l}^t). \end{aligned} \quad (110)$$

A direct proof is provided in Appendix B to show that the 4D-PSAS solution is the same as the FLKS solution derived from cost function  $\mathcal{J}_{FP2}$  at  $t_0$ , and 4D-PSAS can be written as a sequential algorithm as well.

## 2. Linear imperfect model

For the linear imperfect model, under the assumption of Gaussian distribution, the information about mean and covariance is needed for  $p(\mathbf{w}_{k-l}^o, \mathbf{w}_{k-l+1}^o, \dots, \mathbf{w}_k^o | \mathbf{w}_{k-l}^t)$  in equation (106). Since we have

$$\begin{aligned} &\mathcal{E}\{\mathbf{w}_{k-l}^o, \mathbf{w}_{k-l+1}^o, \dots, \mathbf{w}_k^o | \mathbf{w}_{k-l}^t\} \\ &= \begin{pmatrix} \mathcal{E}\{\mathbf{w}_{k-l}^o | \mathbf{w}_{k-l}^t\} \\ \mathcal{E}\{\mathbf{w}_{k-l+1}^o | \mathbf{w}_{k-l}^t\} \\ \vdots \\ \mathcal{E}\{\mathbf{w}_k^o | \mathbf{w}_{k-l}^t\} \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{E}\{\mathbf{h}_{k-l} \mathbf{w}_{k-l}^t + \mathbf{b}_{k-l}^o | \mathbf{w}_{k-l}^t\} \\ \mathcal{E}\{\mathbf{h}_{k-l+1} \mathbf{A}_{k-l+1,k-l} \mathbf{w}_{k-l}^t + \mathbf{h}_{k-l+1} \mathbf{b}_{k-l+1} + \mathbf{b}_{k-l+1}^o | \mathbf{w}_{k-l}^t\} \\ \vdots \\ \mathcal{E}\{\mathbf{h}_k \mathbf{A}_{k,k-l} \mathbf{w}_{k-l}^t + \mathbf{h}_k \mathbf{A}_{k,k-l+1} \mathbf{b}_{k-l+1} + \dots + \mathbf{h}_k \mathbf{A}_{k,k-1} \mathbf{b}_{k-1} + \mathbf{h}_k \mathbf{b}_k + \mathbf{b}_k^o | \mathbf{w}_{k-l}^t\} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{h}_{k-l} \mathbf{w}_{k-l}^t \\ \mathbf{h}_{k-l+1} \mathbf{A}_{k-l+1,k-l} \mathbf{w}_{k-l}^t \\ \vdots \\ \mathbf{h}_k \mathbf{A}_{k,k-l} \mathbf{w}_{k-l}^t \end{pmatrix} \\ &= \bar{\mathbf{G}}' \mathbf{w}_{k-l}^t \end{aligned} \quad (111)$$

and the covariance

$$\begin{aligned} &\mathcal{E}\{(\mathbf{W}^o - \mathcal{E}\{\mathbf{W}^o | \mathbf{w}_{k-l}^t\}) (\mathbf{W}^o - \mathcal{E}\{\mathbf{W}^o | \mathbf{w}_{k-l}^t\})^T | \mathbf{w}_{k-l}^t\} \\ &= \mathcal{E}\left\{ \begin{pmatrix} \mathbf{b}_{k-l}^o \\ \mathbf{h}_{k-l+1} \mathbf{b}_{k-l+1} + \mathbf{b}_{k-l+1}^o \\ \vdots \\ \mathbf{h}_k \mathbf{A}_{k,k-l+1} \mathbf{b}_{k-l+1} + \dots + \mathbf{h}_k \mathbf{A}_{k,k-1} \mathbf{b}_{k-1} + \mathbf{h}_k \mathbf{b}_k + \mathbf{b}_k^o \end{pmatrix} \right\} \end{aligned}$$

$$\begin{aligned}
& \left. \begin{pmatrix} & & & \mathbf{b}_{k-l}^o \\ & & & \mathbf{h}_{k-l+1} \mathbf{b}_{k-l+1} + \mathbf{b}_{k-l+1}^o \\ & & \vdots & \\ \mathbf{h}_k \mathbf{A}_{k,k-l+1} \mathbf{b}_{k-l+1} + \dots + \mathbf{h}_k \mathbf{A}_{k,k-1} \mathbf{b}_{k-1} + \mathbf{h}_k \mathbf{b}_k + \mathbf{b}_k^o & & & \end{pmatrix}^T \mathbf{w}_0^t \right\} \\
& = \mathbf{R}' + \mathcal{G}' \bar{\mathbf{Q}}' \mathcal{G}'^T, \tag{112}
\end{aligned}$$

where  $\mathbf{R}'$  is the block diagonal matrix with  $\mathbf{R}_i$  as the diagonal block elements and  $\mathbf{W}^o$  is the observation vector with  $\mathbf{w}_i^o$  as elements for  $i = k-l, k-l+1, \dots, k$ , and

$$\bar{\mathbf{G}}' = \begin{pmatrix} \mathbf{h}_{k-l} \\ \mathbf{h}_{k-l+1} \mathbf{A}_{k-l+1,k-l} \\ \vdots \\ \mathbf{h}_k \mathbf{A}_{k,k-l} \end{pmatrix}, \quad \bar{\mathbf{Q}}' = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{k-l+1} & \dots & \mathbf{0} \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{Q}_k \end{pmatrix},$$

and

$$\mathcal{G}' = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{h}_{k-l+1} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mathbf{0} & \mathbf{h}_k \mathbf{A}_{k,k-l+1} & \mathbf{h}_k \mathbf{A}_{k,k-l+2} & \dots & \mathbf{h}_k \end{pmatrix},$$

then the probability density function  $p(\mathbf{W}^o | \mathbf{w}_{k-l}^t)$  is proportional to

$$p(\mathbf{W}^o | \mathbf{w}_{k-l}^t) \propto \exp \left\{ -\frac{1}{2} (\mathbf{W}^o - \bar{\mathbf{G}}' \mathbf{w}_{k-l}^t)^T (\mathbf{R}' + \mathcal{G}' \bar{\mathbf{Q}}' \mathcal{G}'^T)^{-1} (\mathbf{W}^o - \bar{\mathbf{G}}' \mathbf{w}_{k-l}^t) \right\}. \tag{113}$$

Therefore, the conditional probability density function  $p(\mathbf{w}_{k-l}^t | \mathbf{W}_k^o)$  is given as

$$p(\mathbf{w}_{k-l}^t | \mathbf{W}_k^o) = \text{const} \cdot \exp(-\mathcal{J}_{FP3}), \tag{114}$$

where the cost function  $\mathcal{J}_{FP3}$  is of the form

$$\begin{aligned}
\mathcal{J}_{FP3} &= \frac{1}{2} (\mathbf{w}_{k-l}^t - \mathbf{w}_{k-l|k-l-1}^f)^T (\mathbf{P}_{k-l|k-l-1}^f)^{-1} (\mathbf{w}_{k-l}^t - \mathbf{w}_{k-l|k-l-1}^f) \\
&+ \frac{1}{2} (\mathbf{W}^o - \bar{\mathbf{G}}' \mathbf{w}_{k-l}^t)^T (\mathbf{R}' + \mathcal{G}' \bar{\mathbf{Q}}' \mathcal{G}'^T)^{-1} (\mathbf{W}^o - \bar{\mathbf{G}}' \mathbf{w}_{k-l}^t). \tag{115}
\end{aligned}$$

Given  $k = l = N$ ,  $\mathbf{w}^b = \mathbf{w}_{k-l|k-l-1}^f$  and  $\mathbf{B} = \mathbf{P}_{k-l|k-l-1}^f$ , it is shown that the cost function  $\mathcal{J}_{FP3}$  for  $p(\mathbf{w}_{k-l}^t | \mathbf{W}_k^o)$  is the same as the cost function  $\mathcal{J}_{FP4}$  for  $p(\mathbf{w}_0^t | \mathbf{W}_N^o)$  derived in Appendix A. Also, it is shown in Appendix A that 4D-PSAS cost function  $\mathcal{J}_N$  (43) has the same solution for the state increment as the cost function  $\mathcal{J}_{FP4}$  (173) does.

## 4 The computational aspects of 4D-PSAS and FLKS algorithms

We discussed the solutions of 4D-VAR, 4D-PSAS and FLKS and their relations in the above sections. In this section we will mainly focus on the computational aspects of the numerical algorithms. Since Courtier (1997) discussed the duality between 4D-VAR and 4D-PSAS and

showed that they are equivalent in terms of results produced and cost, we will only compare 4D-PSAS with FLKS in this section. For both 4D-PSAS and FLKS the minimization calculations are performed in the observational space as in GEOS DAS.

Suppose  $m$  time levels of retrospective analysis are needed to produce by assimilating  $N + 1$  time levels of observations at and beyond the analysis time level, that is, the assimilation time length is  $m$ , and  $m \gg N + 1$ . Then, totally  $m + N$  time levels of observations will be assimilated in order to obtain the retrospective analyses at  $m$  analysis time levels. Here we assume that the observation time coincides with the analysis time. We also assume that  $\mathbf{w}_k^b = \mathbf{w}_{k|k-1}^f$  and  $\mathbf{B}_k = \mathbf{P}_{k|k-1}^f$  ( $k = 1, 2, \dots, m$ ) at time  $t_k$  for each implementation of 4D-PSAS or 4D-VAR.

As we pointed out in the previous section, one implementation of 4D-PSAS or 4D-VAR is required for the retrospective analysis at each time level given  $\mathbf{w}_k^b = \mathbf{w}_{k|k-1}^f$  and  $\mathbf{B}_k = \mathbf{P}_{k|k-1}^f$  ( $k = 1, 2, \dots, m$ ) in order to produce the same quality analysis as that produced by FLKS algorithm with fixed lag  $l = N$ . In the following we will give a detailed description of their implementations for the linear perfect and imperfect model cases, respectively.

## 4.1 Linear perfect model

### 1. 4D-PSAS

From equation (54), we see that the solution of 4D-PSAS can be rewritten as

$$\delta \mathbf{x} = \mathbf{B} \bar{\mathbf{G}}^T \mathbf{q}, \quad (116)$$

where the  $N_p$  - vector  $\mathbf{q}$  ( $N_p = \sum_{k=0}^N p_k$ ) is the vector of analysis increment in observation space, satisfying

$$(\bar{\mathbf{G}} \mathbf{B} \bar{\mathbf{G}}^T + \mathbf{R}) \mathbf{q} = \mathbf{d}. \quad (117)$$

In 4D-PSAS, the  $N_p \times N_p$  linear system (117) is solved by minimizing the functional  $\mathcal{F}$

$$\mathcal{F}(\mathbf{q}) = \frac{1}{2} \mathbf{q}^T (\bar{\mathbf{G}} \mathbf{B} \bar{\mathbf{G}}^T + \mathbf{R}) \mathbf{q} - \mathbf{q}^T \mathbf{d}. \quad (118)$$

Defining the  $N_p$  - vector  $\mathbf{s} = \bar{\mathbf{G}} \mathbf{B} \bar{\mathbf{G}}^T \mathbf{q}$ , which can be calculated as

$$\begin{pmatrix} \mathbf{s}_0 \\ \mathbf{s}_1 \\ \vdots \\ \mathbf{s}_N \end{pmatrix} = \begin{pmatrix} \mathbf{h}_0 \mathbf{B} \mathbf{T} \\ \mathbf{h}_1 \mathbf{A}_1 \mathbf{B} \mathbf{T} \\ \vdots \\ \mathbf{h}_N \mathbf{A}_N \mathbf{B} \mathbf{T} \end{pmatrix}, \quad (119)$$

where

$$\mathbf{T} = \mathbf{h}_0^T \mathbf{q}_0 + \mathbf{A}_1^T \mathbf{h}_1^T \mathbf{q}_1 + \dots + \mathbf{A}_N^T \mathbf{h}_N^T \mathbf{q}_N.$$

**Algorithm 1. 4D-PSAS:**

- (1) Specify the initial guess of vector  $\mathbf{q}_{N_p \times 1}$ .
- (2) An iterative minimization method (e.g. conjugate gradient method) is employed to solve the  $N_p \times N_p$  linear system (117) for quantity  $\mathbf{q}$ , in which the vector  $\mathbf{s}$  and the values of the functional  $\mathcal{F}(\mathbf{q})$  and its gradient need to be evaluated at each iteration as following:
  - (a) Integrate the adjoint model backward in time with null initial condition for the adjoint variable  $\mathbf{p}$  with the forcing term  $\mathbf{h}_i^T \mathbf{q}_i$  at time  $t_i$ . Then, multiply the result of the adjoint integration by  $\mathbf{B}$ , we denote it by  $\hat{\mathbf{z}}_0$ .
  - (b) Integrate the tangent linear model with  $\hat{\mathbf{z}}_0$  as the initial condition. At each time  $t_i$ , compute
 
$$\mathbf{s}_i = \mathbf{h}_i \hat{\mathbf{z}}_i .$$
  - (c) Calculate the values of the functional  $\mathcal{F}(\mathbf{q})$  and the gradient of the functional.
- (3) Integrate the adjoint model backward in time for the adjoint variable  $\mathbf{p}$  with the forcing term  $\mathbf{h}_i^T \mathbf{q}_i$  at time  $t_i$ . Then the retrospective analysis increment at time  $t_0$  is the multiplication of the result of the adjoint integration with  $\mathbf{B}$ .

We can see that one implementation of 4D-PSAS for each time level of retrospective analysis needs one application of modified PSAS (with the integrations of tangent linear model and adjoint model embodied) to a large problem (equation (118) ) with the control variable's size of  $N_p \times 1$ . The computational procedure for one implementation of 4D-PSAS is described as in Algorithm 1.

Totally,  $m$  applications of modified PSAS to a larger problem with control variable  $(\mathbf{q})_{N_p \times 1}$  are needed for the retrospective analyses over the entire assimilation time length. Each implementation of 4D-PSAS requires the memory storage for  $(\mathbf{d})_{N_p \times 1}$ ,  $(\mathbf{q})_{N_p \times 1}$ ,  $(\mathbf{w}^b)_{n \times 1}$ , and the memory storage or calculations of  $(\mathbf{B})_{n \times n}$  and  $(\mathbf{R}_k)_{p_k \times p_k}$  where  $k = 0, 1, \dots, N$ .

## 2. FLKS

The FLKS algorithm for linear perfect model case is derived in section 2.2:

$$\begin{aligned}
 \mathbf{w}_{k-1|k}^a &= \mathbf{w}_{k-1|k-1}^a + \mathbf{A}_{k,k-1}^{-1} \mathbf{P}_{k|k-1}^f \mathbf{h}_k^T \mathbf{\Gamma}_k^{-1} \mathbf{v}_k , \\
 \mathbf{w}_{k-2|k}^a &= \mathbf{w}_{k-2|k-1}^a + \mathbf{A}_{k,k-2}^{-1} \mathbf{P}_{k|k-1}^f \mathbf{h}_k^T \mathbf{\Gamma}_k^{-1} \mathbf{v}_k , \\
 &\vdots \\
 \mathbf{w}_{k-l|k}^a &= \mathbf{w}_{k-l|k-1}^a + \mathbf{A}_{k,k-l}^{-1} \mathbf{P}_{k|k-1}^f \mathbf{h}_k^T \mathbf{\Gamma}_k^{-1} \mathbf{v}_k .
 \end{aligned}$$

It is seen from equation (39a) that the term  $\mathbf{P}_{k|k-1}^f \mathbf{h}_k^T \mathbf{\Gamma}_k^{-1} \mathbf{v}_k$  is equal to the analysis increment  $\mathbf{w}_{k|k}^a - \mathbf{w}_{k|k-1}^f$  of the filter portion, which is already available. It is one application of PSAS (Da Silva et al., 1996), whose algorithm consists of solving one  $p_k \times p_k$  linear system for the  $p_k$  - vector  $\mathbf{q}_k$

$$\mathbf{\Gamma}_k \mathbf{q}_k = \mathbf{v}_k ,$$

and subsequently evaluating the matrix-vector multiplication  $\mathbf{P}_{k|k-1}^f \mathbf{h}_k^T \mathbf{q}_k$ , then obtaining the analysis  $\mathbf{w}_{k|k}^a$  from the equation

$$\mathbf{w}_{k|k}^a = \mathbf{w}_{k|k-1}^f + \mathbf{P}_{k|k-1}^f \mathbf{h}_k^T \mathbf{q}_k .$$

In PSAS the vector  $\mathbf{q}_k$  is solved by a conjugate gradient method which minimizes the functional  $\mathcal{F}(\mathbf{q}_k)$ :

$$\mathcal{F}(\mathbf{q}_k) = \frac{1}{2} \mathbf{q}_k^T (\mathbf{h}_k \mathbf{P}_{k|k-1}^f \mathbf{h}_k^T + \mathbf{R}_k) \mathbf{q}_k - \mathbf{q}_k^T \mathbf{v}_k . \quad (121)$$

It is obvious that, for linear perfect model, only one application of PSAS in a space of dimension  $p_k \times 1$  is needed for the filter analysis at each observational time level, no more PSAS is necessary for the retrospective analyses.

Therefore, the algorithm of FLKS in the linear perfect model case comprises one application of the quasi-inverse model to the analysis increment (Pu et al., 1997), or solving a linear system for a  $p_k$ -vector  $\mathbf{f}_{k-i}$

$$\mathbf{A}_{k,k-i} \mathbf{f}_{k-i} = \mathbf{P}_{k|k-1}^f \mathbf{h}_k^T \Gamma_k^{-1} \mathbf{v}_k ,$$

then calculating the retrospective analysis by

$$\mathbf{w}_{k-i|k}^a = \mathbf{w}_{k-i|k-1}^a + \mathbf{A}_{k,k-i}^{-1} \mathbf{P}_{k|k-1}^f \mathbf{h}_k^T \Gamma_k^{-1} \mathbf{v}_k .$$

Since  $m + N$  time levels of observations are needed to be assimilated in order to obtain the retrospective analyses at  $m$  analysis time levels, total  $m + N$  integrations of quasi-inverse model will have to be carried out. The memory storage required is for the analysis increments,  $(\mathbf{w}_{k-i|k-1}^a)_{n \times 1}$ , and the memory storage or calculations are also required for  $(\mathbf{P}_{k|k-1}^f)_{n \times n}$  and  $(\mathbf{R}_k)_{p_k \times p_k}$ . No additional application of PSAS is needed in the smoother portion. Therefore, it is reasonable to expect that the implementation of FLKS algorithm is much cheaper than that of 4D-PSAS algorithm.

## 4.2 Linear imperfect model

### 1. 4D-PSAS

The solution of the 4D-PSAS formula for the linear imperfect case is given as

$$\mathbf{z} = \mathbf{D} \mathbf{G}^T \mathbf{q} , \quad (122)$$

where  $N_p$ -vector  $\mathbf{q}$  satisfies

$$(\mathbf{G} \mathbf{D} \mathbf{G}^T + \mathbf{R}) \mathbf{q} = \mathbf{d} . \quad (123)$$

Equation (123) is solved by minimizing the functional  $\mathcal{F}$

$$\mathcal{F} = \frac{1}{2} \mathbf{q}^T (\mathbf{G} \mathbf{D} \mathbf{G}^T + \mathbf{R}) \mathbf{q} - \mathbf{q}^T \mathbf{d} . \quad (124)$$

Introducing  $\mathbf{s} = \mathbf{G}\mathbf{D}\mathbf{G}^T\mathbf{q}$ , since

$$\mathbf{G} = \begin{pmatrix} \mathbf{h}_0 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{h}_1\mathbf{A}_1 & \mathbf{h}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{h}_2\mathbf{A}_2 & \mathbf{h}_2\mathbf{A}_{2,1} & \mathbf{h}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mathbf{h}_N\mathbf{A}_N & \mathbf{h}_N\mathbf{A}_{N,1} & \mathbf{h}_N\mathbf{A}_{N,2} & \dots & \mathbf{h}_N \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} \mathbf{B} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \dots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{Q}_N \end{pmatrix},$$

hence, the vector  $\mathbf{s}$  is calculated as

$$\begin{pmatrix} \mathbf{s}_0 \\ \mathbf{s}_1 \\ \mathbf{s}_2 \\ \vdots \\ \mathbf{s}_N \end{pmatrix} = \begin{pmatrix} \mathbf{h}_0\mathbf{B}\mathbf{T} \\ \mathbf{h}_1\mathbf{A}_1\mathbf{B}\mathbf{T} + \mathbf{h}_1\mathbf{Q}_1\mathbf{E}_1 \\ \mathbf{h}_2\mathbf{A}_2\mathbf{B}\mathbf{T} + \mathbf{h}_2\mathbf{A}_{2,1}\mathbf{Q}_1\mathbf{E}_1 + \mathbf{h}_2\mathbf{Q}_2\mathbf{E}_2 \\ \vdots \\ \mathbf{h}_N\mathbf{A}_N\mathbf{B}\mathbf{T} + \mathbf{h}_N\mathbf{A}_{N,1}\mathbf{Q}_1\mathbf{E}_1 + \mathbf{h}_N\mathbf{A}_{N,2}\mathbf{Q}_2\mathbf{E}_2 + \dots + \mathbf{h}_N\mathbf{Q}_N\mathbf{E}_N \end{pmatrix}. \quad (125)$$

Here  $\mathbf{T}$ ,  $\mathbf{E}_1$ ,  $\mathbf{E}_2$ , ..., and  $\mathbf{E}_N$  are defined as

$$\mathbf{T} = \mathbf{h}_0^T\mathbf{q}_0 + \mathbf{A}_1^T\mathbf{h}_1^T\mathbf{q}_1 + \dots + \mathbf{A}_N^T\mathbf{h}_N^T\mathbf{q}_N,$$

$$\mathbf{E}_1 = \sum_{i=1}^N \mathbf{A}_{i,1}^T\mathbf{h}_i^T\mathbf{q}_i,$$

$$\mathbf{E}_2 = \sum_{i=2}^N \mathbf{A}_{i,2}^T\mathbf{h}_i^T\mathbf{q}_i,$$

\vdots

$$\mathbf{E}_N = \mathbf{h}_N^T\mathbf{q}_N.$$

As pointed out in Courtier (1997), the dimension of the control variable is the same as in the linear perfect model case, but one has to store the adjoint variable  $\mathbf{p}_i$  at time  $t_i$  which is used to evaluate the forcing  $\mathbf{Q}_i\mathbf{p}_i$  of the subsequent tangent linear integration.

### Algorithm 3. 4D-PSAS:

- (1) Specify the initial guess of vector  $\mathbf{q}_{N_p \times 1}$ .
- (2) An iterative minimization method (e.g. conjugate gradient method) is employed to solve the  $N_p \times N_p$  linear system (123) for quantity  $\mathbf{q}$ , in which the vector  $\mathbf{s}$  and the values of the functional  $\mathcal{F}(\mathbf{q})$  and its gradient need to be evaluated at each iteration as following:
  - (a) Integrate the adjoint model backward in time for the adjoint variable  $\mathbf{p}$  with  $\mathbf{h}_i^T\mathbf{q}_i$  as forcing term at time  $t_i$ , and store the adjoint variable  $\mathbf{p}_i$ . Then, multiply the result of the adjoint integration by  $\mathbf{B}$ , we denote it by  $\hat{\mathbf{z}}_0$ .
  - (b) Integrate the tangent linear model with  $\hat{\mathbf{z}}_0$  as the initial condition and  $\mathbf{Q}_i\mathbf{p}_i$  as forcing. At time  $t_i$ , compute

$$\mathbf{s}_i = \mathbf{h}_i\hat{\mathbf{z}}_i.$$

- (c) Calculate the values of the functional  $\mathcal{F}(\mathbf{q})$  and the gradient of the functional.
- (3) Integrate the adjoint model backward in time for the adjoint variable  $\mathbf{p}$  with  $\mathbf{h}_i^T \mathbf{q}_i$  as forcing term at time  $t_i$ , and store the adjoint variable  $\mathbf{p}_i$ . Then, the retrospective analysis increment at time  $t_0$  is obtained by multiplying the result of the adjoint integration by  $\mathbf{B}$ , and the model error  $\mathbf{b}_i$  at time  $t_i$  is equal to  $\mathbf{Q}_i \mathbf{p}_i$ .

It is shown that one application of a modified PSAS (with the integrations of the tangent linear model and adjoint model embodied) to a large problem equation (124) has to be performed to obtain the retrospective analysis at one time level. The size of the control variable  $\mathbf{q}$  in this problem is of  $N_p \times 1$ . Totally,  $m$  applications of modified PSAS are needed to produce the retrospective analyses for the entire assimilation time length.

It is noticed that the dimensions of the control variable of problem (124) and (118) are the same, and the total applications of modified PSAS is also the same as that in the 4D-PSAS linear perfect model case. However, compared to the linear perfect model case, much more memory storage is required in the 4D-PSAS linear imperfect model case for all of the intermediate adjoint variables  $\mathbf{p}_i$  and for all of the model error covariances (or one needs to prescribe them) at each model integration time step. In this study the observations are assumed to occur at every analysis time, we should be aware that generally the total number of model integration time step is larger than the total time levels of the observations we use. Therefore, the requirement of the memory storage for the model errors might be tremendously large.

## 2. FLKS

The numerical formula of FLKS in the GEOS DAS framework are presented in section 2.2, which are of the form:

$$\mathbf{w}_{k|k}^a = \mathbf{w}_{k|k-1}^f + \mathbf{P}_{k|k-1}^f \mathbf{g}_0, \quad (126a)$$

$$\mathbf{w}_{k-1|k}^a = \mathbf{w}_{k-1|k-1}^a + \mathbf{P}_{k-1|k-2}^f \mathbf{g}_1, \quad (126b)$$

$$\mathbf{w}_{k-2|k}^a = \mathbf{w}_{k-2|k-1}^a + \mathbf{P}_{k-2|k-3}^f \mathbf{g}_2, \quad (126c)$$

⋮

$$\mathbf{w}_{k-l|k}^a = \mathbf{w}_{k-l|k-1}^a + \mathbf{P}_{k-l|k-l-1}^f \mathbf{g}_l. \quad (126d)$$

where  $\mathbf{\Gamma}_k, \mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_l$  are given by equations (23) and (35), respectively,

$$\mathbf{\Gamma}_k = \mathbf{h}_k \mathbf{P}_{k|k-1}^f \mathbf{h}_k^T + \mathbf{R}_k, \quad (127a)$$

$$\begin{aligned} \mathbf{g}_0 &= \mathbf{h}_k^T \mathbf{\Gamma}_k^{-1} \mathbf{v}_k \\ &= \mathbf{h}_k^T \mathbf{q}_0, \end{aligned} \quad (127b)$$

$$\begin{aligned} \mathbf{g}_1 &= \left[ \mathbf{I} - \mathbf{h}_{k-1}^T \mathbf{\Gamma}_{k-1}^{-1} \mathbf{h}_{k-1} \mathbf{P}_{k-1|k-2}^f \right] \mathbf{A}_{k,k-1}^T \mathbf{g}_0 \\ &= \mathbf{A}_{k,k-1}^T \mathbf{g}_0 - \mathbf{h}_{k-1}^T \mathbf{q}_1, \end{aligned} \quad (127c)$$

⋮



$$\begin{aligned}
\mathbf{g}_l &= \left[ \mathbf{I} - \mathbf{h}_{k-l}^T \Gamma_{k-l}^{-1} \mathbf{h}_{k-l} \mathbf{P}_{k-l|k-l-1}^f \right] \mathbf{A}_{k-l+1,k-l}^T \mathbf{g}_{l-1} \\
&= \mathbf{A}_{k-l+1,k-l}^T \mathbf{g}_{l-1} - \mathbf{h}_{k-l}^T \mathbf{q}_l .
\end{aligned} \tag{127d}$$

Here the  $p_k$  - vectors  $\mathbf{q}_i$  for  $i = 0, 1, \dots, l$  satisfy the following linear systems

$$\Gamma_k \mathbf{q}_0 = \mathbf{v}_k , \tag{128a}$$

$$\Gamma_{k-1} \mathbf{q}_1 = \mathbf{h}_{k-1} \mathbf{P}_{k-1|k-2}^f \mathbf{A}_{k,k-1}^T \mathbf{g}_0 , \tag{128b}$$

⋮

$$\Gamma_{k-l} \mathbf{q}_l = \mathbf{h}_{k-l} \mathbf{P}_{k-l|k-l-1}^f \mathbf{A}_{k-l+1,k-l}^T \mathbf{g}_{l-1} . \tag{128c}$$

In the GEOS DAS framework, the  $l+1$  linear systems for the  $p_k$  - vectors  $\mathbf{q}_i$  for  $i = 0, 1, \dots, l$ , including one for the filter portion (equation (128a)), are solved iteratively by minimizing the functionals  $\mathcal{F}(\mathbf{q}_i)$ , respectively

$$\mathcal{F}(\mathbf{q}_0) = \frac{1}{2} \mathbf{q}_0^T (\mathbf{h}_k \mathbf{P}_{k|k-1}^f \mathbf{h}_k^T + \mathbf{R}_k) \mathbf{q}_0 - \mathbf{q}_0^T \mathbf{v}_k , \tag{129a}$$

$$\begin{aligned}
\mathcal{F}(\mathbf{q}_1) &= \frac{1}{2} \mathbf{q}_1^T (\mathbf{h}_{k-1} \mathbf{P}_{k-1|k-2}^f \mathbf{h}_{k-1}^T + \mathbf{R}_{k-1}) \mathbf{q}_1 \\
&\quad - \mathbf{q}_1^T \mathbf{h}_{k-1} \mathbf{P}_{k-1|k-2}^f \mathbf{A}_{k,k-1}^T \mathbf{g}_0 ,
\end{aligned} \tag{129b}$$

⋮

$$\begin{aligned}
\mathcal{F}(\mathbf{q}_l) &= \frac{1}{2} \mathbf{q}_l^T (\mathbf{h}_{k-l} \mathbf{P}_{k-l|k-l-1}^f \mathbf{h}_{k-l}^T + \mathbf{R}_{k-l}) \mathbf{q}_l \\
&\quad - \mathbf{q}_l^T \mathbf{h}_{k-l} \mathbf{P}_{k-l|k-l-1}^f \mathbf{A}_{k-l+1,k-l}^T \mathbf{g}_{l-1} .
\end{aligned} \tag{129c}$$

It is shown that  $l$  applications of PSAS, are employed for  $l$  lags at each observational time level, with each application being to a small problem with control variable's dimension to be of  $p_k \times 1$ .

#### Algorithm 4. FLKS:

At each observational time level  $k$ , with the availability of  $\mathbf{g}_0$  from the filter portion, for lag  $i = 1, 2, \dots, l$ ,

- (1) Carry out one application of PSAS with modified forcing term as in equations (128b) - (128c), which consists of
  - (a) integrating the adjoint model backward in time from  $t_{k-i}$  to  $t_{k-i+1}$  with  $\mathbf{g}_{i-1}$  as forcing, and storing the result as  $\mathbf{p}$ . Then, multiplying  $\mathbf{p}$  by  $\mathbf{h}_{k-i} \mathbf{P}_{k-i|k-i-1}^f$ , we denote the result as  $\mathbf{c}_i$ .
  - (b) using an iterative minimization procedure to solve one  $p_k \times p_k$  linear system for the  $p_k$ -vector  $\mathbf{q}_i$

$$\Gamma_{k-i} \mathbf{q}_i = \mathbf{c}_i ,$$

and subsequently computing the matrix-vector multiplication  $\mathbf{P}_{k-i|k-i-1}^f \mathbf{h}_{k-i}^T \mathbf{q}_i$ .

- (2) Evaluate the matrix-vector multiplication  $\mathbf{P}_{k-i|k-i-1}^f \mathbf{p}$ , then obtain the retrospective analysis  $\mathbf{w}_{k-i|k}^a$  from the equation

$$\mathbf{w}_{k-i|k}^a = \mathbf{w}_{k-i|k-1}^a + \mathbf{P}_{k-i|k-i-1}^f \mathbf{p} - \mathbf{P}_{k-i|k-i-1}^f \mathbf{h}_{k-i}^T \mathbf{q}_i .$$

- (3) Go back to step (1) for the next lag.

**Remark:**

- At each observational time level  $k$ , one application of PSAS is needed for each lag  $i$ , where  $i = 1, 2, \dots, l$ , that is, given fixed lag  $l = N$ ,  $N$  applications of PSAS are needed for the retrospective analysis. Each PSAS is applied to a small problem with the control variable  $\mathbf{q}_i$  of dimension  $p_k \times 1$ .
- In doing the retrospective analysis over the assimilation time length  $m$  with fixed-lag  $N$ , it is necessary to carry out 1 application of PSAS for lag 1 at the 1st observational time level, 2 applications of PSAS for lags 1 and 2 at the 2nd observational time level, and so on, until  $N$  applications of PSAS for lags 1, 2, ...,  $N$  at the  $N$ th observational time level, then  $N$  applications of PSAS for lags 1, 2, ...,  $N$  at each of the rest  $(N+k)$ -th observational time level where  $k = 1, 2, \dots, m$ . Therefore, the total number of applications of such PSAS is  $N \left( \frac{N+1}{2} + m \right)$  for  $m \geq N$ . Generally,  $m \gg N$  in the practical purpose for the retrospective analysis. Hence, the total number of applications of such PSAS is roughly  $N \times m$ .

Compared with 4D-PSAS algorithm, FLKS algorithm requires more applications of PSAS but to smaller problems, which generally is more feasible. Comparing equation (123) with (128b)-(128c), FLKS looks like a special case of 4D-PSAS with a block diagonal matrix. Moreover, FLKS requires much less memory storage than 4D-PSAS algorithm. Therefore, FLKS algorithm is more suitable for doing retrospective analysis in the GEOS DAS framework from a scientific and computational standpoint.

## 5 Extensions to nonlinear model

It is clear that if the forward operator is not a linear function of  $\mathbf{w}_k^t$ , then the posterior probability density function  $p$  is not Gaussian. The more nonlinear the forward operator is, the more remote is  $p$  from a Gaussian function. Tarantola (1987) presented a detailed discussion about how the nonlinearity affects the posterior probability density function  $p$  away from a Gaussian function.

In this section we extend the linear FLKS and 4D-PSAS algorithms to nonlinear cases, discussing two cases with different degrees of nonlinearity for FLKS and 4D-PSAS, respectively.

## 5.1 FLKS algorithm

As shown in section 2.1, the cost function  $\mathcal{J}_{FLKS}$  for FLKS algorithm is given as

$$\begin{aligned} \mathcal{J}_{FLKS} &= \frac{1}{2}(\mathbf{W}^t - \bar{\mathbf{W}})^T (\mathbf{P}^t)^{-1} (\mathbf{W}^t - \bar{\mathbf{W}}) \\ &\quad + \frac{1}{2}(\mathbf{w}_k^o - \mathbf{h}_k(\mathbf{w}_k^t))^T \mathbf{R}_k^{-1} (\mathbf{w}_k^o - \mathbf{h}_k(\mathbf{w}_k^t)) . \end{aligned}$$

If  $\mathbf{h}_k(\mathbf{w}_k^t)$  can be linearized around  $\mathbf{w}_{k|k-1}^f$ , i.e.,

$$\begin{aligned} \mathbf{h}_k(\mathbf{w}_k^t) &\simeq \mathbf{h}_k(\mathbf{w}_{k|k-1}^f) + \bar{\mathbf{h}}_k(\mathbf{w}_k^t - \mathbf{w}_{k|k-1}^f) \\ &= \mathbf{h}_k(\mathbf{w}_{k|k-1}^f) + \mathcal{H}_0(\mathbf{W}^t - \bar{\mathbf{W}}) , \end{aligned} \quad (130)$$

where

$$\bar{\mathbf{h}}_k = \left( \frac{\partial \mathbf{h}_k}{\partial \mathbf{w}_k^t} \right)_{\mathbf{w}_{k|k-1}^f} \quad (131)$$

and  $\mathcal{H}_0 = (\bar{\mathbf{h}}_k, \mathbf{0}, \dots, \mathbf{0})$ . This is the weakest nonlinearity case. The a posteriori probability density function is approximately Gaussian, with its maximum likelihood point being given as

$$\mathbf{W}^a = \bar{\mathbf{W}} + \mathbf{P}^t \mathcal{H}_0^T (\mathcal{H}_0 \mathbf{P}^t \mathcal{H}_0^T + \mathbf{R}_k)^{-1} (\mathbf{w}_k^o - \mathbf{h}_k(\mathbf{w}_{k|k-1}^f)) , \quad (132)$$

and the a posteriori covariance being given as

$$\mathbf{P} = \left[ (\mathbf{P}^t)^{-1} + \mathcal{H}_0^T \mathbf{R}_k^{-1} \mathcal{H}_0 \right]^{-1} . \quad (133)$$

Similar to the derivations in the linear case, the gain matrices of the FLKS can be derived as

$$\mathcal{K}_{k|k} = \mathbf{P}_{k|k-1}^f \bar{\mathbf{h}}_k^T \mathbf{\Gamma}_k^{-1} , \quad (134a)$$

$$\mathcal{K}_{k-1|k} = \mathbf{P}_{k-1,k|k-1}^{af} \bar{\mathbf{h}}_k^T \mathbf{\Gamma}_k^{-1} , \quad (134b)$$

$\vdots$

$$\mathcal{K}_{k-l|k} = \mathbf{P}_{k-l,k|k-1}^{af} \bar{\mathbf{h}}_k^T \mathbf{\Gamma}_k^{-1} , \quad (134c)$$

where

$$\mathbf{\Gamma}_k = \bar{\mathbf{h}}_k \mathbf{P}_{k|k-1}^f \bar{\mathbf{h}}_k^T + \mathbf{R}_k , \quad (135)$$

then the FLKS analysis equation (132) can be rewritten as

$$\mathbf{w}_{k|k}^a = \mathbf{w}_{k|k-1}^f + \mathcal{K}_{k|k} \mathbf{v}_k , \quad (136a)$$

$$\mathbf{w}_{k-1|k}^a = \mathbf{w}_{k-1|k-1}^a + \mathcal{K}_{k-1|k} \mathbf{v}_k , \quad (136b)$$

$\vdots$

$$\mathbf{w}_{k-l|k}^a = \mathbf{w}_{k-l|k-1}^a + \mathcal{K}_{k-l|k} \mathbf{v}_k , \quad (136c)$$

where  $\mathbf{v}_k$  is the innovation vector defined as

$$\mathbf{v}_k = \mathbf{w}_k^o - \mathbf{h}_k(\mathbf{w}_{k|k-1}^f) . \quad (137)$$

The error covariance equation (133) can also be expanded as

$$\mathbf{P}_{k|k}^a = (\mathbf{I} - \mathcal{K}_{k|k} \bar{\mathbf{h}}_k) \mathbf{P}_{k|k-1}^f, \quad (138a)$$

$$\mathbf{P}_{k-l|k}^a = \mathbf{P}_{k-l|k-1}^a - \mathcal{K}_{k-l|k} \bar{\mathbf{h}}_k \mathbf{P}_{k,k-l|k-1}^{fa}, \quad (138b)$$

$$\mathbf{P}_{k,k-l|k}^{aa} = (\mathbf{I} - \mathcal{K}_{k|k} \bar{\mathbf{h}}_k) \mathbf{P}_{k,k-l|k-1}^{fa}. \quad (138c)$$

If we also assume  $\mathbf{A}_{k,k-1}(\mathbf{w}_{k-1}^t)$  can be linearizable around  $\mathbf{w}_{k-1|k-1}^a$ , then

$$\begin{aligned} \mathbf{P}_{k,k-l|k-1}^{fa} &= \mathcal{E} \left\{ \mathbf{e}_{k|k-1}^f (\mathbf{e}_{k-l|k-1}^a)^T \right\} \\ &= \mathcal{E} \left\{ \mathcal{A}_{k,k-1} e_{k-1|k-1}^a (e_{k-l|k-1}^a)^T \right\} \\ &= \mathcal{A}_{k,k-1} \mathbf{P}_{k-1,k-l|k-1}^{aa} \\ &= \left( \mathbf{P}_{k-l,k|k-1}^{af} \right)^T, \end{aligned} \quad (139)$$

where

$$\mathcal{A}_{k,k-1} = \left( \frac{\partial \mathbf{A}_{k,k-1}}{\partial \mathbf{w}_{k-1}^t} \right)_{\mathbf{w}_{k-1|k-1}^a}.$$

For this weakest nonlinearity case, we see that solving such a problem is not more difficult than solving a linear problem. The algorithm is the same as that for the strict linear case except that  $\mathcal{A}$  and  $\bar{\mathbf{h}}$  replace  $\mathbf{A}$  and  $\mathbf{h}$ , respectively.

If the linearization of  $\mathbf{h}_k(\mathbf{w}_k^t)$  around  $\mathbf{w}_{k|k-1}^f$  is no longer acceptable, but it is still linearizable in the region of significant posterior probability density, i.e., quasi-linear around the true maximum likelihood point  $\mathbf{w}_{k|k}^a$ ,

$$\begin{aligned} \mathbf{h}_k(\mathbf{w}_k^t) &\simeq \mathbf{h}_k(\mathbf{w}_{k|k}^a) + \bar{\mathbf{h}}_\infty (\mathbf{w}_k^t - \mathbf{w}_{k|k}^a) \\ &= \mathbf{h}_k(\mathbf{w}_{k|k}^a) + \mathcal{H}_\infty (\mathbf{W}^t - \mathbf{W}^a), \end{aligned} \quad (140)$$

where  $\mathcal{H}_\infty = (\bar{\mathbf{h}}_\infty, \mathbf{0}, \dots, \mathbf{0})$  and

$$\bar{\mathbf{h}}_\infty = \left( \frac{\partial \mathbf{h}_k}{\partial \mathbf{w}_k^t} \right)_{\mathbf{w}_{k|k}^a}, \quad (141)$$

then the maximum likelihood point of this case is given as

$$\mathbf{W}^a = \bar{\mathbf{W}} + \mathbf{P}^t \mathcal{H}_\infty^T (\mathcal{H}_\infty \mathbf{P}^t \mathcal{H}_\infty^T + \mathbf{R}_k)^{-1} \left( (\mathbf{w}_k^o - \mathbf{h}_k(\mathbf{w}_{k|k}^a)) + \mathcal{H}_\infty (\mathbf{W}^a - \bar{\mathbf{W}}) \right). \quad (142)$$

Defining the gain matrices of the FLKS as

$$\mathcal{K}_{k|k} = \mathbf{P}_{k|k-1}^f \bar{\mathbf{h}}_\infty^T \Gamma_k^{-1}, \quad (143a)$$

$$\mathcal{K}_{k-1|k} = \mathbf{P}_{k-1,k|k-1}^{af} \bar{\mathbf{h}}_\infty^T \Gamma_k^{-1}, \quad (143b)$$

$\vdots$

$$\mathcal{K}_{k-l|k} = \mathbf{P}_{k-l,k|k-1}^{af} \bar{\mathbf{h}}_\infty^T \Gamma_k^{-1}, \quad (143c)$$

where

$$\Gamma_k = \bar{\mathbf{h}}_\infty \mathbf{P}_{k|k-1}^f \bar{\mathbf{h}}_\infty^T + \mathbf{R}_k, \quad (144)$$

then the FLKS analysis equation (142) can be rewritten as

$$\mathbf{w}_{k|k}^a = \mathbf{w}_{k|k-1}^f + \mathcal{K}_{k|k} \mathbf{v}_k, \quad (145a)$$

$$\mathbf{w}_{k-1|k}^a = \mathbf{w}_{k-1|k-1}^a + \mathcal{K}_{k-1|k} \mathbf{v}_k, \quad (145b)$$

⋮

$$\mathbf{w}_{k-l|k}^a = \mathbf{w}_{k-l|k-1}^a + \mathcal{K}_{k-l|k} \mathbf{v}_k, \quad (145c)$$

where  $\mathbf{v}_k$  is of the form

$$\mathbf{v}_k = \mathbf{w}_k^o - \mathbf{h}_k(\mathbf{w}_{k|k}^a) + \bar{\mathbf{h}}_\infty(\mathbf{w}_{k|k}^a - \mathbf{w}_{k|k-1}^f). \quad (146)$$

It is seen that  $\mathbf{w}_{k|k}^a$  is also implicitly involved in equation (145a). Usually, a nonlinear iterative procedure is employed for equation (145a), for instance,

$$\begin{aligned} \mathbf{w}_{k|k,r+1}^a &= \mathbf{w}_{k|k-1}^f + \mathbf{P}_{k|k-1}^f \bar{\mathbf{h}}_r^T (\bar{\mathbf{h}}_r \mathbf{P}_{k|k-1}^f \bar{\mathbf{h}}_r^T + \mathbf{R}_k)^{-1} \\ &\quad \cdot \left[ (\mathbf{w}_k^o - \mathbf{h}_k(\mathbf{w}_{k|k,r}^a)) + \bar{\mathbf{h}}_r (\mathbf{w}_{k|k,r}^a - \mathbf{w}_{k|k-1}^f) \right], \end{aligned} \quad (147)$$

where  $r$  denotes the  $r$ -th iteration, and

$$\bar{\mathbf{h}}_r = \left( \frac{\partial \mathbf{h}_k}{\partial \mathbf{w}_k^a} \right)_{\mathbf{w}_{k|k,r}^a}. \quad (148)$$

Once the maximum likelihood point has been approached, the a posteriori covariance can be computed as

$$\mathbf{P} \simeq \left[ (\mathbf{P}^t)^{-1} + \mathcal{H}_\infty^T \mathbf{R}_k^{-1} \mathcal{H}_\infty \right]^{-1}. \quad (149)$$

Same equations for error covariances can be obtained as equations (138a) – (138c) in the previous case except that  $\mathbf{h}_k$  is replaced by  $\bar{\mathbf{h}}_\infty$ .

The calculation of  $\mathbf{P}_{k,k-l|k-1}^{fa}$  can also be based on the linearization of  $\mathbf{A}_{k,k-1}(\mathbf{w}_{k-1}^t)$  about  $\mathbf{w}_{k-1|k-1}^a$  as equation (139) in the weakest nonlinearity case. The function can also be linearized around the latest lag results available, it depends on whether we want it to be consistent with the filter portion, and to be consistent among the calculations for different lags.

Generally, the second case is more expensive than the first case due to iteration procedure for the filter portion in which the counter parts of the observational variables have to be computed in every iteration. Of course, a linear/linearizable (around the a priori estimation) problem can also be solved nonlinearly. This kind of trade-off between computational cost and accuracy depends on how much we can gain. This is the case for 4D-VAR algorithm, in which the forward operator is always assumed to be quasi-linear around the maximum likelihood point  $\mathbf{W}^a$  for all

linear/linearizable and nonlinear problems, and the maximum likelihood point is always obtained as the limit point of an iterative algorithm.

Moreover, we should be aware of that for the problem where the nonlinearity is too strong, the linearizable approximation is no longer acceptable.

## 5.2 4D-PSAS algorithm

As seen in section 2.3, 4D-PSAS is derived from the cost function (41):

$$\begin{aligned} \mathcal{J}_N &= \frac{1}{2}(\mathbf{w}_0^t - \mathbf{w}^b)^T \mathbf{B}^{-1}(\mathbf{w}_0^t - \mathbf{w}^b) \\ &+ \frac{1}{2} \sum_{k=0}^N (\mathbf{w}_k^o - \mathbf{h}_k(\mathbf{w}_k^t))^T \mathbf{R}_k^{-1}(\mathbf{w}_k^o - \mathbf{h}_k(\mathbf{w}_k^t)) + \frac{1}{2} \sum_{k=1}^N \mathbf{b}_k^T \mathbf{Q}_k^{-1} \mathbf{b}_k . \end{aligned}$$

Let

$$\begin{aligned} \mathbf{d}_k &= \mathbf{w}_k^o - \mathbf{h}_k(\mathbf{A}_k(\mathbf{w}^b)) , \\ \delta \mathbf{x} &= \mathbf{w}_0^t - \mathbf{w}^b , \end{aligned}$$

and

$$\delta \mathbf{x}(t_k) = \mathbf{w}_k^t - \mathbf{A}_k(\mathbf{w}^b) = \mathbf{A}_{k,k-1}(\mathbf{w}_{k-1}^t) - \mathbf{A}_{k,k-1}(\mathbf{A}_{k-1}(\mathbf{w}^b)) + \mathbf{b}_k ,$$

then, for the weakest nonlinear case, if  $\mathbf{A}_{k,k-1}(\mathbf{w}_{k-1}^t)$  can be linearized around  $\mathbf{A}_{k-1}(\mathbf{w}^b)$ , we have

$$\delta \mathbf{x}(t_k) = \mathcal{A}_{k,k-1} \delta \mathbf{x}(t_{k-1}) + \mathbf{b}_k , \quad (150)$$

where

$$\mathcal{A}_{k,k-1} = \left( \frac{\partial \mathbf{A}_{k,k-1}}{\partial \mathbf{w}_{k-1}^t} \right)_{\mathbf{A}_{k-1}(\mathbf{w}^b)} . \quad (151)$$

Also, since

$$\mathbf{h}_k(\mathbf{w}_k^t) - \mathbf{w}_k^o = \mathbf{h}_k(\mathbf{w}_k^t) - \mathbf{h}_k(\mathbf{A}_k(\mathbf{w}^b)) - \mathbf{d}_k ,$$

if  $\mathbf{h}_k(\mathbf{w}_k^t)$  is linearizable around  $\mathbf{A}_k(\mathbf{w}^b)$ , then

$$\mathbf{h}_k(\mathbf{w}_k^t) - \mathbf{w}_k^o = \hat{\mathbf{h}}_k \delta \mathbf{x}(t_k) - \mathbf{d}_k , \quad (152)$$

where

$$\hat{\mathbf{h}}_k = \left( \frac{\partial \mathbf{h}_k}{\partial \mathbf{w}_k^t} \right)_{\mathbf{A}_k(\mathbf{w}^b)} . \quad (153)$$

Hence, equation (41) can be rewritten as

$$\mathcal{J}_N = \frac{1}{2} \delta \mathbf{x}^T \mathbf{B}^{-1} \delta \mathbf{x} + \frac{1}{2} \sum_{k=0}^N (\hat{\mathbf{h}}_k \delta \mathbf{x}(t_k) - \mathbf{d}_k)^T \mathbf{R}_k^{-1} (\hat{\mathbf{h}}_k \delta \mathbf{x}(t_k) - \mathbf{d}_k) + \frac{1}{2} \sum_{k=1}^N \mathbf{b}_k^T \mathbf{Q}_k^{-1} \mathbf{b}_k . \quad (154)$$

It is seen that equation (154) is the same as equation (43) except that  $\hat{\mathbf{h}}_k$  replaces  $\mathbf{h}_k$  and  $\mathcal{A}$  replaces  $\mathbf{A}$ . Therefore, solving such a linearizable problem is similar to solving a strictly linear

problem. However, comparing the 4D-PSAS and FLKS algorithm, it is clearly shown that the forward operators are linearized around the current best estimate  $\bar{\mathbf{W}}$  in FLKS algorithm, while they are linearized around the trajectory starting from the initial guess  $\mathbf{w}^b$ . Therefore, the requirement of linearization approximation for 4D-PSAS is much more strict than for FLKS algorithm.

If the linearization around the trajectory starting from  $\mathbf{w}^b$  is no longer acceptable, but the forward operators are still quasi-linear in the region of significant posterior probability density, then we can linearize the forward operators around the maximum likelihood point  $\mathbf{w}^a$ . Introducing

$$\begin{aligned}\mathbf{d}_k &= \mathbf{w}_k^o - \mathbf{h}_k(\mathbf{A}_k(\mathbf{w}^b)) , \\ \delta\mathbf{x} &= \mathbf{w}_0^t - \mathbf{w}^b ,\end{aligned}$$

and

$$\begin{aligned}\delta\mathbf{x}(t_k) &= \mathbf{w}_k^t - \mathbf{A}_k(\mathbf{w}^b) \\ &\simeq [\mathbf{A}_k(\mathbf{w}^a) - \mathbf{A}_k(\mathbf{w}^b) - \mathcal{A}_{k,k-1}(\mathbf{A}_{k-1}(\mathbf{w}^a) - \mathbf{A}_{k-1}(\mathbf{w}^b))] \\ &\quad + \mathcal{A}_{k,k-1}\delta\mathbf{x}(t_{k-1}) + \mathbf{b}_k ,\end{aligned}$$

where

$$\mathcal{A}_{k,k-1} = \left( \frac{\partial \mathbf{A}_{k,k-1}}{\partial \mathbf{w}_{k-1}^t} \right)_{\mathbf{A}_{k-1}(\mathbf{w}^a)} , \quad (155)$$

also since

$$\begin{aligned}\mathbf{h}_k(\mathbf{w}_k^t) - \mathbf{w}_k^o &= \mathbf{h}_k(\mathbf{w}_k^t) - \mathbf{h}_k(\mathbf{A}_k(\mathbf{w}^b)) - \mathbf{d}_k \\ &\simeq \hat{\mathbf{h}}_k\delta\mathbf{x}(t_k) - \bar{\mathbf{d}}_k ,\end{aligned} \quad (156)$$

where  $\bar{\mathbf{d}}_k$  and  $\hat{\mathbf{h}}_k$  are defined as

$$\bar{\mathbf{d}}_k = \mathbf{d}_k - [\mathbf{h}_k(\mathbf{A}_k(\mathbf{w}^a)) - \mathbf{h}_k(\mathbf{A}_k(\mathbf{w}^b)) - \hat{\mathbf{h}}_k(\mathbf{A}_k(\mathbf{w}^a) - \mathbf{A}_k(\mathbf{w}^b))] , \quad (157)$$

$$\hat{\mathbf{h}}_k = \left( \frac{\partial \mathbf{h}_k}{\partial \mathbf{w}_k^t} \right)_{\mathbf{A}_k(\mathbf{w}^a)} , \quad (158)$$

then equation (41) can be rewritten as

$$\mathcal{J}_N = \frac{1}{2}\delta\mathbf{x}^T \mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}\sum_{k=0}^N (\hat{\mathbf{h}}_k\delta\mathbf{x}(t_k) - \bar{\mathbf{d}}_k)^T \mathbf{R}_k^{-1} (\hat{\mathbf{h}}_k\delta\mathbf{x}(t_k) - \bar{\mathbf{d}}_k) + \frac{1}{2}\sum_{k=1}^N \mathbf{b}_k^T \mathbf{Q}_k^{-1} \mathbf{b}_k . \quad (159)$$

Applying equation (155), equation (159) can also be rewritten as the following form, which is a functional of  $\delta\mathbf{x}$  and  $\{\mathbf{b}_k\}$ ,

$$\begin{aligned}\mathcal{J}_N &= \frac{1}{2}\delta\mathbf{x}^T \mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}\sum_{k=1}^N \mathbf{b}_k^T \mathbf{Q}_k^{-1} \mathbf{b}_k + \\ &\quad \frac{1}{2}\sum_{k=0}^N (\hat{\mathbf{h}}_k \mathcal{A}_k \delta\mathbf{x} + \hat{\mathbf{h}}_k \sum_{l=0}^{k-1} \mathcal{A}_{k,k-l} \mathbf{b}_{k-l} - \mathbf{d}'_k)^T \mathbf{R}_k^{-1} (\hat{\mathbf{h}}_k \mathcal{A}_k \delta\mathbf{x} + \hat{\mathbf{h}}_k \sum_{l=0}^{k-1} \mathcal{A}_{k,k-l} \mathbf{b}_{k-l} - \mathbf{d}'_k) ,\end{aligned}$$

where

$$\mathbf{d}'_k = \bar{\mathbf{d}}_k - \hat{\mathbf{h}}_k \sum_{l=0}^{k-1} \mathcal{A}_{k,k-l} \left[ \mathbf{A}_{k-l}(\mathbf{w}^a) - \mathbf{A}_{k-l}(\mathbf{w}^b) - \mathcal{A}_{k-l,k-l-1}(\mathbf{A}_{k-l-1}(\mathbf{w}^a) - \mathbf{A}_{k-l-1}(\mathbf{w}^b)) \right]. \quad (160)$$

Using the definitions of  $\mathbf{G}$  and  $\mathbf{d}$  in section 2.3 with  $\mathbf{h}_k$ ,  $\mathbf{A}$  and  $\mathbf{d}_k$  being replaced by  $\hat{\mathbf{h}}_k$ ,  $\mathcal{A}$  and  $\mathbf{d}'_k$ , respectively, and the definitions of  $\mathbf{z}$ ,  $\mathbf{D}$  and  $\mathbf{R}$ , the cost function  $\mathcal{J}_N$  can then be written as a compact form:

$$\mathcal{J}_N = \frac{1}{2} \mathbf{z}^T \mathbf{D}^{-1} \mathbf{z} + \frac{1}{2} (\mathbf{Gz} - \mathbf{d})^T \mathbf{R}^{-1} (\mathbf{Gz} - \mathbf{d}). \quad (161)$$

Then the maximum likelihood point can be obtained using an iterative optimization algorithm, e.g., quasi-Newton method,

$$\begin{aligned} \mathbf{z}_{r+1} &= \mathbf{z}_r - (\mathbf{D}^{-1} + \mathbf{G}_r^T \mathbf{R}^{-1} \mathbf{G}_r)^{-1} \left\{ \mathbf{D}^{-1} \mathbf{z}_r + \mathbf{G}_r^T \mathbf{R}^{-1} (\mathbf{G}_r \mathbf{z}_r - \mathbf{d}_r) \right\} \\ &= \mathbf{D} \mathbf{G}_r^T (\mathbf{R} + \mathbf{G}_r \mathbf{D} \mathbf{G}_r^T)^{-1} \mathbf{d}_r. \end{aligned} \quad (162)$$

Like FLKS algorithm, other function linearizations around the current best estimates (the forecasts starting from the retrospective analysis at  $t_0$ ) are also possible, but the solution formula will be a little different.



## **Acknowledgments**

This research was supported by the NASA EOS Interdisciplinary Project on Data Assimilation.

## Appendix A

### Comparison between the analysis estimates of fixed-interval and fixed-point smoothers

In section 2.3, it is shown that the cost function  $\mathcal{J}_N$  of incremental 4D-VAR and 4D-PSAS derived from the conditional probability density  $p(\mathbf{w}_0^t, \mathbf{w}_1^t, \mathbf{w}_2^t, \dots, \mathbf{w}_N^t | \mathbf{W}_N^o)$  is given as equation (43):

$$\mathcal{J}_N = \frac{1}{2} \delta \mathbf{x}^T \mathbf{B}^{-1} \delta \mathbf{x} + \frac{1}{2} \sum_{k=0}^N (\mathbf{h}_k \delta \mathbf{x}(t_k) - \mathbf{d}_k)^T \mathbf{R}_k^{-1} (\mathbf{h}_k \delta \mathbf{x}(t_k) - \mathbf{d}_k) + \frac{1}{2} \sum_{k=1}^N \mathbf{b}_k^T \mathbf{Q}_k^{-1} \mathbf{b}_k \quad (163)$$

or equation (44):

$$\mathcal{J}_N = \frac{1}{2} \mathbf{z}^T \mathbf{D}^{-1} \mathbf{z} + \frac{1}{2} (\mathbf{G} \mathbf{z} - \mathbf{d})^T \mathbf{R}^{-1} (\mathbf{G} \mathbf{z} - \mathbf{d}), \quad (164)$$

and the solution of 4D-PSAS which minimizes the cost function  $\mathcal{J}_N$ , therefore, is given as equation (46):

$$\mathbf{z} = \mathbf{D} \mathbf{G}^T (\mathbf{G} \mathbf{D} \mathbf{G}^T + \mathbf{R})^{-1} \mathbf{d}. \quad (165)$$

Using the definitions of  $\mathbf{D}$ ,  $\mathbf{G}$  and the matrix calculations, it follows that

$$\mathbf{D} \mathbf{G}^T = \begin{pmatrix} \mathbf{B} \mathbf{h}_0^T & \mathbf{B} (\mathbf{h}_1 \mathbf{A}_1)^T & \mathbf{B} (\mathbf{h}_2 \mathbf{A}_2)^T & \dots & \mathbf{B} (\mathbf{h}_N \mathbf{A}_N)^T \\ \mathbf{0} & \mathbf{Q}_1 \mathbf{h}_1^T & \mathbf{Q}_1 (\mathbf{h}_2 \mathbf{A}_{2,1})^T & \dots & \mathbf{Q}_1 (\mathbf{h}_N \mathbf{A}_{N,1})^T \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}_2 \mathbf{h}_2^T & \dots & \mathbf{Q}_2 (\mathbf{h}_N \mathbf{A}_{N,2})^T \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{Q}_N \mathbf{h}_N^T \end{pmatrix} \quad (166)$$

and

$$\mathbf{G} \mathbf{D} \mathbf{G}^T = \bar{\mathbf{G}} \bar{\mathbf{B}} \bar{\mathbf{G}}^T + \mathcal{G} \bar{\mathbf{Q}} \mathcal{G}^T, \quad (167)$$

where

$$\mathcal{G} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{h}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{h}_2 \mathbf{A}_{2,1} & \mathbf{h}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mathbf{0} & \mathbf{h}_N \mathbf{A}_{N,1} & \mathbf{h}_N \mathbf{A}_{N,2} & \dots & \mathbf{h}_N \end{pmatrix}, \quad \bar{\mathbf{Q}} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{Q}_N \end{pmatrix}.$$

Therefore, the solution of 4D-PSAS can be rewritten as

$$\mathbf{z} = \mathbf{D} \mathbf{G}^T (\bar{\mathbf{G}} \bar{\mathbf{B}} \bar{\mathbf{G}}^T + \mathcal{G} \bar{\mathbf{Q}} \mathcal{G}^T + \mathbf{R})^{-1} \mathbf{d}, \quad (168)$$

and it is easy to prove by using equation (166) that the analysis incremental  $\delta \mathbf{x}$  is given as

$$\delta \mathbf{x} = \mathbf{B} \bar{\mathbf{G}}^T (\bar{\mathbf{G}} \bar{\mathbf{B}} \bar{\mathbf{G}}^T + \mathcal{G} \bar{\mathbf{Q}} \mathcal{G}^T + \mathbf{R})^{-1} \mathbf{d}. \quad (169)$$

Now let's take a look at the conditional probability density function  $p(\mathbf{w}_0^t | \mathbf{W}_N^o)$  which gives the estimate of  $\mathbf{w}_0^t$  conditioned on  $N + 1$  observations:

$$p(\mathbf{w}_0^t | \mathbf{W}_N^o) = \frac{1}{p(\mathbf{W}_N^o)} p(\mathbf{W}_N^o | \mathbf{w}_0^t) p(\mathbf{w}_0^t). \quad (170)$$

Similar to equation (113), under the assumption of Gaussian distribution, the probability density function  $p(\mathbf{W}_N^o | \mathbf{w}_0^t)$  is proportional to

$$p(\mathbf{W}_N^o | \mathbf{w}_0^t) \propto \exp \left\{ -\frac{1}{2} (\mathbf{W}_N^o - \bar{\mathbf{G}} \mathbf{w}_0^t)^T (\mathbf{R} + \mathcal{G} \bar{\mathbf{Q}} \mathcal{G}^T)^{-1} (\mathbf{W}_N^o - \bar{\mathbf{G}} \mathbf{w}_0^t) \right\}, \quad (171)$$

therefore, the conditional probability density function  $p(\mathbf{w}_0^t | \mathbf{W}_N^o)$  is given as

$$p(\mathbf{w}_0^t | \mathbf{W}_N^o) = \text{const} \cdot \exp(-\mathcal{J}_{FP4}), \quad (172)$$

where the cost function  $\mathcal{J}_{FP4}$  is

$$\begin{aligned} \mathcal{J}_{FP4} &= \frac{1}{2} (\mathbf{w}_0^t - \mathbf{w}^b)^T \mathbf{B}^{-1} (\mathbf{w}_0^t - \mathbf{w}^b) \\ &\quad + \frac{1}{2} (\mathbf{W}_N^o - \bar{\mathbf{G}} \mathbf{w}_0^t)^T (\mathbf{R} + \mathcal{G} \bar{\mathbf{Q}} \mathcal{G}^T)^{-1} (\mathbf{W}_N^o - \bar{\mathbf{G}} \mathbf{w}_0^t). \end{aligned} \quad (173)$$

Applying the same definitions of  $\delta \mathbf{x}$  and  $\mathbf{d}$  as in section 2.3, i.e.,  $\delta \mathbf{x} = \mathbf{w}_0^t - \mathbf{w}^b$ ,  $\mathbf{d}_k = \mathbf{w}_k^o - \mathbf{h}_k \mathbf{A}_k \mathbf{w}^b$ , the cost function  $\mathcal{J}_{FP4}$  is rewritten as

$$\mathcal{J}_{FP4} = \frac{1}{2} \delta \mathbf{x}^T \mathbf{B}^{-1} \delta \mathbf{x} + \frac{1}{2} (\bar{\mathbf{G}} \delta \mathbf{x} - \mathbf{d})^T (\mathbf{R} + \mathcal{G} \bar{\mathbf{Q}} \mathcal{G}^T)^{-1} (\bar{\mathbf{G}} \delta \mathbf{x} - \mathbf{d}). \quad (174)$$

The minimum of the cost function, which is the maximum likelihood point of the conditional probability density function  $p(\mathbf{w}_0^t | \mathbf{W}_N^o)$ , is given as:

$$\delta \mathbf{x} = \mathbf{B} \bar{\mathbf{G}}^T (\bar{\mathbf{G}} \mathbf{B} \bar{\mathbf{G}}^T + \mathcal{G} \bar{\mathbf{Q}} \mathcal{G}^T + \mathbf{R})^{-1} \mathbf{d}. \quad (175)$$

### Remarks:

- Comparing equations (169) and (175), we notice that the 4D-VAR or 4D-PSAS cost function  $\mathcal{J}_N$  (43) has the same solution for the state increment as the cost function  $\mathcal{J}_{FP4}$  (173) does. This means that given the same amount of observations, the fixed-interval smoother  $p(\mathbf{w}_0^t, \mathbf{w}_1^t, \mathbf{w}_2^t, \dots, \mathbf{w}_N^t | \mathbf{W}_N^o)$  yields the same estimate of  $\mathbf{w}_0^t$  as the fixed-point smoother  $p(\mathbf{w}_0^t | \mathbf{W}_N^o)$  does.
- From the 4D-VAR or 4D-PSAS cost function (43) we can get the solutions for both the state increment and the model error at every time step, while from the cost function (173) we can only obtain the solution for the state increment. In the other words,  $\mathcal{J}_N = \mathcal{J}_N(\delta \mathbf{x}, \{\mathbf{b}_k\})$ ,  $\mathcal{J}_{FP4} = \mathcal{J}_{FP4}(\delta \mathbf{x})$ .
- The question remained is: do we really need those model error estimates at every model integration time step for reanalysis purpose? With the availability of those model error estimates, one can issue a forecast from the initial condition to obtain the analyses within the interval. However, the subsequent analyses within the interval produced by a forecast comprise information of varied amount of future observational time levels. In other words, the analysis at time  $t_0$  contains information of current and future observations at times  $t_0$ ,  $t_1$ ,  $\dots$ , and  $t_N$ ; the analysis at time  $t_1$  contains information of observations at times  $t_1$ ,  $\dots$ , and  $t_N$ ; and so on, until the analysis at time  $t_N$  contains information of observations at

time  $t_N$ , which is just a filter solution. To the users of the reanalysis products, generally, the model error estimates are less useful, since their most concern is the qualities of the products and also they usually don't have (and also it is not necessary for them to have ) access to the data assimilation system which produces the products. The reanalysis products should be produced by incorporating the fixed future time levels of observations rather than being produced by issuing a forecast from an initial condition. Also, the requirement of the memory storage for the model errors is huge. In this study, we assume the observations occur at every time step, but, the total number of model integration time steps is generally much larger than the total number of the times when the observations occur. In fact, the model dynamics is a continuous process. One way to reduce the storage requirement is to relax the assumption of the model error whiteness in time and to redefine the cost function  $\mathcal{J}_N$ . But on the other hand, the forecast issued from the optimal initial condition can be used as a better priori estimation for the next implementation of the algorithm.

## Appendix B

### Sequential form of 4D-PSAS

It is shown that the cost function of 4D-PSAS for the linear perfect model is given as equation (49):

$$\begin{aligned} \mathcal{J}_N &= \frac{1}{2}(\mathbf{w}_0^t - \mathbf{w}^b)^T \mathbf{B}^{-1}(\mathbf{w}_0^t - \mathbf{w}^b) \\ &\quad + \frac{1}{2} \sum_{k=0}^N (\mathbf{w}_k^o - \mathbf{h}_k \mathbf{w}_k^t)^T \mathbf{R}_k^{-1} (\mathbf{w}_k^o - \mathbf{h}_k \mathbf{w}_k^t), \end{aligned}$$

which is identical with the FLKS cost function in the fixed-point smoother perspective -  $\mathcal{J}_{FP1}$  (equation (108)) given  $k = l = N$ ,  $\mathbf{w}^b = \mathbf{w}_{k-l|k-l-1}^f$ , and  $\mathbf{B} = \mathbf{P}_{k-l|k-l-1}^f$ . The solution of 4D-PSAS which minimizes the cost function  $\mathcal{J}_N$  is calculated as equation (54):

$$\delta \mathbf{x} = (\mathbf{B}^{-1} + \bar{\mathbf{G}}^T \mathbf{R}^{-1} \bar{\mathbf{G}})^{-1} \bar{\mathbf{G}}^T \mathbf{R}^{-1} \mathbf{d}, \quad (176)$$

i.e., the solution at time  $t_0$  is expressed as

$$\mathbf{w}_{0|N}^a = \mathbf{w}^b + (\mathbf{B}^{-1} + \bar{\mathbf{G}}^T \mathbf{R}^{-1} \bar{\mathbf{G}})^{-1} \bar{\mathbf{G}}^T \mathbf{R}^{-1} (\mathbf{W}_N^o - \bar{\mathbf{G}} \mathbf{w}^b) \quad (177)$$

with the error covariance

$$\mathbf{P}_{0|N}^a = (\mathbf{B}^{-1} + \bar{\mathbf{G}}^T \mathbf{R}^{-1} \bar{\mathbf{G}})^{-1}. \quad (178)$$

In the following, we will show that for a linear perfect model, 4D-PSAS can be written as a sequential algorithm as well.

Substituting the definitions of matrices  $\bar{\mathbf{G}}$  and  $\mathbf{R}$  into the 4D-PSAS solution - equation (177), we have

$$\begin{aligned} \mathbf{w}_{0|N}^a &= \mathbf{w}^b + \left( \mathbf{B}^{-1} + \begin{pmatrix} \mathbf{G}_0 \\ \mathbf{G}_1 \\ \vdots \\ \mathbf{G}_N \end{pmatrix}^T \begin{pmatrix} \mathbf{R}_0 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_1 & \dots & \mathbf{0} \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{R}_N \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{G}_0 \\ \mathbf{G}_1 \\ \vdots \\ \mathbf{G}_N \end{pmatrix} \right)^{-1} \\ &\quad \cdot \begin{pmatrix} \mathbf{G}_0 \\ \mathbf{G}_1 \\ \vdots \\ \mathbf{G}_N \end{pmatrix}^T \begin{pmatrix} \mathbf{R}_0 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_1 & \dots & \mathbf{0} \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{R}_N \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{w}_0^o - \mathbf{G}_0 \mathbf{w}^b \\ \mathbf{w}_1^o - \mathbf{G}_1 \mathbf{w}^b \\ \vdots \\ \mathbf{w}_N^o - \mathbf{G}_N \mathbf{w}^b \end{pmatrix} \\ &= \mathbf{w}^b + \left( \mathbf{B}^{-1} + \sum_{k=0}^N \mathbf{G}_k^T \mathbf{R}_k^{-1} \mathbf{G}_k \right)^{-1} \left[ \sum_{k=0}^N \mathbf{G}_k^T \mathbf{R}_k^{-1} (\mathbf{w}_k^o - \mathbf{G}_k \mathbf{w}^b) \right]. \quad (179) \end{aligned}$$

Since

$$\left( \mathbf{B}^{-1} + \sum_{k=0}^N \mathbf{G}_k^T \mathbf{R}_k^{-1} \mathbf{G}_k \right)^{-1}$$

$$\begin{aligned}
&= \left( \mathbf{B}^{-1} + \sum_{k=0}^N \mathbf{G}_k^T \mathbf{R}_k^{-1} \mathbf{G}_k \right)^{-1} \\
&\cdot \left[ \left( \mathbf{B}^{-1} + \sum_{k=0}^N \mathbf{G}_k^T \mathbf{R}_k^{-1} \mathbf{G}_k \right) - \sum_{k=1}^N \mathbf{G}_k^T \mathbf{R}_k^{-1} \mathbf{G}_k \right] \left( \mathbf{B}^{-1} + \mathbf{G}_0^T \mathbf{R}_0^{-1} \mathbf{G}_0 \right)^{-1} \\
&= \left[ \mathbf{I} - \left( \mathbf{B}^{-1} + \sum_{k=0}^N \mathbf{G}_k^T \mathbf{R}_k^{-1} \mathbf{G}_k \right)^{-1} \left( \sum_{k=1}^N \mathbf{G}_k^T \mathbf{R}_k^{-1} \mathbf{G}_k \right) \right] \left( \mathbf{B}^{-1} + \mathbf{G}_0^T \mathbf{R}_0^{-1} \mathbf{G}_0 \right)^{-1} \\
&= \left( \mathbf{B}^{-1} + \mathbf{G}_0^T \mathbf{R}_0^{-1} \mathbf{G}_0 \right)^{-1} - \left( \mathbf{B}^{-1} + \sum_{k=0}^N \mathbf{G}_k^T \mathbf{R}_k^{-1} \mathbf{G}_k \right)^{-1} \\
&\cdot \left( \sum_{k=1}^N \mathbf{G}_k^T \mathbf{R}_k^{-1} \mathbf{G}_k \right) \left( \mathbf{B}^{-1} + \mathbf{G}_0^T \mathbf{R}_0^{-1} \mathbf{G}_0 \right)^{-1}, \tag{180}
\end{aligned}$$

equation (179) becomes

$$\begin{aligned}
\mathbf{w}_{0|N}^a &= \mathbf{w}^b + \left[ \left( \mathbf{B}^{-1} + \mathbf{G}_0^T \mathbf{R}_0^{-1} \mathbf{G}_0 \right)^{-1} \right. \\
&\quad \left. - \left( \mathbf{B}^{-1} + \sum_{k=0}^N \mathbf{G}_k^T \mathbf{R}_k^{-1} \mathbf{G}_k \right)^{-1} \left( \sum_{k=1}^N \mathbf{G}_k^T \mathbf{R}_k^{-1} \mathbf{G}_k \right) \left( \mathbf{B}^{-1} + \mathbf{G}_0^T \mathbf{R}_0^{-1} \mathbf{G}_0 \right)^{-1} \right] \\
&\quad \cdot \left[ \mathbf{G}_0^T \mathbf{R}_0^{-1} (\mathbf{w}_0^o - \mathbf{G}_0 \mathbf{w}^b) + \sum_{k=1}^N \mathbf{G}_k^T \mathbf{R}_k^{-1} (\mathbf{w}_k^o - \mathbf{G}_k \mathbf{w}^b) \right]. \tag{181}
\end{aligned}$$

Rearranging the above equation, we obtain

$$\mathbf{w}_{0|N}^a = \mathbf{w}_{0|0}^a + \left( (\mathbf{P}_{0|0}^a)^{-1} + \sum_{k=1}^N \mathbf{G}_k^T \mathbf{R}_k^{-1} \mathbf{G}_k \right)^{-1} \left[ \sum_{k=1}^N \mathbf{G}_k^T \mathbf{R}_k^{-1} (\mathbf{w}_k^o - \mathbf{G}_k \mathbf{w}_{0|0}^a) \right], \tag{182}$$

where

$$\mathbf{w}_{0|0}^a = \mathbf{w}^b + \left[ \mathbf{B}^{-1} + \mathbf{G}_0^T \mathbf{R}_0^{-1} \mathbf{G}_0 \right]^{-1} \mathbf{G}_0^T \mathbf{R}_0^{-1} (\mathbf{w}_0^o - \mathbf{G}_0 \mathbf{w}^b), \tag{183}$$

$$\mathbf{P}_{0|0}^a = \left( \mathbf{B}^{-1} + \mathbf{G}_0^T \mathbf{R}_0^{-1} \mathbf{G}_0 \right)^{-1}. \tag{184}$$

Keeping doing this manipulation, we have

$$\mathbf{w}_{0|N}^a = \mathbf{w}_{0|1}^a + \left( (\mathbf{P}_{0|1}^a)^{-1} + \sum_{k=2}^N \mathbf{G}_k^T \mathbf{R}_k^{-1} \mathbf{G}_k \right)^{-1} \left[ \sum_{k=2}^N \mathbf{G}_k^T \mathbf{R}_k^{-1} (\mathbf{w}_k^o - \mathbf{G}_k \mathbf{w}_{0|1}^a) \right], \tag{185}$$

where

$$\mathbf{w}_{0|1}^a = \mathbf{w}_{0|0}^a + \left[ (\mathbf{P}_{0|0}^a)^{-1} + \mathbf{G}_1^T \mathbf{R}_1^{-1} \mathbf{G}_1 \right]^{-1} \mathbf{G}_1^T \mathbf{R}_1^{-1} (\mathbf{w}_1^o - \mathbf{G}_1 \mathbf{w}_{0|0}^a), \tag{186}$$

$$\mathbf{P}_{0|1}^a = \left( (\mathbf{P}_{0|0}^a)^{-1} + \mathbf{G}_1^T \mathbf{R}_1^{-1} \mathbf{G}_1 \right)^{-1}. \tag{187}$$

Inductively, we obtain

$$\mathbf{w}_{0|N}^a = \mathbf{w}_{0|N-1}^a + \left( (\mathbf{P}_{0|N-1}^a)^{-1} + \mathbf{G}_N^T \mathbf{R}_N^{-1} \mathbf{G}_N \right)^{-1} \left[ \mathbf{G}_N^T \mathbf{R}_N^{-1} (\mathbf{w}_N^o - \mathbf{G}_N \mathbf{w}_{0|N-1}^a) \right], \tag{188}$$

where

$$\mathbf{w}_{0|N-1}^a = \mathbf{w}_{0|N-2}^a + \left[ (\mathbf{P}_{0|N-2}^a)^{-1} + \mathbf{G}_{N-1}^T \mathbf{R}_{N-1}^{-1} \mathbf{G}_{N-1} \right]^{-1} \mathbf{G}_{N-1}^T \mathbf{R}_{N-1}^{-1} (\mathbf{w}_{N-1}^o - \mathbf{G}_{N-1} \mathbf{w}_{0|N-2}^a), \quad (189)$$

$$\mathbf{P}_{0|N-1}^a = \left( (\mathbf{P}_{0|N-2}^a)^{-1} + \mathbf{G}_{N-1}^T \mathbf{R}_{N-1}^{-1} \mathbf{G}_{N-1} \right)^{-1}. \quad (190)$$

**Remark:**

- For linear perfect model, if the observations are white in time, then the result of 4D-PSAS obtained by assimilating all of the observations simultaneously is identical to that by assimilating the observations sequentially, i.e., one time level of observations at a time. In other words, the 4D-PSAS algorithm can also be rewritten as a sequential algorithm.
- Comparing the formula with the sequential FLKS formula, we see that they are the same given  $k = l = N$ ,  $\mathbf{w}_{k-l|k-l-1}^f = \mathbf{w}^b$ , and  $\mathbf{P}_{k-l|k-l-1}^f = \mathbf{B}$ .

## References

- Anderson, B.D.O., and J.B. Moore, 1979: *Optimal filtering*. Prentice-Hall, 357pp.
- Bennett, A.F., 1992: *Inverse methods in physical oceanography*. Cambridge University Press, 346pp.
- Bennett, A.F., B.S. Chua, and L.M. Leslie, 1996: Generalized inversion of a global numerical weather prediction model. *Meteorol. Atmos. Phys.*, **60**, 165–178.
- Bryson, JR., A.E. and Y.C. Ho, 1975: *Applied optimal control*. Blaisdell Publishing Company, Waltham, Mass., 481pp.
- Cohn, S.E., N.S. Sivakumaran, and R. Todling, 1994: A fixed-lag Kalman smoother for retrospective data assimilation. *Mon. Wea. Rev.*, **122**, 2838–2867.
- Cohn, S.E., 1997: An introduction to Estimation theory. *J. Meteorol. Soc. Japan*, **75**, No. 1B, 257–288.
- Courtier, P., 1997: Dual formulation of four dimensional variational assimilation. *Quart. J. Roy. Meteor. Soc.*, **123**, 2449–2461.
- Da Silva, Arlindo, and Jing Guo, 1996: Documentation of the Physical-space Statistical Analysis System (PSAS). Part I: The conjugate gradient solver version PSAS - 1.00. DAO Office Note 96-2.
- Evensen, Geir, 1997: Advanced data assimilation for strongly nonlinear dynamics. *Mon. Wea. Rev.*, **125**, 1342–1354.
- Jazwinski, A.H., 1970: *Stochastic Processes and filtering theory*. Academic Press, 376pp.
- Menard, R., and R. Daley, 1996: The application of Kalman smoother theory to the estimation of 4DVAR error statistics. *Tellus*, **48A**, 221–237.
- Pu, Z-X, K. Eugenia, and J. Sela, 1997: Sensitivity of forecast errors to initial conditions with a quasi-inverse linear method. *Mon. Wea. Rev.*, **125**, 2479–2503.
- Todling, R., S.E. Cohn and N.S. Sivakumaran, 1998: Suboptimal Schemes for retrospective data assimilation based on the fixed-lag Kalman smoother. *Mon. Wea. Rev.*, **126**, 247–259.
- Tarantola, Albert, 1987: *Inverse problem theory*. Elsevier Science Publishing Company Inc., 613pp.