

# *Exploring strategies for coupled 4D-Var data assimilation using an idealised atmosphere-ocean model*

Article

Accepted Version

Smith, P. J. ORCID: <https://orcid.org/0000-0003-4570-4127>,  
Fowler, A. M. ORCID: <https://orcid.org/0000-0003-3650-3948>  
and Lawless, A. S. ORCID: <https://orcid.org/0000-0002-3016-6568> (2015) Exploring strategies for coupled 4D-Var data assimilation using an idealised atmosphere-ocean model. *Tellus A*, 67. 27025. ISSN 1600-0870 doi: <https://doi.org/10.3402/tellusa.v67.27025> Available at <https://centaur.reading.ac.uk/40637/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.3402/tellusa.v67.27025>

Publisher: Co-Action Publishing

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

# Exploring strategies for coupled 4D-Var data assimilation using an idealised atmosphere-ocean model

By POLLY J. SMITH\*<sup>†‡</sup>, ALISON M. FOWLER<sup>†‡</sup> and AMOS S. LAWLESS<sup>†‡</sup> <sup>†</sup>. *School of Mathematical and Physical Sciences, University of Reading, UK; ‡. National Centre for Earth Observation, University of Reading, UK.*

(Manuscript received xx xxxx xx; in final form xx xxxx xx)

## ABSTRACT

Operational forecasting centres are currently developing data assimilation systems for coupled atmosphere-ocean models. Strongly coupled assimilation, in which a single assimilation system is applied to a coupled model, presents significant technical and scientific challenges. Hence weakly coupled assimilation systems are being developed as a first step, in which the coupled model is used to compare the current state estimate with observations, but corrections to the atmosphere and ocean initial conditions are then calculated independently. In this paper we provide a comprehensive description of the different coupled assimilation methodologies in the context of four dimensional variational assimilation (4D-Var) and use an idealised framework to assess the expected benefits of moving towards coupled data assimilation.

We implement an incremental 4D-Var system within an idealised single column atmosphere-ocean model. The system has the capability to run both strongly and weakly coupled assimilations as well as uncoupled atmosphere or ocean only assimilations, thus allowing a systematic comparison of the different strategies for treating the coupled data assimilation problem. We present results from a series of identical twin experiments devised to investigate the behaviour and sensitivities of the different approaches. Overall, our study demonstrates the potential benefits that may

be expected from coupled data assimilation. When compared to uncoupled initialisation, coupled assimilation is able to produce more balanced initial analysis fields, thus reducing initialisation shock and its impact on the subsequent forecast. Single observation experiments demonstrate how coupled assimilation systems are able to pass information between the atmosphere and ocean and therefore use near-surface data to greater effect. We show that much of this benefit may also be gained from a weakly coupled assimilation system, but that this can be sensitive to the parameters used in the assimilation.

*Keywords: incremental four dimensional variational data assimilation, single column model, KPP mixed layer model, initialisation, strongly coupled, weakly coupled*

## 1. Introduction

The successful application of data assimilation techniques to operational numerical weather prediction and ocean forecasting systems, together with increasing availability of near surface observations from new satellite missions, has led to an increased interest in their potential for use in the initialisation of coupled atmosphere-ocean models. To produce reliable predictions across seasonal to decadal time scales we need to simulate the evolution of the atmosphere and ocean coupled together. Coupled models have been used operationally for seasonal and longer range forecasting for a number of years, and are now being considered for shorter term prediction. Typically, the initial conditions for these forecasts are provided by combining analyses from independent (uncoupled) ocean and atmosphere assimilation systems (Balmaseda and Anderson, 2009). This approach ignores interactions between the systems and this inconsistency can cause imbalance such that the initial conditions are far from the natural state of the coupled system. When the coupled forecast is initialised the model ad-

---

\* Corresponding author.

e-mail: p.j.smith@reading.ac.uk

14 justs itself towards its preferred climatology; this adjustment can produce rapid shocks at the  
15 air-sea interface during the early stages of the forecast, a process referred to as initialisation  
16 shock (Balmaseda, 2012). It also means that near-surface data are not fully utilised.

17 The development of coupled atmosphere-ocean data assimilation systems presents a num-  
18 ber of scientific and technical challenges (Murphy et al., 2010; Lawless, 2012) and requires  
19 a significant amount of resources to be made possible operationally. Yet such systems offer a  
20 long list of potential benefits including improved use of near-surface observations, reduction  
21 of initialisation shocks in coupled forecasts, and generation of a consistent system state for  
22 the initialisation of coupled forecasts across all timescales. In addition, coupled reanalyses  
23 offer the potential for greater understanding and representation of air-sea exchange processes  
24 in turn facilitating more accurate prediction of phenomena such as El Niño and the Madden-  
25 Julian Oscillation (MJO) in which air-sea interaction plays an important role.

26 Studies have shown that, in certain regions, the initialisation of coupled models can enhance  
27 the skill of decadal predictions for the first 5 or so years of the forecast (Meehl et al. (2014)  
28 and references therein). Although it is widely accepted that coupled data assimilation has a  
29 central role in improving our ability to generate consistent and accurate initial conditions for  
30 coupled atmosphere-ocean forecasting it is still a relatively young area of research. Hence  
31 there has so far only been limited amount of work in this field. An assortment of strategies  
32 for using observed data to improve coupled model initialisation have been explored with  
33 varying degrees of success; these include sea surface temperature (SST) nudging or relaxation  
34 (e.g. Keenlyside et al. (2008)), anomaly initialisation/ bias-blind assimilation (e.g. Pierce  
35 et al. (2004)), anomaly coupling (Pohlmann et al. (2009)), and variants of the full uncoupled  
36 initialisation approach (Balmaseda and Anderson, 2009). Work has mainly been focussed on

37 improving ocean initial conditions with a lack of fully consistent treatment of air-sea feedback  
38 mechanisms.

39 There are groups exploring more comprehensive approaches that aim to produce more dy-  
40 namically balanced initial ocean-atmosphere states. The Japan Agency for Marine-Earth Sci-  
41 ence and Technology (JAMSTEC) are working towards a coupled 4D-Var data assimilation  
42 system for their Coupled model for the Earth Simulator (CFES), a fully coupled global cli-  
43 mate model. Sugiura et al. (2008) describes the development of a first step 4D-Var system  
44 for estimating ocean initial conditions together with adjustment parameters of the bulk flux  
45 formulae. Their approach is focussed on representing slow time scales only, filtering out fast  
46 atmospheric modes by using 10 day mean states. Whilst this enables them to better represent  
47 several key seasonal to interannual climate events in the tropical Pacific and Indian Ocean  
48 region, including the El Niño, it would not be suitable for atmospheric reanalyses or for ini-  
49 tialising medium range forecasts.

50 So far, most coupled data assimilation work in the published literature has employed ensem-  
51 ble rather than variational based assimilation methods. Tardif et al. (2014) use an idealised  
52 low dimension atmosphere-ocean climate model with an Ensemble Kalman Filter (EnKF) to  
53 explore strategies for ensemble coupled data assimilation. Their model represents an idealisa-  
54 tion of the midlatitude North Atlantic climate system and is designed to allow experiments on  
55 very long timescales in order to assess the EnKF approach in terms of effectiveness for initial-  
56 isation of the meridional overturning circulation (MOC). In twin experiments using 50 year  
57 windows from a 5000 year reference simulation, they found that forcing the idealised ocean  
58 model with atmospheric analyses was inefficient at recovering the MOC due to slow conver-

59 gence of the solutions. In contrast, coupled assimilation produced accurate MOC analyses,  
60 even when only atmospheric observations were assimilated.

61 In a larger scale study, Zhang et al. (2007) describe a coupled assimilation system consist-  
62 ing of an EnKF applied to the National Oceanic and Atmospheric Administration (NOAA)  
63 Geophysical Fluid Dynamics Laboratory (GFDL) global fully coupled climate model for the  
64 initialisation of seasonal and decadal forecasts. The system is evaluated in a series of twin  
65 experiments assimilating atmosphere-only or ocean-only observations but not both together.  
66 Although the system shows good skill in reconstructing seasonal and decadal ocean vari-  
67 ability and trends it fails to fully realise the potential benefit of surface and near surface  
68 observational data.

69 In this paper we explore some of the fundamental questions in the design of coupled varia-  
70 tional data assimilation systems within the context of an idealised one-dimensional (1D) col-  
71 umn coupled atmosphere-ocean model. The system is designed to enable the effective explo-  
72 ration of various approaches to performing coupled model data assimilation whilst avoiding  
73 many of the issues associated with more complex models and allows us to perform exper-  
74 iments that would not be feasible in operational scale systems. We employ an incremental  
75 four dimensional variational data assimilation (4D-Var) scheme (Courtier et al., 1994; Law-  
76 less et al., 2005; Lawless, 2013) to reflect the coupled assimilation systems currently being  
77 developed at the European Centre for Medium Range Weather Forecasts (ECMWF) and UK  
78 Met Office. The problem of variational data assimilation is to find the initial state such that  
79 the model forecast best fits the available observations over a given time window, subject to  
80 the state remaining close to a given *a priori*, or background, estimate and allowing for the  
81 errors in each. This best estimate is known as the *analysis* and should be consistent with both

82 the observations and the system dynamics. The standard 4D-Var problem is formulated as  
83 the minimisation of a non-linear weighted least squares cost function; in the incremental ap-  
84 proach the non-linear problem is instead approximated by a sequence of linear least squares  
85 problems. Rather than search for the initial state directly, we solve in terms of increments  
86 with respect to an initial background state; this is done iteratively in a series of linearised  
87 inner-loop cost function minimisations and non-linear outer-loop update steps.

88 *Strongly or fully coupled* variational assimilation treats the atmosphere and ocean as a single  
89 coherent system, using the coupled model in both the inner- and outer-loops. This approach  
90 is able to pass information between the atmosphere and ocean, and therefore enables obser-  
91 vations of atmospheric variables to influence the ocean increments and vice versa. This is  
92 expected to lead to better use of near-surface observations, such as scatterometer winds and  
93 SST, that depend on both the atmosphere and ocean state, and to produce a more physically-  
94 balanced analysis. Although there are currently no plans to move towards strongly coupled  
95 systems at operational centres, this approach represents the quintessential coupled assimi-  
96 lation system and implementing it in our idealised system allows us to better assess the potential  
97 of intermediate, or *weakly coupled*, approaches.

98 As a first step towards the implementation of operational coupled data assimilation, centres  
99 such as the ECMWF and UK Met Office are developing prototype weakly coupled assimi-  
100 lation systems (Brassington et al., 2015; Laloyaux et al., 2014; 2015; Lea et al., 2015). Weakly  
101 coupled incremental 4D-Var makes use of the incremental inner and outer-loop structure; the  
102 coupled model is used to provide the initial atmosphere and ocean background states and non-  
103 linear trajectory for separate (uncoupled) inner-loop atmosphere and ocean minimisations; the  
104 uncoupled analysis increments are then fed back into the coupled model for the next outer-

105 loop forecast. Unlike strongly coupled assimilation, the weakly coupled approach does not  
106 allow for cross-covariance information between the atmosphere and ocean. This means that  
107 the atmosphere (ocean) observations cannot affect the ocean (atmosphere) analyses unless  
108 multiple outer-loops are performed.

109 The purpose of this paper is to (i) provide a comprehensive description of the different cou-  
110 pled 4D-Var data assimilation methodologies, and (ii) use our idealised framework to assess  
111 the benefits expected in moving towards coupled data assimilation systems. Although the  
112 greatest benefits are anticipated to be attained with strongly coupled assimilation, we investi-  
113 gate whether the weakly coupled approaches being pursued by operational centres are likely  
114 to provide a determinable improvement on the current uncoupled systems. We consider if the  
115 potential added benefits of strongly coupled assimilation ultimately outweigh the challenges  
116 their development presents.

117 We begin, in section 2., with the formulation of the general incremental 4D-Var algorithm  
118 and a description of the different approaches to coupled atmosphere-ocean 4D-Var data as-  
119 simulation. We introduce our coupled 1D model system in section 3.. In section 4. we give  
120 details of a set of identical twin experiments designed to investigate and compare the be-  
121 haviour and sensitivities of the different approaches. Results are presented in section 5.. A  
122 summary and conclusions are given in section 6..

## 123 **2. Incremental 4D-Var data assimilation**

124 Variational methods form the basis of most operational numerical weather prediction (NWP)  
125 data assimilation systems (Gauthier et al., 1999; 2007; Rabier et al., 2000; Rawlins et al.,  
126 2007; Huang et al., 2009). Our system has therefore been designed using the incremental  
127 4D-Var approach. In this formulation the solution to the full non-linear 4D-Var minimisation

128 problem is replaced by a sequence of minimisations of linear quadratic cost functions such  
 129 that the control variable in the minimisation problem is the increment to the current estimate  
 130 rather than the model state itself. The method was originally developed to overcome the  
 131 cost and practical difficulties involved in solving the complete non-linear problem directly  
 132 in large scale systems (Courtier et al., 1994). We choose to employ the incremental 4D-  
 133 Var formulation for this study as it allows us not only to emulate the methodologies being  
 134 developed for operational systems, but also to explore the type of benefits that could be gained  
 135 by moving towards strongly coupled assimilation systems, thereby providing a benchmark  
 136 for the assessment of weakly coupled assimilation systems. We describe each of the different  
 137 coupled 4D-Var assimilation strategies in detail in section 2.1.. To aid these descriptions, we  
 138 begin with an outline of the steps of the general incremental 4D-Var algorithm.

139 Let  $\mathbf{x}_i \in \mathbb{R}^m$  denote the model state vector, representing the system state at a given time  $t_i$ .  
 140 Then given the discrete non-linear dynamical system model

$$141 \quad \mathbf{x}_i = \mathcal{M}(t_i, t_0, \mathbf{x}_0), \quad i = 0, \dots, n, \quad (1)$$

142 a background, or first guess,  $\mathbf{x}_0^b \in \mathbb{R}^m$ , at  $t_0$ , and imperfect observations  $\mathbf{y}_i \in \mathbb{R}^{r_i}$  at times  
 143  $t_i, i = 0, \dots, n$ , the full non-linear 4D-Var problem is to find the initial model state,  $\mathbf{x}_0$ , that  
 144 minimises the cost function

$$145 \quad J(\mathbf{x}_0) = \frac{1}{2} (\mathbf{x}_0^b - \mathbf{x}_0)^T \mathbf{B}_0^{-1} (\mathbf{x}_0^b - \mathbf{x}_0) \\
 146 \quad + \frac{1}{2} \sum_{i=0}^n (\mathbf{y}_i - h_i(\mathbf{x}_i))^T \mathbf{R}_i^{-1} (\mathbf{y}_i - h_i(\mathbf{x}_i)). \quad (2)$$

147 Here  $h_i : \mathbb{R}^m \rightarrow \mathbb{R}^{r_i}$  is a (generally) non-linear observation operator and  $\mathbf{B}_0 \in \mathbb{R}^{m \times m}$  and  
 148  $\mathbf{R}_i \in \mathbb{R}^{r_i \times r_i}$  are the background and observation error covariance matrices respectively.

149 For the incremental 4D-Var approach, rather than minimise (2) directly we define the incre-

150 ment, at a given time  $t_i$  and outer-loop iteration  $k$ , as

$$151 \quad \delta \mathbf{x}_i^{(k)} = \mathbf{x}_i^{(k+1)} - \mathbf{x}_i^{(k)}, \quad (3)$$

152 and solve iteratively as described below (Lawless et al., 2005).

153 For  $k = 0, 1, \dots, K$  outer-loops, or until desired convergence is reached:

154 (i) For the first iteration set  $\mathbf{x}_0^{(0)} = \mathbf{x}_0^b$ .

155 (ii) Run the non-linear model (1) to obtain  $\mathbf{x}_i^{(k)}$  at each time  $t_i$ .

156 (iii) Compute the innovations

$$157 \quad \mathbf{d}_i^{(k)} = \mathbf{y}_i - h_i(\mathbf{x}_i^{(k)}). \quad (4)$$

158 (iv) Minimise the least squares cost function

$$159 \quad J^{(k)}(\delta \mathbf{x}_0^{(k)}) = \frac{1}{2} \left( (\mathbf{x}_0^b - \mathbf{x}_0^{(k)}) - \delta \mathbf{x}_0^{(k)} \right)^T \mathbf{B}_0^{-1} \left( (\mathbf{x}_0^b - \mathbf{x}_0^{(k)}) - \delta \mathbf{x}_0^{(k)} \right) \\ 160 \quad + \frac{1}{2} \sum_{i=0}^n \left( \mathbf{d}_i^{(k)} - \mathbf{H}_i \delta \mathbf{x}_i^{(k)} \right)^T \mathbf{R}_i^{-1} \left( \mathbf{d}_i^{(k)} - \mathbf{H}_i \delta \mathbf{x}_i^{(k)} \right), \\ 161 \quad := J_b^{(k)} + J_o^{(k)}, \quad (5)$$

162 subject to

$$163 \quad \delta \mathbf{x}_i^{(k)} = \mathbf{M}(t_i, t_0, \mathbf{x}^{(k)}) \delta \mathbf{x}_0^{(k)}. \quad (6)$$

164 (v) Update  $\mathbf{x}_0^{(k+1)} = \mathbf{x}_0^{(k)} + \delta \mathbf{x}_0^{(k)}$ , and return to step (ii).

165 In (5) the operator  $\mathbf{H}_i \in \mathbb{R}^{r_i \times m}$  is the tangent linear of the non-linear observation operator,  
166  $h_i$ , and  $\mathbf{M}$  is the tangent linear (TL) of the non-linear model (NLM) operator  $\mathcal{M}$ . For each  
167 outer-loop,  $k$ , the linearised operators  $\mathbf{H}$  and  $\mathbf{M}$  are evaluated at the current estimate of the  
168 non-linear trajectory,  $\mathbf{x}^{(k)}$ , referred to as the linearisation state.

169 Step (iv) is referred to as the ‘*inner-loop*’. The minimisation of the cost function (5) is  
170 performed iteratively using a gradient descent algorithm. For each iteration of the inner-loop  
171 minimisation the tangent linear model (6) is integrated to give the evolution of the increment

172 for the cost function computation (5) and the adjoint of the TL model,  $M^T$ , is integrated  
173 to obtain the cost function gradient. In this study we employ an off the shelf optimisation  
174 algorithm based on the conjugate gradient method (Shanno, 1978; Shanno and Phua, 1980).

### 175 *2.1. Coupled data assimilation*

176 Our system has been designed to enable several different 4D-Var configurations: an un-  
177 coupled atmosphere-only or ocean-only assimilation, a weakly coupled assimilation and a  
178 strongly coupled assimilation, thus allowing a systematic comparison of the different strate-  
179 gies for treating the coupled 4D-Var data assimilation problem. In this section we give details  
180 of each algorithm and highlight the main differences between them.

#### 181 *2.1.1. Strongly coupled incremental 4D-Var*

182 For the strongly (or fully) coupled assimilation system, the state vector,  $\mathbf{x}$ , and the incremen-  
183 tal 4D-Var control vector,  $\delta\mathbf{x}$ , consist of both the atmosphere and ocean prognostic variables.  
184 The coupled model is used in both the outer and inner-loops; the non-linear coupled model  
185 is used in the outer-loops to generate the linearisation trajectory (1) and compute the innova-  
186 tion vectors (4), and the inner-loop cost function minimisation is performed using the tangent  
187 linear and adjoint of the coupled non-linear model in step (iv). Information is exchanged  
188 between the atmosphere and ocean components at regular, specified time intervals; the SST  
189 from the ocean model is used in the computation of the atmospheric lower boundary con-  
190 ditions, and the surface heat, moisture and momentum fluxes from the atmosphere model  
191 provide the ocean surface boundary conditions.

192 The incremental 4D-Var algorithm implicitly evolves the background error covariances  
193 across the assimilation window according to the TL model dynamics (e.g. Thépaut et al.

194 (1993; 1996)). This acts to modify the prior background error variance estimates and induce  
195 non-zero correlations between model variables. The use of the fully coupled TL and adjoint  
196 models in the inner-loops of the strongly coupled assimilation system means that we expect  
197 cross-covariance information to be generated between the atmosphere and ocean fields. This  
198 allows observations of one fluid to produce analysis increments in the other and is there-  
199 fore expected to generate more consistent analyses. The design of this system also has the  
200 advantage of allowing for cross-covariances between the atmosphere and ocean errors to be  
201 explicitly prescribed *a priori*. We note, however, that this is a non trivial matter; research on  
202 how to characterise and represent atmosphere-ocean cross covariances within coupled data  
203 assimilation systems is currently underway and will be addressed in a future paper.

#### 204 2.1.2. *Uncoupled incremental 4D-Var*

205 The uncoupled atmosphere and ocean assimilation systems are completely independent. Here,  
206 the state and incremental 4D-Var control vectors are comprised of the atmosphere or ocean  
207 prognostic variables only and a separate inner-loop cost function is used for each model. For  
208 the atmosphere (ocean) the outer-loop linearisation trajectory (1) is taken from a run of the  
209 atmosphere-only (ocean-only) non-linear model, the innovation vectors (step (iii)) are com-  
210 puted using the available atmosphere (ocean) observations and the inner-loop minimisation  
211 (step (iv)) uses the corresponding uncoupled atmosphere (ocean) tangent linear and adjoint  
212 model. There is no exchange of information between the two systems at any stage; the SST  
213 used at the atmosphere bottom boundary and the momentum, heat and freshwater fluxes at  
214 the ocean surface boundary are prescribed. Although this approach has its advantages, such  
215 as ease of implementation and modularity, it does not allow for cross-covariances between  
216 the atmosphere and ocean fields and atmospheric (ocean) observations cannot influence the

217 ocean (atmosphere) analysis. The lack of feedback means that the atmosphere and ocean  
218 analysis states are unlikely to be in balance and this can have a negative impact if they are  
219 used to initialise a coupled model forecast (Balmaseda and Anderson, 2009).

### 220 *2.1.3. Weakly coupled incremental 4D-Var*

221 The weakly coupled 4D-Var algorithm is a combination of the strongly and uncoupled al-  
222 gorithms; the system uses a coupled full state vector but uncoupled atmosphere and ocean  
223 incremental 4D-Var control vectors. This approach has the advantage that it limits the amount  
224 of new technical development required when independent atmosphere and ocean assimilation  
225 systems are already in place. The outer-loop linearisation trajectory (1) is generated using the  
226 coupled non-linear model but separate inner-loop cost functions (step (iv)) are defined for the  
227 atmosphere and ocean, using the respective uncoupled atmosphere- or ocean-only tangent lin-  
228 ear and adjoint models and assimilating the atmosphere or ocean observations only. Although  
229 the computation of the innovations (step (iii)) uses only the atmosphere or ocean observa-  
230 tions, the observation-model fit is measured against the coupled model state. The ocean SST  
231 from the coupled outer-loop linearisation trajectory is used in the computation of the bottom  
232 boundary conditions for the uncoupled atmospheric TL model and the surface heat, moisture  
233 and momentum fluxes from the coupled outer-loop linearisation trajectory are used in the  
234 computation of the surface boundary conditions for the uncoupled ocean TL model. Once the  
235 uncoupled atmosphere and ocean inner-loop minimisations have been performed, the uncou-  
236 pled atmosphere and ocean analysis increments are combined and added to the current guess  
237 to provide the initial coupled state for the next outer-loop iteration.

238 Analogous to the uncoupled case, the separation of the atmosphere and ocean TL model  
239 components in the inner-loops of the weakly coupled system means that cross-covariances

240 between the atmosphere and ocean are ignored; they can only be generated between atmo-  
241 sphere fields or between ocean fields. However, as we demonstrate in sections 5.4. and 5.5.,  
242 observations of one fluid are able to influence the analysis of the other if multiple outer-loops  
243 are performed due to the linearisation state being updated.

244  
245 Here, we have presented the coupled 4D-Var assimilation strategies in their cleanest forms.  
246 It should be noted that in practice there will be variations in their application. For example, the  
247 uncoupled analysis systems at the ECMWF and Met Office run the atmosphere component  
248 with a prescribed SST but then use the updated fluxes from this analysis to constrain the  
249 ocean assimilation system. It is also not necessary for the uncoupled and weakly coupled  
250 systems to use the same assimilation window length for the atmosphere and ocean, or even  
251 the same assimilation scheme. Both the existing uncoupled analysis systems and the weakly  
252 coupled systems currently under development at the ECMWF and Met Office use 4D-Var for  
253 the atmosphere and 3D-Var FGAT (first guess at appropriate time) for the ocean (Laloyaux  
254 et al., 2014; 2015; Lea et al., 2015).

### 255 **3. The coupled model**

256 The objective of this study is to gain a greater theoretical understanding of the coupled  
257 atmosphere-ocean data assimilation problem by exploring and comparing the behaviours of  
258 the coupling strategies presented in section 2.1.. Idealised models offer an effective frame-  
259 work for investigating and advancing new methods, avoiding unnecessary complexities that  
260 can obscure results. Using a simplified system allows us to perform a range and quantity of  
261 experiments that would require a significant amount of technical development and resources  
262 to execute in a full scale system.

263 In many cases, finding balanced solutions to the coupled atmosphere-ocean assimilation  
264 problem is primarily a vertical problem of the two boundary layers. A 1D column atmosphere-  
265 ocean model framework therefore offers a tractable and relevant approach. Whilst it is prefer-  
266 able to keep the model as simple as possible from a developmental point of view, it is impor-  
267 tant to ensure that processes crucial to realistic air-sea interaction, such as the diurnal SST  
268 cycle and evolution of surface forcing, are adequately represented. Our new system has been  
269 built by coupling the ECMWF single-column atmospheric model (SCM) to a single-column  
270 K-Profile Parameterisation (KPP) ocean mixed layer model. The use of these models ensures  
271 that the simplified system retains the key elements of coupling processes in a fully coupled  
272 ocean-atmosphere model without being overly complex.

### 273 *3.1. Non-linear models*

274 The atmosphere model solves the primitive equations for temperature,  $T$ , specific humid-  
275 ity,  $q$ , and zonal,  $u$ , and meridional,  $v$ , wind components, formulated in non-spherical co-  
276 ordinates (Simmons and Burridge, 1981; Ritchie et al., 1995). Compared to the original  
277 SCM, our model does not include the parameterisation of physical processes such as radi-  
278 ation, subgrid-scale orographic drag, convection, clouds and surface/soil processes; we also  
279 use a simplified vertical diffusion scheme. For the ocean, the evolution of the temperature,  $\theta$ ,  
280 salinity,  $s$ , and zonal and meridional currents  $u_o, v_o$  are described following the formulation  
281 of (Large et al., 1994). Further details, including the model equations, are given in appendix  
282 A.

283 The coupling of the two models takes place at the atmosphere-ocean boundary. In the at-  
284 mosphere, the lower boundary conditions for temperature and specific humidity depend on  
285 the SST (temperature at top level of ocean model which corresponds to a depth of 1m) and

286 saturation specific humidity,  $q_{sat}(SST)$ , and a no-slip condition is used for the  $u$  and  $v$  wind  
287 components. The ocean surface boundary conditions for temperature and salinity are given  
288 by the surface kinematic fluxes of heat and salt which in turn depend on the net long wave  
289 radiation and latent and sensible heat fluxes. The surface boundary conditions for the  $u$  and  
290  $v$  components of the current are given by the zonal and meridional components of the kine-  
291 matic momentum flux. At each model time step, the atmosphere model computes and passes  
292 the latent and sensible heat fluxes and the horizontal components of the surface momentum  
293 flux to the ocean model. The updated ocean model SST is passed back to the atmosphere  
294 where it is used in the computation of the atmosphere lower boundary conditions for the next  
295 step.

296 A fuller description of the individual atmosphere and ocean non-linear model components  
297 and their coupling is given in appendix A. This system combined with our 4D-Var schemes  
298 provides a unique and tractable framework for addressing the coupled atmosphere-ocean as-  
299 simulation problem.

300

301 As part of the assimilation system development, the simplified non-linear model was vali-  
302 dated against the original (full physics) version of the ECMWF SCM code. As expected, we  
303 see small differences in the evolution of both the prognostic variables and surface fluxes but  
304 in general we find that using the simplified physics provides a good approximation to the full  
305 physics in the coupled model. Where there are differences the simplified model still produces  
306 an evolution that is physically reasonable, with a diurnal cycle in the ocean SST and mixed  
307 layer depth and appropriate atmosphere-ocean fluxes; see, for example, the truth trajectory

308 (black line) in figure 4. We are therefore confident that the model is sufficient for assessing  
309 the different assimilation strategies.

### 310 3.2. *Tangent linear and adjoint models*

311 In order to be able to compute the cost function and its gradient for each inner-loop we need  
312 to develop the tangent linear (TL) and adjoint models. A particular issue worth noting is the  
313 linearisation of the atmosphere and ocean vertical turbulent flux parameterisations. The for-  
314 mulation of the diffusion coefficients  $K_\phi$  in both the atmosphere and ocean vertical diffusion  
315 schemes (equations (A6), (A13) of appendix A) is strongly non-linear and its linearisation  
316 has been shown to be unstable (Laroche et al., 2002). The simplest way to avoid difficulties  
317 associated with this linearisation is to neglect the perturbation of the  $K_\phi$  coefficients. Studies  
318 such as Janisková et al. (1999); Mahfouf (1999) and Laroche et al. (2002) have shown that  
319 a TL diffusion scheme can still produce reasonable and useful behaviour under this assump-  
320 tion and this approach has been widely adopted in both atmosphere and ocean assimilation  
321 systems (e.g. Mahfouf (1999); Weaver et al. (2003)). We are therefore satisfied that this sim-  
322 plification is appropriate for our system. During the assimilation the  $K_\phi$  are computed for  
323 each non-linear outer-loop and then held constant for the inner-loop minimisation.

324 Although this means we are using an approximate TL model rather than the exact TL, since  
325 the adjoint model is derived from the approximate TL model, the inner-loop cost function  
326 gradient calculation contains the correct information for convergence of the minimisation  
327 problem. The correctness of the tangent linear and adjoint model codes, and the gradient  
328 calculation were all verified using standard tests (e.g. Navon et al. (1992); Lawless (2013)).

## 329 4. Experimental design

330 We compare the performance of the strongly coupled, weakly coupled and uncoupled 4D-  
331 Var systems via a series of identical twin experiments in which the coupled non-linear model  
332 is used to forecast a reference or ‘truth’ trajectory from which synthetic observations are  
333 generated. The assimilation systems are then assessed on how well they approximate the  
334 initial reference state and subsequent forecast.

### 335 4.1. Initial conditions and forcing

336 The atmospheric initial conditions, surface pressures, and SST data used to force the un-  
337 coupled atmosphere system are taken from the ERA Interim Re-analysis<sup>1</sup>(Dee et al., 2011).  
338 Fields are available at 6 hourly intervals and can be extracted on model levels so that they  
339 do not need pre-processing. These data are also used to estimate the geostrophic wind com-  
340 ponents,  $u_g, v_g$  in equations (A1) and (A2), and large scale horizontal forcing terms (see  
341 appendix A) using simple centred finite difference approximations across adjacent latitude  
342 and longitude points. .

343 Initial ocean fields are produced by interpolating Mercator Ocean reanalysis data<sup>2</sup>(Lellouche  
344 et al., 2013), onto the KPP model grid. The surface short and long wave radiation forc-  
345 ing fields are computed by running the full physics version of the coupled single column  
346 model and taking 6 hourly snapshots of the diagnostic clear-sky radiation flux fields that are  
347 computed as part of the radiation scheme (ECMWF IFS documentation, 2001–2013b). The  
348 geostrophic components of the ocean currents (section A2.1.) are estimated by computing a  
349 10 day rolling average of the Mercator ocean currents. The surface heat, moisture and mo-

<sup>1</sup> ERA Interim Re-analysis data can be downloaded via the ECMWF data server at [www.ecmwf.int](http://www.ecmwf.int)

350 momentum fluxes required for uncoupled ocean model integrations are taken from the ERA  
 351 Interim Re-analysis.

352 For the experiments presented here, the true initial state,  $\mathbf{x}_0$ , is a 24 hour coupled model  
 353 forecast from 00:00 UTC on 2nd June 2013 to 00:00 UTC on 3rd June 2013 for the point  
 354 (188.75°E, 25°N) which is located in the north west Pacific ocean. This forecast was ini-  
 355 tialised using ERA interim and Mercator Ocean reanalysis data; we denote this initial fore-  
 356 cast state as  $\mathbf{x}_{-24}$ . We run a forecast rather than initialise from these data directly in order to  
 357 generate an initial state that is consistent with the coupled model dynamics.

#### 358 4.2. Background

359 The initial background state,  $\mathbf{x}_0^b$ , is generated by running a second 24 hour coupled model  
 360 forecast from 00:00 UTC on 2nd June 2013 with perturbed initial data; this data, denoted  
 361  $\hat{\mathbf{x}}_{-24}$ , is generated by adding random Gaussian noise to  $\mathbf{x}_{-24}$ ,

$$362 \quad \hat{\mathbf{x}}_{-24} = \mathbf{x}_{-24} + \boldsymbol{\sigma} \circ \delta\mathbf{x}. \quad (7)$$

363 Here, the  $\delta\mathbf{x}$  are normally distributed, random perturbations and  $\boldsymbol{\sigma} \in \mathbb{R}^m$  is a vector of  
 364 standard deviations; these are computed as the sample standard deviation of the unperturbed  
 365 coupled model forecast states at each model time step from  $\mathbf{x}_{-24}$  to  $\mathbf{x}_0$ . The initial background  
 366 guess for the assimilation is then given by the perturbed coupled model forecast state after 24  
 367 hours, i.e.  $\mathbf{x}_0^b = \hat{\mathbf{x}}_0$ . The true and initial background states are shown in figure 1.

368 For the purposes of this study, we assume that the background error covariance matrix  $\mathbf{B}_0$   
 369 is diagonal, that is, the initial background errors are univariate and spatially uncorrelated; the  
 370 diagonal elements,  $\sigma_b^2$ , representing the error variances, are assumed to vary for each model

<sup>2</sup> Mercator Ocean re-analysis data are available via the MyOcean project Web Portal at [www.myocean.eu.org](http://www.myocean.eu.org).

371 field and vertical level and are taken to be the squared values of the standard deviations  $\sigma$   
372 from (7). The initial background error standard deviation profiles are shown in figure 2. The  
373 relative magnitude of the standard deviations of the atmosphere fields are greater than the  
374 ocean due to the faster timescales. For the ocean temperature and salinity, the variability,  
375 and thus the prescribed background error standard deviation, is largest in the turbulent mixed  
376 layer region ( $\sim$ top 50 m) where timescales are shortest. Moving deeper into the ocean the  
377 timescales become longer and the standard deviations become very small.

378 The assumption of a diagonal matrix  $\mathbf{B}$  is a great simplification but is used here as an aid  
379 to understanding the implicit evolution of the error covariances by the 4D-Var algorithm.  
380 Although we assume that the prior atmosphere and ocean fields are uncorrelated, the incre-  
381 mental 4D-Var algorithm implicitly propagates the background error covariances across the  
382 assimilation window according to the TL model dynamics (see Bannister (2008a) and refer-  
383 ences therein). This acts to modify the prior background error variance estimates and induce  
384 non-zero correlations between model variables.

385 A simple preconditioning of the inner-loop cost function using the square root of the back-  
386 ground error covariance matrix was found to be beneficial in terms of improving the con-  
387 ditioning of the system and allowing convergence of the inner-loop minimisation within a  
388 reasonable number of iterations (Courtier, 1997). Preconditioning is common in most op-  
389 erational variational assimilation systems and is often implemented using a control variable  
390 transform (Bannister, 2008b).

### 391 *4.3. Observations*

392 We assume that the model state variables are observed directly to avoid the additional com-  
393 plexity of a non-linear observation operator. Observations are generated by adding uncorre-

lated random Gaussian errors, with given standard deviations (see table 1) to the reference trajectory at constant time and space intervals. Observations of atmospheric temperature, and  $u$  and  $v$  wind components are assimilated at 17 of the 60 atmosphere model levels; these are chosen to approximately correspond to the standard pressure levels used, for example, to report radiosonde data (see table 2). Observations of ocean temperature, salinity, and zonal and meridional currents are assimilated at 23 of the 35 ocean model levels giving vertical frequency comparable to a XBT profile (see table 3). Note that since the atmospheric model does not include the parameterisation of processes such as moist convection, clouds and precipitation we do not assimilate observations of specific humidity,  $q$ . Ocean observational data are typically available less frequently than atmospheric observational data, particularly for certain operational observing systems. The atmosphere and ocean observation frequencies used in our assimilation experiments were chosen to reflect this disparity. Unless otherwise stated, results refer to experiments run with atmosphere observations at 3, 6, 9 and 12 hours, and ocean observations at 6 and 12 hours.

Although it is generally accepted that observation error covariances exist it is typical to ignore them and in practice it is assumed that the errors in the observational data are spatially and temporally uncorrelated so that the observation error covariance matrices  $\mathbf{R}_i$  is diagonal (Daley, 1991). We follow the same approach here but also keep the observation network fixed for the duration of each experiment so that the number of observations  $r_i = r$  and  $\mathbf{R}_i = \mathbf{R}$  for all  $i$ . The observation error variances,  $\sigma_o^2$ , are assumed to be constant across all vertical levels for each observation type.

## 415 5. Assimilation results

416 Since our aim is to examine the impact of coupled assimilation on the atmosphere and ocean  
417 boundary layers we limit our discussion to this region and focus on the results in the bot-  
418 tom  $\sim 200$  hPa of the atmosphere model ( $\sim 15$  levels), and top 50 m (26 levels) of the ocean  
419 model. We use a 12 hour assimilation window with 3 outer-loops and a model time-step of  
420 15 minutes. A 12 hour window length is common for atmospheric data assimilation systems,  
421 such as the ECMWF IFS. Initial experiments using a greater number of outer-loops showed  
422 that the cost function usually converged after around 3 loops. The inner-loop minimisation  
423 is terminated when the relative change in gradient is less than 0.001 (Lawless and Nichols,  
424 2006); for our system this is typically after around 10-20 iterations.

425 Figure 3 shows the absolute (truth - analysis) error profiles at initial time  $t_0$  for each of the  
426 prognostic model variables for the three assimilation systems. The differences between the  
427 analyses are most pronounced in the upper ocean temperature and  $u$ ,  $v$  current fields. The  
428 atmospheric temperature and specific humidity analysis errors are very similar to the initial  
429 background errors for all three systems. For specific humidity this is expected since we do  
430 not observe this field. For atmospheric temperature this may be in part due to the fact that the  
431 initial background errors are small in this region. There are also relatively fewer observations  
432 in the lower atmosphere compared to the upper ocean. There are clearer improvements in  
433 the near surface  $u$  and  $v$  wind fields. Here, the analysis errors for the uncoupled and weakly  
434 coupled systems are very alike whereas the strongly coupled system appears to use the near  
435 surface ocean current observations to further correct the  $u$  and  $v$  wind analyses. This is, in  
436 part, due to the way the coupled and uncoupled non-linear models and assimilation systems  
437 are formulated.

438 A notable aspect in the ocean analysis errors is at approximately 20m, which coincides  
439 with the mixed layer depth. The mixed layer depth is characterised by a sharp gradient in  
440 the temperature and salinity profiles. In the background estimate the position of this feature  
441 is incorrect. When the assimilation of observations attempts to correct this positional error,  
442 instead of shifting the profiles, it erroneously changes the structure of the temperature and  
443 salinity profiles so that the error in the analysis is actually increased compared to the back-  
444 ground. This is an issue for all three coupling strategies and is a well documented problem  
445 in the atmosphere when assimilating observations of the analogous boundary layer capping  
446 inversion (Fowler et al., 2012).

447 It is not possible to draw conclusions on the performance of each approach from the anal-  
448 ysis errors alone. In particular, these results do not give any indication of whether the initial  
449 atmosphere and ocean analysis states are in balance. Since one of the key drivers behind the  
450 development of coupled data assimilation systems is generation of a consistent system state  
451 for the initialisation of coupled model forecasts, we use the analysis fields at the beginning  
452 of the assimilation window to initialise a series of coupled model forecasts; the results are  
453 discussed in sections 5.1. to 5.3..

### 454 *5.1. Initialisation shock*

455 A major problem with using analysis states from uncoupled assimilation systems to initialise  
456 a coupled model forecast is that the atmosphere and ocean fields may not be balanced and  
457 this can lead to initialisation shock. If the initial conditions are not on the coupled model  
458 attractor (in these twin experiments also the true attractor) the forecast will experience an  
459 adjustment process. In some cases the adjustment towards the model attractor solution occurs  
460 asymptotically but in others it manifests itself as a rapid change in the model fields in the

461 early stages of the forecast (Balmaseda, 2012). The skill of a coupled model forecast depends  
462 strongly on the way it is initialised, thus the reduction or elimination of initialisation shock is  
463 particularly important in seasonal forecasting (Balmaseda and Anderson, 2009).

464 Figure 4 compares the SST and surface fluxes for the first 48 hours of each coupled model  
465 forecast against the truth trajectory and also a forecast initialised from the initial background  
466 state (i.e. no assimilation). In all cases, the forecast eventually tracks the true trajectory fairly  
467 well but there is variation in behaviour during the first part of the forecast window. There is  
468 evidence of initialisation shock in the SST field. The initial SST from the uncoupled ocean  
469 analysis is furthest from the true initial SST ( $\sim 0.5$  K warmer) and when the coupled model  
470 is initialised from the combined uncoupled atmosphere and ocean analysis states the forecast  
471 SST increases sharply, even further away from the true SST, over the first 5 model time-steps  
472 before gradually converging back towards the true trajectory. We also see jumps in the SST  
473 forecasts initialised from the strongly and weakly coupled analyses but these are much smaller  
474 suggesting that the coupled analyses are more balanced. Note that the differences between the  
475 SST forecasts can most easily be seen in figure 6 which focuses in on the first 12 hours of  
476 the forecast window. In this example, the error in the weakly coupled SST analysis at the  
477 initial time is actually smaller than the strongly coupled SST analysis and the SST forecast  
478 from the weakly coupled analysis initially tracks the truth more closely. However, later in  
479 the forecast window, at the peak of the diurnal cycle ( $\sim 25$  hours), the SST forecasts from  
480 both the weakly and uncoupled analyses unexpectedly diverge from the truth, whereas the  
481 strongly coupled analysis continues to track it closely. This could be interpreted as a further  
482 indication of greater balance in the strongly coupled analysis; although the initial error in  
483 the SST forecast from the strongly coupled analysis is greater than the weakly coupled, for

484 this example, it appears to be in better balance with rest of the model. The error in the initial  
485 temperature at the bottom atmosphere level is very similar for all three forecasts but if we  
486 examine the atmosphere-ocean temperature difference (figure 5) we see that although at times  
487 the weakly and uncoupled systems are closer to the truth than the strongly coupled system,  
488 the temperature difference for the strongly coupled system is more stable across the whole  
489 forecast period. It also tracks the truth very accurately during the first 12 hours of the forecast  
490 which is the period corresponding to the assimilation window. The pattern seen in the sensible  
491 heat flux forecasts in figure 4 would also support this.

492 The strongly coupled analysis also produces better forecasts of the surface wind speed and  
493  $u$ ,  $v$  wind stress components (figure 4). The forecasts initialised from the weakly coupled  
494 and uncoupled analyses capture the general phasing of these fields but their magnitudes are  
495 overestimated to a greater extent than in the strongly coupled case over the first 24-48 hours  
496 of the forecast.

497 The latent heat flux forecasts are the slowest to stabilise; this is likely to be due to the fact  
498 that we are not assimilating observations of specific humidity. However, as the forecasts adjust  
499 towards the model attractor, there is a clear pattern of increasing accuracy as we progress from  
500 uncoupled to weakly to strongly coupled initialisation.

501 Overall our experiments have shown that, when compared to uncoupled initialisation, ini-  
502 tialisation using the analysis from a coupled assimilation can help to reduce initialisation  
503 shock and its impact on the subsequent forecast. For our model, the benefit appears to be  
504 greatest with the strongly coupled system; the weakly coupled assimilation system is also ca-  
505 pable of reducing shock, but its behaviour is less consistent. It is worth noting that we would  
506 expect the type and size of shocks produced by each assimilation system to vary depending

507 on factors such as season, location, assimilation start time and initial background state, and  
508 whether a full diurnal cycle is being observed.

## 509 *5.2. Uncoupled assimilations with ‘true’ SST and surface fluxes*

510 To understand how much the accuracy of the prescribed SST and surface fluxes may affect  
511 the results of the uncoupled assimilation we repeat the experiments performed in section  
512 5.1. using 6 hourly snapshots of the fluxes from the ‘truth’ trajectory in place of the forcing  
513 ERA-interim fields. In some sense this is the best we may expect the uncoupled assimila-  
514 tion system to be. Figure 6 shows the forecast fluxes for this case alongside those from the  
515 original experiment for the first 12 hours of the forecast. Although we see an improvement,  
516 with reduced shocks in the SST and latent and sensible heat fluxes, the strongly coupled as-  
517 simulation still generally performs better than the uncoupled assimilation. Even with the true  
518 forcing data, the uncoupled systems suffer from the lack of atmosphere-ocean feedback. The  
519 weakly coupled SST analysis still produces a better SST forecast than the uncoupled ocean  
520 SST analysis, but the corresponding surface flux forecasts are either very similar or slightly  
521 worse for the weakly coupled case. This was also verified in experiments with different test  
522 cases (not shown). Overall, the performance of our weakly coupled assimilation system is  
523 usually comparable to the uncoupled system with the ‘true’ forcing. This indicates that even  
524 moving to a weakly coupled assimilation system may be of benefit. Furthermore, if multiple  
525 outer-loops are used, the update of the SST and surface fluxes can provide useful information  
526 not available to the uncoupled systems, as we demonstrate in section 5.4.

### 5.3. Temporal frequency of observations

To test the sensitivity of the different approaches to the frequency of observations we repeat the assimilation experiments with the frequency of the atmosphere observations reduced by half to 6 hourly, so that we are observing the atmosphere and ocean with the same frequency. The strongly coupled system still performs better than the weakly coupled and uncoupled systems. The most significant effect is that differences between forecasts initialised from the weakly coupled and uncoupled analyses are less pronounced over the first 12 hours. This is best illustrated through the SST and surface flux forecasts (figures 7 and 8). If we compare these with figures 4 and 6 we see that both the strongly and weakly coupled  $t_0$  SST estimates are further from the true value than in the previous case. The forecast initialised from the strongly coupled analysis adjusts itself smoothly, but the forecast initialised from the weakly coupled analysis exhibits a much larger shock with amplitude similar to the uncoupled case. There is also more drift in the SST forecasts in the second half of the forecast window for this case.

The change in the SST trajectories means that there is now also less of a clear gap between the latent and sensible heat fluxes for the forecasts initialised from the uncoupled and weakly coupled analyses. There is no real change to the pattern of behaviour in the wind stresses, but the over estimation of magnitude is slightly larger due to the increased errors in the near surface wind forecasts.

These experiments have shown that the weakly coupled assimilation system appears to be much more sensitive to the observation frequency than the strongly coupled system. This is because the weakly coupled assimilation system is, unlike the strongly coupled system, predominantly exposed to the coupling of the atmosphere and ocean through the innovations

550 which are by definition in observation space. Therefore, if the number of observations is de-  
551 creased, either spatially or temporally, this will clearly impact the greatest on the weakly  
552 coupled assimilation system. Although weak coupling can reduce shock, as seen in section  
553 5.1., at its worst it can produce results that are very similar to the uncoupled assimilation sys-  
554 tem. Similar behaviour was found in experiments varying the background error variances (not  
555 shown). Since changing the background error standard deviations has the effect of changing  
556 the relative weight given to the observations, the performance of the weakly coupled system  
557 was found to be more sensitive than the strongly coupled system to a change in the prescribed  
558 background error standard deviations.

#### 559 *5.4. Single observation experiments*

560 A big hope for coupled data assimilation systems is that they will enable greater use of near-  
561 surface observations, such as satellite SST measurements and scatterometer data, by allowing  
562 cross-covariance information between the atmosphere and ocean. We investigate this prop-  
563 erty by assimilating single observations of near-surface variables. Since the initial background  
564 error covariance matrix is diagonal, increments from a single observation at the end of the as-  
565 similation window provide insight into the implicit covariances generated by the 4D-Var sys-  
566 tem. The purpose of the experiments presented here is to illustrate the ability of the strongly  
567 and weakly coupled systems to induce cross-correlations between the atmosphere and ocean  
568 rather than investigate their size and structure.

569 To understand the impact of SST observations, we assimilate the temperature from the top  
570 ocean level (1 m depth) at the end of the 12 hour assimilation window. Figure 9 shows the  
571 (analysis-background) increments produced by the strongly and weakly coupled systems at  
572 initial time,  $t_0$ . We see initial increments in atmospheric temperature, specific humidity, ocean

573 temperature and salinity with the strongly coupled system, but only ocean temperature and  
574 salinity for the weakly coupled system.

575 We can relate the behaviour observed in this experiment back to the model equations and  
576 assimilation system design. The strongly coupled system uses the coupled tangent linear and  
577 adjoint models in its inner-loops. Since the boundary conditions for the atmospheric temper-  
578 ature and specific humidity depend on the SST and the boundary conditions for the ocean  
579 temperature and salinity depend on the atmospheric temperature and humidity via the latent  
580 and sensible heat fluxes, an increment or perturbation to the SST should produce increments  
581 in the atmosphere temperature and specific humidity fields. We do not see increments in the  
582 initial  $u$  and  $v$  wind fields because assumptions made in the development of the tangent lin-  
583 ear and adjoint models mean that the SST does not directly depend on them. The weakly  
584 coupled system cannot produce initial increments in the atmosphere fields because it runs  
585 separate inner-loops for the atmosphere and ocean which use the uncoupled atmosphere only  
586 and uncoupled ocean only tangent linear and adjoint models. The SST used in the bound-  
587 ary conditions for the inner-loop uncoupled atmosphere model and the surface fluxes used  
588 in the boundary conditions of the inner-loop uncoupled ocean model are prescribed from the  
589 outer-loop linearisation trajectory.

590 The initial analysis increments modify the coupled model trajectory and produce incre-  
591 ments to the background fields for all variables across the rest of the assimilation window;  
592 figure 10 shows the analysis increments in the centre of the assimilation window ( $t = 6$  hr)  
593 as an example. The changes in the initial atmospheric temperature, specific humidity, ocean  
594 temperature and salinity fields subsequently produce increments to all the model variables.  
595 However, these increments are relatively small compared to the full fields and so when we

596 examine the analysis trajectories across the whole assimilation window we see that the ocean  
597  $u$ ,  $v$  current and atmosphere  $u$ ,  $v$  wind trajectories are qualitatively very similar for both  
598 the strongly and weakly coupled systems (not shown). There are more visible differences  
599 between the strongly and weakly coupled atmospheric temperature and specific humidity  
600 analysis trajectories due to the difference in increments at  $t_0$  (not shown).

601

602 Scatterometer data provide information on ocean surface wind speed and direction via mea-  
603 surements of backscatter from surface waves. Since our system is only currently designed to  
604 handle direct observations we use (i) horizontal wind components,  $u$  and  $v$ , at the bottom  
605 level of the atmosphere model ( $\sim 10$  m height); (ii) zonal and meridional ocean currents at  
606 the top level of the ocean model as a proxy.

607 With single  $u$  and  $v$  wind observations only the  $u$  and  $v$  wind fields are updated at  $t_0$  for  
608 both the strongly and weakly coupled systems (not shown). These initial wind increments  
609 do, however, produce increments to all of the atmosphere and ocean background fields over  
610 the remainder of the assimilation window. In this case, the strongly and weakly coupled  
611 analyses are identical; this is due to the model formulation and the fact that we are ignoring  
612 perturbations to the diffusion coefficients in the tangent linear and adjoint models as described  
613 in section 3.2.. The  $u$  and  $v$  winds only depend on each other and so are essentially decoupled  
614 from the rest of the model in both the coupled model and atmosphere only model.

615 With  $u$  and  $v$  ocean surface current observations, the strongly coupled system produces ini-  
616 tial increments in all fields, although these are very small for atmospheric temperature and  
617 specific humidity. The weakly coupled system only produces initial increments in the ocean  
618 fields and these are larger than in the strongly coupled case (figure 11). Again, the update of

619 the initial state gives rise to increments in all fields across the assimilation window for both  
620 systems. Although the  $t_0$  ocean analysis increments are larger in the weakly coupled system  
621 the increments across the assimilation window are generally smaller, particularly for the at-  
622 mospheric fields, where there is no change to the initial states (results not shown). There is  
623 no difference in the strongly and weakly coupled SST and surface fluxes when we assimilate  
624 single wind observations and only very small differences in the single SST observation exper-  
625 iment. However, for this case the strongly coupled system produces a much better analysis of  
626 the true surface wind stress and wind speed than the weakly coupled system (figure 12). As  
627 described in section 2.1.1., the strongly coupled system is able to generate cross-covariances  
628 between the atmosphere and ocean fields and thus improve the wind analysis using the ocean  
629 current observations. Improved near-surface wind conditions can have a positive impact on  
630 air-sea exchange and thus both the atmosphere and ocean analyses. This result clearly demon-  
631 strates the potential for greater use of near surface data with strongly coupled assimilation.

632 These experiments have provided a valuable illustration of the ability of a strongly cou-  
633 pled assimilation system to induce cross-covariance information between the atmosphere and  
634 ocean variables, such that a single observation of a variable in one fluid at the end of the  
635 assimilation window can produce increments to variables in the other fluid at initial time  $t_0$ .  
636 Although the structure of the weakly coupled assimilation system does not allow atmosphere-  
637 ocean cross-covariances, there is benefit to be gained from this approach if more than one  
638 outer-loop is used, and particularly if both the atmosphere and ocean are well observed (see  
639 section 5.3.). An analysis increment from an observation in one system will change the lin-  
640 earisation state for both the uncoupled TL models used in the next inner-loop minimisation  
641 and thus has the potential to influence the subsequent analysis across the whole atmosphere-

642 ocean system. In the next section we present a simple illustration of how assimilating a single  
643 observation of both fluids enables the weakly coupled system to modify the initial analysis  
644 increments across the air-sea interface.

### 645 *5.5. Double observation experiment*

646 Due to the inability of the weakly coupled system to produce an initial increment in the un-  
647 observed fluid it is not possible to see explicitly how information is spread across the air-sea  
648 interface in the analysis increments when only one fluid is observed. Unlike in the uncou-  
649 pled system, in the weakly coupled system observations are able to influence the analysis  
650 across the atmosphere-ocean interface. This can be illustrated by assimilating two observa-  
651 tions of temperature at the end of the 12 hour assimilation window; one at the lowest level  
652 of the atmosphere, and one at the top level of the ocean. The atmospheric temperature anal-  
653 ysis increments after three outer-loops are illustrated in figure 13 (solid lines). These can be  
654 compared to the analysis increments when only the atmosphere temperature observation is  
655 assimilated (dashed lines) and when only the ocean temperature observation is assimilated  
656 (dot-dash lines, same as figure 9).

657 It can be seen in the strongly coupled case that only assimilating the atmosphere tempera-  
658 ture observation produces a negative increment, peaking at approximately 980hPa, and assim-  
659 ilating only the ocean temperature observation produces a positive increment, peaking again  
660 at approximately 980hPa. When both these observations are assimilated the strongly coupled  
661 system is able to make use of the overlapping information they provide and the result is a  
662 positive increment which is slightly reduced in magnitude compared to when only the ocean  
663 observation is assimilated. For the weakly coupled case, when only the atmosphere temper-  
664 ature observation is assimilated the atmospheric temperature analysis increment is nearly

665 identical to that produced by the strongly coupled system. When only the ocean temperature  
666 observation is assimilated there is no initial analysis increment in the atmosphere tempera-  
667 ture field (as discussed in section 5.4.). However, when both temperature observations are as-  
668 similated the atmospheric temperature analysis increment becomes larger in magnitude than  
669 when only the atmosphere temperature observation was assimilated. Therefore, the weakly  
670 coupled system is able to use the ocean observation to alter the initial analysis increment in  
671 the atmosphere when atmosphere observations are also assimilated. This is not possible in  
672 the uncoupled system.

## 673 **6. Summary**

674 We have developed an idealised coupled atmosphere-ocean model system and used it to study  
675 different formulations of the coupled atmosphere-ocean data assimilation problem. By em-  
676 ploying the incremental 4D-Var algorithm we have built the capability to run both strongly  
677 and weakly coupled assimilations as well as uncoupled atmosphere or ocean only assim-  
678 ilations. This has provided a flexible framework for comparing the behaviours of varying  
679 degrees of coupling.

680 A key motivation for the development of coupled data assimilation systems is the potential  
681 for the reduction or elimination of initialisation shock in coupled model forecasts via the gen-  
682 eration of more balanced initial conditions, and the positive impact this is expected to have  
683 in terms of forecast skill. Initialisation shocks were seen in SST in our simple system and  
684 experiments showed that, when compared to uncoupled initialisation, coupled assimilation is  
685 able to reduce initialisation shock and its impact on the subsequent forecast, although it may  
686 not eliminate it completely. Whilst this improvement was clearly evident when using analy-  
687 ses from the strongly coupled system, it was not always so obvious with the weakly coupled

688 system in our experiments. The ability of the weakly coupled assimilation system to reduce  
689 initialisation shock was found to be sensitive to the input parameters, such as observation  
690 frequency. In the best cases the behaviour of the SST and surface fluxes in the initial stages  
691 of the forecast (used to identify shock) followed those from the strongly coupled assimila-  
692 tion. In other cases the weakly coupled assimilation did not show the same improvement as  
693 strongly coupled assimilation. However, the weakly coupled system was usually comparable  
694 to uncoupled assimilations in which the atmosphere and ocean models were forced using the  
695 ‘true’ SST and surface fluxes. This illustrates that even moving to a weakly coupled assim-  
696 ilation system should be of benefit, as the update of the SST and surface fluxes through the  
697 outer-loop step can provide useful information not available to the uncoupled assimilation  
698 systems.

699 Single observation experiments were used to demonstrate how coupled assimilation sys-  
700 tems offer the potential for improved use of near-surface observations via the generation of  
701 cross covariance information. Although the possible cross-covariances that can be generated  
702 are partly limited by the simplified dynamics of our model, the effect of coupled assimilation  
703 can clearly be seen. The strongly coupled assimilation system is able to implicitly induce  
704 cross-covariance information between the atmosphere and ocean at the initial time, such that  
705 a single ocean observation can generate analysis increments in the initial atmospheric fields  
706 and vice-versa. While the design of the weakly coupled incremental 4D-Var assimilation al-  
707 gorithm does not allow this, the use of the coupled model in the outer-loop update step means  
708 that if more than one outer-loop is run, an observation in one system can affect the other sys-  
709 tem by changing the linearisation state. With observations in both fluids, the weakly coupled  
710 system is also able to alter the analysis increments across the atmosphere-ocean interface

711 at the initial time and this was illustrated via a double observation experiment. Thus infor-  
712 mation from near-surface observations can be used to greater effect compared to uncoupled  
713 assimilation systems.

714 Overall, the results from experiments with this idealised system support the belief that ben-  
715 efits can be expected from coupled data assimilation systems. In the experiments presented  
716 here and others performed using variations of the set-up described, the strongly coupled as-  
717 simulation system generally outperforms both the weakly coupled and uncoupled systems,  
718 in terms of producing more balanced initial analysis fields, and extracting more information  
719 from observations through the implicit generation of cross-covariances. The results from the  
720 weakly coupled assimilation experiments show that benefit can be gained from such a system,  
721 but that it is unlikely to be as large as that from strongly coupled assimilation. Nevertheless,  
722 even with a weak coupling we may expect some reduction in initialisation shock and the gen-  
723 eration of some cross-covariance information. Thus the current efforts of operational centres  
724 to develop weakly-coupled assimilation systems are a step in the right direction.

725 Further work is required to better understand the sensitivity of the weakly coupled system to  
726 the input parameters of the assimilation. In particular, this study used a diagonal background  
727 error covariance matrix in order to understand more cleanly the covariances generated by the  
728 coupled assimilation. If the weakly-coupled assimilation included non-diagonal background  
729 error covariance matrices in the atmosphere and ocean inner-loop cost functions, then better  
730 balance would be expected in the increments of the individual systems. This may in itself  
731 help to reduce initialisation shock and make better use of observations.

732 Work is now underway to investigate the nature and structure of the atmosphere-ocean  
733 cross-covariances and how they should be represented in both strongly and weakly coupled

734 systems. An increased understanding of the covariance information arising from atmosphere-  
735 ocean coupling will provide valuable guidance for the design of more balanced covariances  
736 for future full scale coupled data assimilation systems.

## 737 **7. Acknowledgments**

738 This work was funded by the European Space Agency (ESA) as part of the Data Assimila-  
739 tion projects - Coupled Model Data Assimilation initiative (contract no. 4000104980/11/I-  
740 LG), and the UK Natural Environment Research Council (NERC) via research grant no.  
741 NE/J005835/1 and as part of NERC's support of the National Centre for Earth Observation  
742 (NCEO). We are also grateful to Keith Haines and colleagues at the ECMWF and UK Met Of-  
743 fice for useful discussions, and to Irina Sandu and Glenn Carver of the ECMWF for providing  
744 source code and technical support. We would like to thank the anonymous reviewers for their  
745 helpful comments and in particular the reviewer who suggested the experiment presented in  
746 section 5.5..

## 747 **APPENDIX A: Atmosphere and ocean model components**

748 This appendix provides further details of the atmosphere and ocean components of the  
749 coupled 1D model system.

### 750 *A1. Atmosphere model*

751 The atmospheric component of the model is a stripped-down version of the ECMWF single-  
752 column model which originates from an early cycle of the IFS (Integrated Forecasting Sys-  
753 tem) code. The model solves the primitive equations for temperature,  $T$ , specific humidity,  $q$ ,  
754 and zonal,  $u$ , and meridional,  $v$ , wind components, formulated in non-spherical co-ordinates,

(Simmons and Burridge, 1981; Ritchie et al., 1995) and using a hybrid vertical co-ordinate,  $\eta$  (see section A1.2.),

$$\frac{\partial u}{\partial t} + \dot{\eta} \frac{\partial u}{\partial \eta} = f(v - v_g) + F_u + P_u, \quad (\text{A1})$$

$$\frac{\partial v}{\partial t} + \dot{\eta} \frac{\partial v}{\partial \eta} = -f(u - u_g) + F_v + P_v, \quad (\text{A2})$$

$$\frac{\partial T}{\partial t} + \dot{\eta} \frac{\partial T}{\partial \eta} = \frac{R}{c_p} T \frac{\omega}{p} + F_T + P_T, \quad (\text{A3})$$

$$\frac{\partial q}{\partial t} + \dot{\eta} \frac{\partial q}{\partial \eta} = F_q + P_q. \quad (\text{A4})$$

Here  $t$  is time,  $f$  is the Coriolis parameter,  $u_g$  and  $v_g$  are prescribed geostrophic wind components,  $p$  is pressure,  $R$  is the gas constant for air,  $c_p$  is the specific heat at constant pressure for air,  $\dot{\eta}$  is the vertical velocity in  $\eta$  co-ordinates and  $\omega$  is the prescribed vertical velocity in pressure co-ordinates. The  $F_\phi$  ( $\phi = u, v, T, q$ ) are forcing terms representing the horizontal advection of the mean variables and the  $P_\phi$  terms represent tendencies due to the parameterisation of sub-grid scale physical processes.

The vertical advection terms in (A1)–(A4) are computed using a two time level Eulerian (upwind) scheme with a semi-implicit treatment of the right hand sides (ECMWF IFS documentation, 2001–2013a).

In the original ECMWF SCM code (the ‘full physics’ version), the  $P_\phi$  terms in (A1)–(A4) incorporate the effects of processes such as radiation, turbulent mixing, moist convection and clouds. For the simplified system, the code was stripped back to include just advection and turbulent mixing; the  $P_\phi$  terms then represent physical tendencies due to vertical exchange by turbulent processes only. This was done in order to simplify the derivation of the adjoint model whilst ensuring that the evolution of the atmosphere was sufficiently realistic for purposes of this study. The turbulent mixing is parameterised using a  $k$ -diffusion approach

777 (Louis, 1979)

$$778 \quad \frac{\partial \phi}{\partial t} = \frac{1}{\rho_a} \frac{\partial J_\phi}{\partial z}, \quad (\text{A5})$$

779 where  $\phi$  is the prognostic variable ( $T$ ,  $q$ ,  $u$ , or  $v$ ),  $\rho_a$  is the density of air,  $z$  is height, and the  
780 vertical turbulent flux  $J_\phi$  (positive downward) is given by

$$781 \quad J_\phi = \rho_a K_\phi \frac{\partial \phi}{\partial z}, \quad (\text{A6})$$

782 where  $K_\phi$  is the turbulent exchange coefficient.

783 The exchange coefficients between the surface and lowest model level ( $\sim 10$  m above sur-  
784 face) are expressed as functions of the bulk Richardson number, determined according to  
785 the formulation of Louis et al. (1982). Above the surface layer, the turbulent transports are  
786 based on local stability and the coefficients are defined using a combined Louis-Tiedtke-  
787 Geleyn (LTG) - Monin-Obukov (MO) formulation (Beljaars, 1995; Beljaars and Viterbo,  
788 1998; Viterbo et al., 1999). The physical tendencies are computed using an implicit time-  
789 stepping procedure. Full details of the turbulent diffusion scheme together with further refer-  
790 ences can be found in the ECMWF IFS documentation (2001–2013b).

### 791 *A1.1. Boundary conditions*

792 The upper and lower boundary conditions are given by

$$793 \quad K_\phi \frac{\partial \phi}{\partial z} = 0, \quad \text{at } p = p_{\text{top}} \quad (\text{A7})$$

$$794 \quad K_\phi \frac{\partial \phi}{\partial z} \rightarrow C_\phi |U(z)| (\phi(z) - \phi_{\text{surf}}) \quad \text{as } z \rightarrow 0. \quad (\text{A8})$$

795 where  $p_{\text{top}}$  is the pressure at the top of the atmosphere (set at 0.1 hPa),  $C_\phi$  is the transfer coef-  
796 ficient at the lowest model level, and  $\phi_{\text{surf}}$  represents the value of the prognostic variable  $\phi$  at  
797 the surface. The SST and saturation specific humidity are used as  $\phi_{\text{surf}}$  values for temperature  
798 and specific humidity, and a no-slip condition is used for the  $u$  and  $v$  wind components. The

799 surface turbulent fluxes are passed to the ocean model where they are used in the computation  
800 of the ocean surface boundary conditions (section A2.2.).

### 801 *A1.2. Vertical discretisation*

802 The atmosphere is divided into 60 unequally spaced layers, extending from the surface up  
803 to  $p_{\text{top}}$ , with the finest vertical resolution (measured in geometric height) in the planetary  
804 boundary layer. The model uses the hybrid vertical co-ordinate of Simmons and Burridge  
805 (1981). The co-ordinate  $\eta = \eta(p, p_s)$  is a monotonic function of the pressure, and is also  
806 dependent on the surface pressure,  $p_s$ . There is no staggering of prognostic model variables;  
807  $T, q, u$  and  $v$  are all represented at ‘full-level’ pressures  $p_k$  and the model layers are defined  
808 by the pressures at the interfaces between them (termed ‘half levels’).

### 809 *A2. Ocean mixed layer model*

810 The ocean mixed layer model is based on the KPP vertical mixing scheme of Large et al.  
811 (1994). The code was originally developed by the NCAS Centre for Global Atmospheric  
812 Modelling at the University of Reading (Woolnough et al., 2007) and incorporated into the  
813 ECMWF SCM code by Takaya et al. (2010) as part of a study into the impact of better  
814 representation of coupled atmosphere-upper ocean processes in the ECMWF medium-range  
815 forecasts. In this section we summarise the components of the scheme most relevant to this  
816 study; a comprehensive description of the model is given in Large et al. (1994).

817 The KPP model describes the evolution of the mean values of temperature,  $\theta$ , salinity,  $s$ ,  
818 and zonal and meridional currents  $u_o, v_o$ . The time evolution of each field is expressed as  
819 the vertical divergence of the kinematic turbulent fluxes,  $\overline{w'\phi'}$  ( $\phi = \theta, s, u_o, v_o$ ), giving the

820 following set of equations

$$821 \quad \frac{\partial \bar{\theta}}{\partial t} = -\frac{\partial \overline{w'\theta'}}{\partial z} - \frac{\partial Q_n}{\partial z}, \quad (\text{A9})$$

$$822 \quad \frac{\partial \bar{s}}{\partial t} = -\frac{\partial \overline{w's'}}{\partial z}, \quad (\text{A10})$$

$$823 \quad \frac{\partial \bar{u}_o}{\partial t} = -\frac{\partial \overline{w'u'_o}}{\partial z} + f v_o, \quad (\text{A11})$$

$$824 \quad \frac{\partial \bar{v}_o}{\partial t} = -\frac{\partial \overline{w'v'_o}}{\partial z} - f u_o. \quad (\text{A12})$$

825 Here, an overbar denotes a time average, primed variables represents turbulent fluctuations  
 826 from this average,  $w$  is the turbulent vertical velocity and  $Q_n$  is the non-turbulent heat flux  
 827 (solar irradiance) which is modelled using an empirical function of short wave radiation,  
 828  $Q_{SW}$ , and ocean depth,  $d$  (distance from ocean surface boundary).

829 The ocean surface boundary layer is defined as the region where  $d$  is less than or equal to  
 830 the ocean boundary layer depth  $h$ , the value of which is based on the depth at which the bulk  
 831 Richardson number equals the prescribed critical Richardson number. Within this region the  
 832 kinematic fluxes  $\overline{w'\phi'}$  are parameterised using K-profiles

$$833 \quad \overline{w'\phi'} = -K_\phi \frac{\partial \bar{\phi}}{\partial z}, \quad (\text{A13})$$

834 where  $\bar{\phi}$  represents a mean quantity. The  $K_\phi$  are expressed as product of a depth dependent  
 835 turbulent velocity scale and a smooth non-dimensional shape function such that they are  
 836 directly proportional to  $h$  at all depths.

837 In the ocean interior ( $d > h$ ) the turbulent vertical fluxes are parameterised as

$$838 \quad \overline{w'\phi'} = -\nu_\phi(d) \frac{\partial \bar{\phi}}{\partial z}, \quad (\text{A14})$$

839 where the interior diffusivity  $\nu_\phi$  is the sum of resolved shear instability and unresolved shear  
 840 instability due to internal wave breaking; we neglect the effect of double diffusion for reasons  
 841 described in Takaya et al. (2010).

### 842 *A2.1. Geostrophic currents*

843 For a 1D water column, the ocean currents are essentially governed by Ekman flow (Stew-  
 844 art, 2008). Without pressure gradient terms, the water moves under the sole influence of the  
 845 Coriolis force and the ocean momentum equations reduce to the equation for the harmonic  
 846 oscillator (Stewart, 2008); the solution takes the form of an inertial oscillation or inertial cur-  
 847 rent. In reality, we expect the ocean currents to be approximately geostrophically balanced.  
 848 To alleviate the unrealistic behaviour that this produces we use the method of Takaya et al.  
 849 (2010) and decompose the currents into slow and fast varying flows. The fast varying flow is  
 850 assumed to be mainly the Ekman (or ageostrophic) flow simulated by the KPP model. The  
 851 slow varying geostrophic component is prescribed and not modelled.

### 852 *A2.2. Boundary conditions*

853 The ocean surface boundary conditions are given by the surface kinematic fluxes of heat, salt  
 854 and momentum

$$855 \quad \overline{w'\theta'_0} = -Q_t/(\rho_0 c_{p0}), \quad (\text{A15})$$

$$856 \quad \overline{w's'_0} = -F_t s_0/\rho_0(0), \quad (\text{A16})$$

$$857 \quad \overline{w'u'_0} = -\tau_x/\rho_0, \quad (\text{A17})$$

$$858 \quad \overline{w'v'_0} = -\tau_y/\rho_0, \quad (\text{A18})$$

859 where  $Q_t$  is the net turbulent heat flux,  $F_t$  is the net turbulent freshwater flux,  $\tau_x$  and  $\tau_y$  are  
 860 the zonal and meridional components of the surface wind stress,  $s_0$ ,  $\rho_0$ ,  $c_{p0}$  are the salinity,  
 861 density and specific heat at constant pressure at the ocean surface, and  $\rho_0(0)$  is the density of

862 surface water with zero salinity (i.e. pure water). The fluxes  $Q_t$  and  $F_t$  are computed as

$$863 \quad Q_t = Q_{LW} + Q_E + Q_H, \quad (\text{A19})$$

$$864 \quad F_t = Q_E/L_\nu, \quad (\text{A20})$$

865 where  $Q_{LW}$  is net long wave radiation,  $Q_E$ ,  $Q_H$  are the latent and sensible heat fluxes and  
 866  $L_\nu$  is the latent heat of evaporation. The latent and sensible heat and momentum fluxes are  
 867 the surface turbulent fluxes from (A8); these are computed within the atmosphere model  
 868 component using the formulae given in section A3..

### 869 *A2.3. Vertical discretisation*

870 The ocean model uses a stretched vertical grid (Takaya et al., 2010) with 35 levels from  
 871 the surface to a depth of 250 m. The resolution is increased in the upper layers in order to  
 872 simulate the diurnal SST variability; the top model layer is chosen to be 1m thick and there  
 873 are 19 levels in the top 25 m. The largest depth is fixed so that the ocean model levels do not  
 874 vary with time. As with the atmosphere component, there is no staggering of the prognostic  
 875 model variables,  $\theta$ ,  $s$ ,  $u_o$  and  $v_o$  are all represented at full model level depths.

### 876 *A3. Atmosphere-Ocean coupling*

877 The atmosphere-ocean fluxes are estimated from bulk formulae

$$878 \quad \tau_x = \rho_a C_D |U_n| u_n, \quad (\text{A21})$$

$$879 \quad \tau_y = \rho_a C_D |U_n| v_n, \quad (\text{A22})$$

$$880 \quad Q_H = \rho_a C_H |U_n| (T_n - SST), \quad (\text{A23})$$

$$881 \quad Q_E = \rho_a L_\nu C_E |U_n| (q_n - q_{sat}(SST)), \quad (\text{A24})$$

882 where the subscript  $n$  represents the lowest atmosphere model level,

$$883 \quad |U_n| = \sqrt{u_n^2 + v_n^2}, \quad (\text{A25})$$

884 is the ( $\sim 10\text{m}$ ) windspeed and  $q_{sat}(SST)$  is the surface saturation specific humidity. The drag  
885 coefficient,  $C_D$ , and the transfer coefficients for heat,  $C_H$ , and moisture,  $C_E$ , are computed  
886 using the method of Louis et al. (1982).

## 887 **References**

- 888 M. Balmaseda. Initialization techniques in seasonal forecasts. In *Proceedings of the ECMWF*  
889 *Seminar on Seasonal prediction: science and applications*, pages 217–236, ECMWF, Read-  
890 ing, UK, 3–7 September 2012, 2012.
- 891 M. Balmaseda and D. Anderson. Impact of initialization strategies and observations on sea-  
892 sonal forecast skill. *Geophysical Research Letters*, 36:L01701, 2009.
- 893 R.N. Bannister. A review of forecast error covariance statistics in atmospheric variational data  
894 assimilation. I: Characteristics and measurements of forecast error covariances. *Quarterly*  
895 *Journal of Royal Meteorological Society*, 134:1955–1970, 2008a.
- 896 R.N. Bannister. A review of forecast error covariance statistics in atmospheric variational  
897 data assimilation. II: Modelling the forecast error covariance statistics. *Quarterly Journal*  
898 *of Royal Meteorological Society*, 134:1971–1996, 2008b.
- 899 A.C.M. Beljaars. The impact of some aspects of the boundary layer scheme in the ECMWF  
900 model. In *Proceedings of the ECMWF Seminar on Parametrization of Sub-grid Scale Phys-*  
901 *ical Processes*, pages 125–161, 1995.
- 902 A.C.M. Beljaars and P. Viterbo. The role of the boundary layer in a numerical weather  
903 prediction model. In A.A.M. Holtslag and P.G. Duynkerke, editors, *Clear and Cloudy*  
904 *Boundary Layers*, pages 287–304. Royal Netherlands Academy of Arts and Sciences, 1998.
- 905 G.B. Brassington, M.J. Martin, H.L. Tolman, S. Akella, M. Balmeseda, C.R.S. Chambers,  
906 J.A. Cummings, Y. Drillet, P.A.E.M. Jansen, P. Laloyaux, D. Lea, A. Mehra, I. Mirouze,

- 907 H. Ritchie, G. Samson, P.A. Sandery, G.C. Smith, Suarez M., and R. Todling. Progress and  
908 challenges in short- to medium-range coupled prediction. *Accepted for publication in the*  
909 *Journal of Operational Oceanography*, 2015.
- 910 P. Courtier. Dual formulation of four-dimensional variational assimilation. *Quarterly Journal*  
911 *of the Royal Meteorological Society*, 123:2449–2461, 1997.
- 912 P. Courtier, J-N. Thépaut, and A. Hollingsworth. A strategy for operational implementation  
913 of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological*  
914 *Society*, 120:1367–1387, 1994.
- 915 R. Daley. *Atmospheric Data Analysis*. Cambridge University Press, 1991.
- 916 D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A.  
917 Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot,  
918 N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy,  
919 H. Hersbach, E. V. Hólm, L. Isaksen, P. Køollberg, M. Köhler, M. Matricardi, A. P. McNally,  
920 B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N.  
921 Thépaut, and F. Vitart. The ERA-Interim reanalysis: configuration and performance of the  
922 data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):  
923 553–597, 2011.
- 924 ECMWF IFS documentation. *Part III: Dynamics and numerical procedures*. Eu-  
925 ropean Centre for Medium-Range Weather Forecasts, 2001–2013a. Available at  
926 <http://old.ecmwf.int/research/ifsdocs/>.
- 927 ECMWF IFS documentation. *Part IV: Physical Processes*. European Centre for Medium-  
928 Range Weather Forecasts, 2001–2013b. Available at <http://old.ecmwf.int/research/ifsdocs/>.
- 929 A. Fowler, R. Bannister, and J. Eyre. A new floating model level scheme for the assimilation  
930 of boundary layer top inversions: The univariate assimilation of temperature. *Quarterly*  
931 *Journal of the Royal Meteorological Society*, 138:682–698, 2012.
- 932 P. Gauthier, C. Charette, L. Fillion, P. Koclas, and S. Laroche. Implementation of a 3D  
933 variational data assimilation system at the Canadian Meteorological Centre. Part 1: The

- 934 global analysis. *Atmosphere-Ocean*, 37:103–156, 1999.
- 935 P. Gauthier, M. Tanguay, S. Laroche, S. Pellerin, and J. Morneau. Extension of 3DVAR to  
936 4DVAR: Implementation of 4DVAR at the Meteorological Service of Canada. *Monthly*  
937 *Weather Review*, 135:2339–2354, 2007.
- 938 X-Y. Huang, Q. Xiao, D.M. Barker, X. Zhang, J. Michalakes, W. Huang, T. Henderson,  
939 J. Bray, Y. Chen, Z. Ma, J. Dudhia, Y. Guo, X. Zhang, D-J. Won, H-C. Lin, and Y-H. Kuo.  
940 Four-dimensional variational data assimilation for WRF: Formulation and preliminary re-  
941 sults. *Monthly Weather Review*, 137:299–314, 2009.
- 942 M. Janisková, J-N. Thépaut, and J-F. Geleyn. Simplified and regular physical parameteriza-  
943 tions for incremental four-dimensional variational assimilation. *Monthly Weather Review*,  
944 127:26–45, 1999.
- 945 N.S. Keenlyside, M. Latif, J. Jungclaus, L. Kornblueh, and E. Roeckner. Advancing decadal-  
946 scale climate prediction in the North Atlantic sector. *Nature*, 453:84–88, 2008.
- 947 P. Laloyaux, M. Balmaseda, K. Mogensen, P. Janssen, and D. Dee. The ECMWF coupled  
948 assimilation system. In P. Soille and Marchetti P.G., editors, *Proceedings of the 2014 confer-*  
949 *ence on Big Data from Space (BiDS'14)*, pages 16–19, ESA-ESRIN, Frascati, Italy, 12-14  
950 November 2014, 2014. URL <http://dx.doi.org/10.2788/1823>.
- 951 P. Laloyaux, M. Balmaseda, D. Dee, K. Mogensen, and P. Janssen. The ECMWF prototype  
952 for a coupled assimilation system. *Submitted to Quarterly Journal of the Royal Meteor-*  
953 *ological Society*, 2015.
- 954 W.G. Large, J.C. McWilliams, and S.C. Doney. Oceanic vertical mixing: a review and a  
955 model with non-local boundary layer parameterization. *Reviews of Geophysics*, 32:363–  
956 403, 1994.
- 957 S. Laroche, M. Tanguay, and Y. Delage. Linearization of a simplified planetary boundary  
958 layer parameterization. *Monthly Weather Review*, 130:2074–2087, 2002.
- 959 A.S. Lawless. International workshop on coupled data assimilation, University of Reading,  
960 UK. Final report, September 2012. Available at <http://www.esa-da.org/content/d1-report->

- 961 international-workshop-coupled-data-assimilation.
- 962 A.S. Lawless. Variational data assimilation for very large environmental problems. In M.J.P.  
963 Cullen, M. A. Freitag, S. Kindermann, and R. Scheichl, editors, *Large Scale Inverse Prob-*  
964 *lems: Computational Methods and Applications in the Earth Sciences*, Radon Series on  
965 Computational and Applied Mathematics 13., pages 55–90. De Gruyter, 2013.
- 966 A.S. Lawless and N.K. Nichols. Inner loop stopping criteria for incremental four-dimensional  
967 variational data assimilation. *Monthly Weather Review*, 134:3425–3435, 2006.
- 968 A.S. Lawless, S. Gratton, and N.K. Nichols. An investigation of incremental 4D-Var using  
969 non-tangent linear models. *Quarterly Journal of the Royal Meteorological Society*, 131:  
970 459–476, 2005.
- 971 D.J. Lea, I. Mirouze, M.J. Martin, R.R. King, A. Hines, and D. Walters. Using a new cou-  
972 pled data assimilation system to assess the HadGEM3 coupled model. *In preparation for*  
973 *submission to Monthly Weather Review*, 2015.
- 974 J.-M. Lellouche, O. Le Galloudec, M. Drévillon, C. Régnier, E. Greiner, G. Garric, N. Ferry,  
975 C. Desportes, C.-E. Testut, C. Bricaud, R. Bourdallé-Badie, B. Tranchant, M. Benkiran,  
976 Y. Drillet, A. Daudin, and C. De Nicola. Evaluation of global monitoring and forecasting  
977 systems at Mercator Océan. *Ocean Science*, 9:57–81, 2013.
- 978 J.F. Louis. A parametric model of vertical eddy fluxes in the atmosphere. *Boundary Layer*  
979 *Meteorology*, 17:187–202, 1979.
- 980 J.F. Louis, M. Tiedtke, and J.F. Geleyn. A short history of the operational PBL parameteriza-  
981 tion at ECMWF. In *ECMWF Workshop on Boundary Layer Parametrization*, pages 59–80,  
982 Reading, 25-27 November 1982.
- 983 J-F. Mahfouf. Influence of physical processes on the tangent-linear approximation. *Tellus*,  
984 51A:147–166, 1999.
- 985 G.A. Meehl, L. Goddard, G. Boer, R. Burgman, G. Branstator, C. Cassou, S. Corti, G. Dan-  
986 abasoglu, F. Doblas-Reyes, E. Hawkins, A. Karspeck, M. Kimoto, A. Kumar, D. Matei,  
987 J. Mignot, R. Msadek, A. Navarra, H. Pohlmann, M. Rienecker, T. Rosati, E. Schneider,

- 988 D. Smith, R. Sutton, H. Teng, G.J. van Oldenborgh, G. Gabriel Vecchi, and S. Stephen Yea-  
989 ger. Decadal climate prediction: an update from the trenches. *Bulletin of the American*  
990 *Meteorological Society*, 95:243–267, 2014.
- 991 J. Murphy, V. Kattsov, N. Keenlyside, M. Kimoto, G. Meehl, H. Mehta, A. Pohlmann,  
992 A. Scaife, and D. Smith. Towards prediction of decadal climate variability and change.  
993 *Procedia Environmental Sciences 1*, pages 287–304, 2010.
- 994 I.M. Navon, X. Zou, J. Derber, and J. Sela. Variational data assimilation with an adiabatic  
995 version of the NMC spectral model. *Monthly Weather Review*, 120:1433–1466, 1992.
- 996 D.W. Pierce, T.P. Barnett, R. Tokmakian, A. Semtner, M. Maltrud, J. Lysne, and A. Craig. The  
997 ACPI project, element 1: Initializing a coupled climate model from observed conditions.  
998 *Climatic Change*, 62:13–28, 2004.
- 999 H. Pohlmann, J.H. Jungclaus, A. Köhl, D. Stammer, and J. Marotzke. Initializing decadal cli-  
1000 mate predictions with the GECCO oceanic synthesis: Effects on the North Atlantic. *Journal*  
1001 *of Climate*, 22:3926–3938, 2009.
- 1002 F. Rabier, H. Jarvinen, E. Klinker, J.F. Mahfouf, and A. Simmons. The ECMWF operational  
1003 implementation of four-dimensional variational assimilation. I: Experimental results with  
1004 simplified physics. *Quarterly Journal of the Royal Meteorological Society*, 126:1143–1170,  
1005 2000.
- 1006 F. Rawlins, S.P. Ballard, K.J. Bovis, A.M. Clayton, D. Li, G.W. Inverarity, A.C. Lorenc, and  
1007 T.J. Payne. The Met Office global four-dimensional variational data assimilation scheme.  
1008 *Quarterly Journal of the Royal Meteorological Society*, 133:347–362, 2007.
- 1009 H. Ritchie, C. Temperton, A. Simmons, M. Hortal, T. Davies, D. Dent, and M. Hamrud. Im-  
1010 plementation of the Semi-Lagrangian method in a high-resolution version of the ECMWF  
1011 forecast model. *Monthly Weather Review*, 123:489–514, 1995.
- 1012 D.F. Shanno. On the convergence of a new conjugate gradient algorithm. *SIAM Journal on*  
1013 *Numerical Analysis*, 15:1247–1257, 1978.
- 1014 D.F. Shanno and K.H. Phua. Remark on algorithm 500 - a variable method subroutine for

- 1015 unconstrained nonlinear minimization. *ACM Transactions on Mathematical Software*, 6:  
1016 618–622, 1980.
- 1017 A.J. Simmons and D.M. Burridge. An energy and angular-momentum conserving vertical  
1018 finite-difference scheme and hybrid vertical coordinates. *Monthly Weather Review*, 109:  
1019 758–766, 1981.
- 1020 R.H. Stewart. *Introduction to Physical Oceanography*. Department of Oceanography, Texas  
1021 A&M University, 2008. Available from <http://oceanworld.tamu.edu/index.html>.
- 1022 N. Sugiura, T. Awaji, S. Masuda, T. Mochizuki, T. Toyoda, T. Miyama, H. Igarashi, and  
1023 Y. Ishikawa. Development of a four-dimensional variational coupled data assimilation sys-  
1024 tem for enhanced analysis and prediction of seasonal to interannual climate variations. *Jour-  
1025 nal of Geophysical Research*, 113, 2008.
- 1026 Y. Takaya, F. Vitart, G. Balsamo, M. Balmaseda, M. Leutbecher, and F. Molteni. Implemen-  
1027 tation of an ocean mixed layer model in IFS. Technical Memorandum No. 622, European  
1028 Centre for Medium-Range Weather Forecasts, Reading, U.K, 2010.
- 1029 R. Tardif, G.J. Hakim, and C. Snyder. Coupled atmosphere-ocean data assimilation experi-  
1030 ments with a low-order climate model. *Climate Dynamics*, 43:1631–1643, 2014.
- 1031 J-N. Thépaut, R.N. Hoffman, and P. Courtier. Interactions of dynamics and observations in a  
1032 four-dimensional variational assimilation. *Monthly Weather Review*, 121:3393–3414, 1993.
- 1033 J-N. Thépaut, P. Courtier, G. Belaud, and G. Lemaître. Dynamical structure functions in a  
1034 four-dimensional variational assimilation: A case study. *Quarterly Journal of the Royal  
1035 Meteorological Society*, 122:535–561, 1996.
- 1036 P. Viterbo, A. Beljaars, J-F. Mahfouf, and J. Teixeira. The representation of soil moisture  
1037 freezing and its impact on the stable boundary layer. *Quarterly Journal of the Royal Mete-  
1038 orological Society*, 125:2401–2426, 1999.
- 1039 A.T. Weaver, J. Vialard, and D.L.T. Anderson. Three- and four-dimensional variational assim-  
1040 ilation with a general circulation model of the tropical pacific ocean. Part I: Formulation,  
1041 internal diagnostics, and consistency checks. *Monthly Weather Review*, 131:1360–1378,

1042 2003.

1043 S. Woolnough, F. Vitart, and M.A. Balmaseda. The role of the ocean in the Madden-Julian  
1044 Oscillation: implications for MJO prediction. *Quarterly Journal of the Royal Meteorologi-  
1045 cal Society*, 133:117–128, 2007.

1046 S. Zhang, M.J. Harrison, A. Rosati, and A. Wittenberg. System design and evaluation of  
1047 coupled ensemble data assimilation for global oceanic climate studies. *Monthly Weather  
1048 Review*, 135:3541–3564, 2007.

*Table 1.* observation error standard deviations by field

atmosphere temperature (K)	$u$ wind ( $\text{ms}^{-1}$ )	$v$ wind ( $\text{ms}^{-1}$ )	ocean temperature (K)	salinity (psu)	$u$ current ( $\text{ms}^{-1}$ )	$v$ current ( $\text{ms}^{-1}$ )
1.0	1.5	1.5	0.01	0.003	0.01	0.01

*Table 2.* atmosphere observation locations

model level	standard pressure level (hPa)	model full level pressure value (hPa) <sup>3</sup>
14	10	9.893
17	20	18.815
19	30	28.882
22	50	54.624
23	70	66.623
25	100	95.980
28	150	154.038
30	200	202.230
32	250	257.685
33	300	288.093
36	400	389.233
39	500	501.637
44	700	694.696
49	850	861.497
52	925	935.065
56	1000	995.055
60	n/a	1017.293

<sup>3</sup> values based on a surface pressure value of 1018.5 hPa; model full level pressure values vary with surface pressure. These levels have been chosen to approximately correspond to the standard pressure levels (hPa).

*Table 3.* ocean observation locations

model level	depth (m)
1	1.000
3	3.069
5	5.277
8	8.848
10	11.406
13	15.538
16	20.173
18	23.762
20	28.100
22	33.760
23	37.366
24	41.703
25	46.985
26	53.475
27	61.498
28	71.452
29	83.818
30	99.175
31	118.214
32	141.758
33	170.778
34	206.414
35	250.000

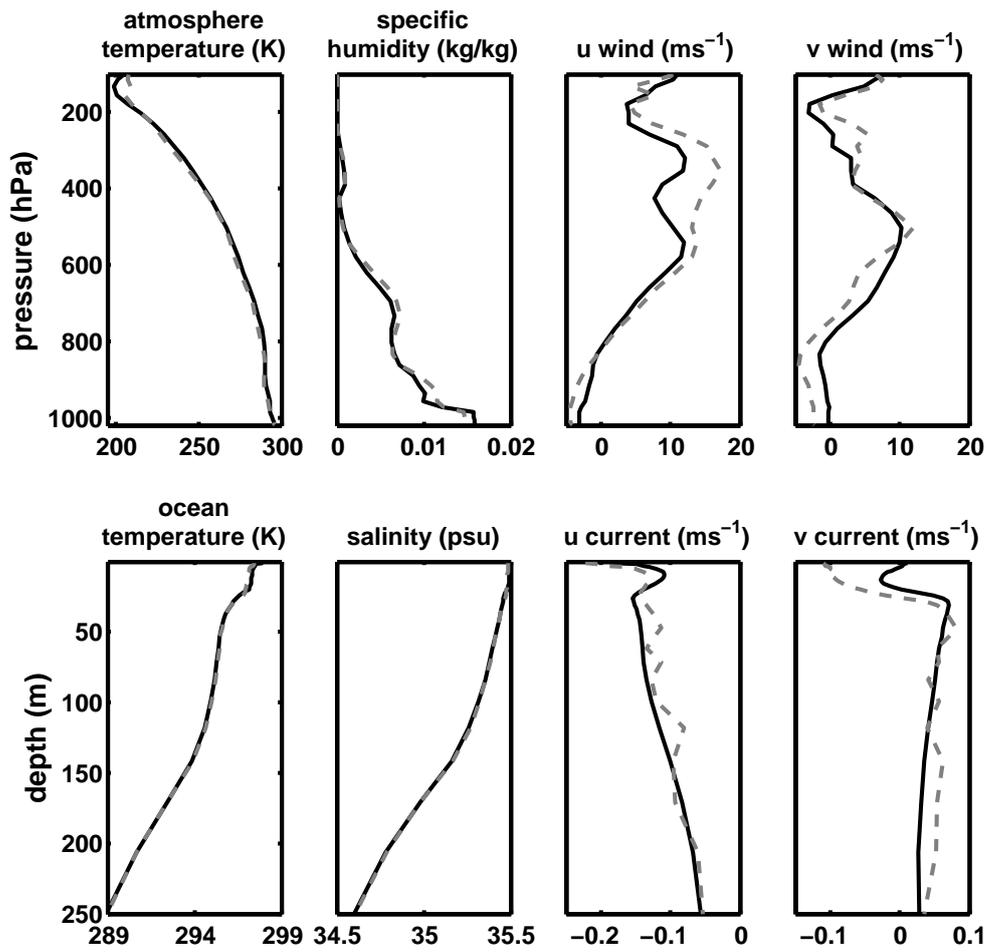


Fig. 1. True (solid black line) and background (dashed grey line) initial states.

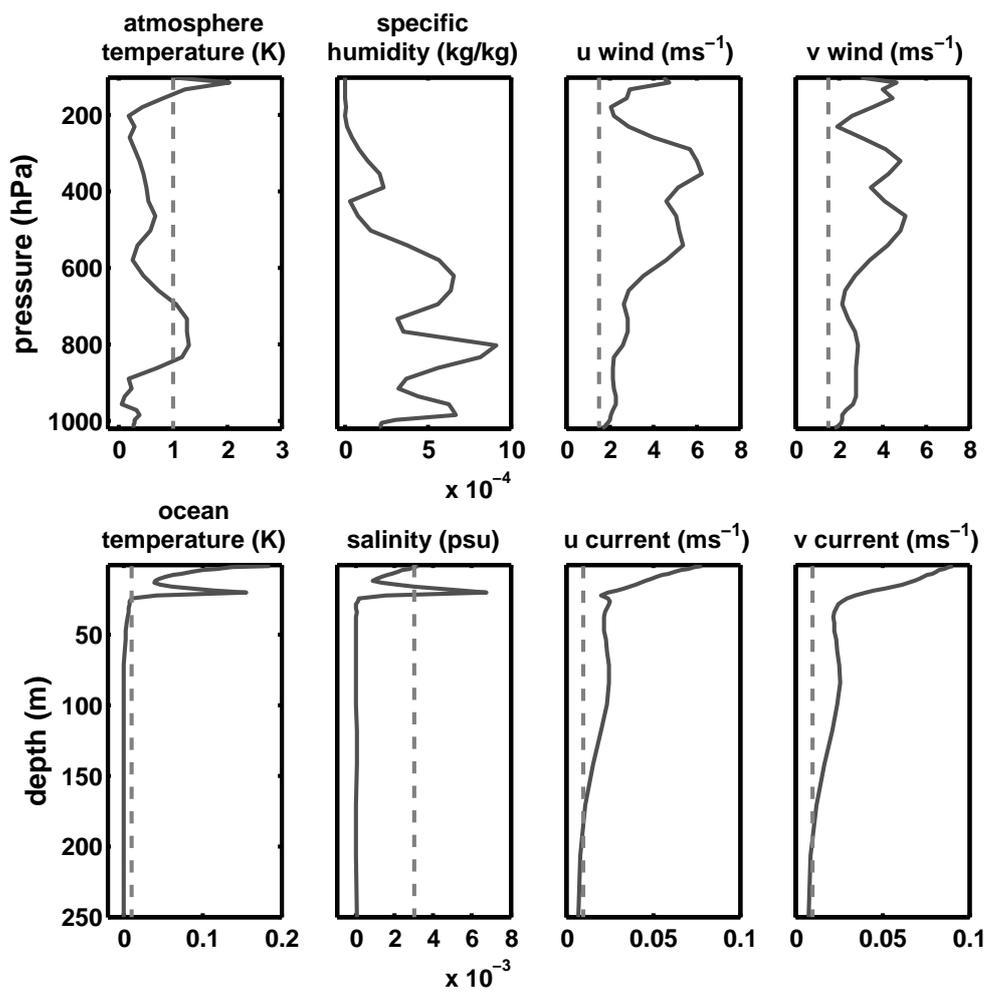
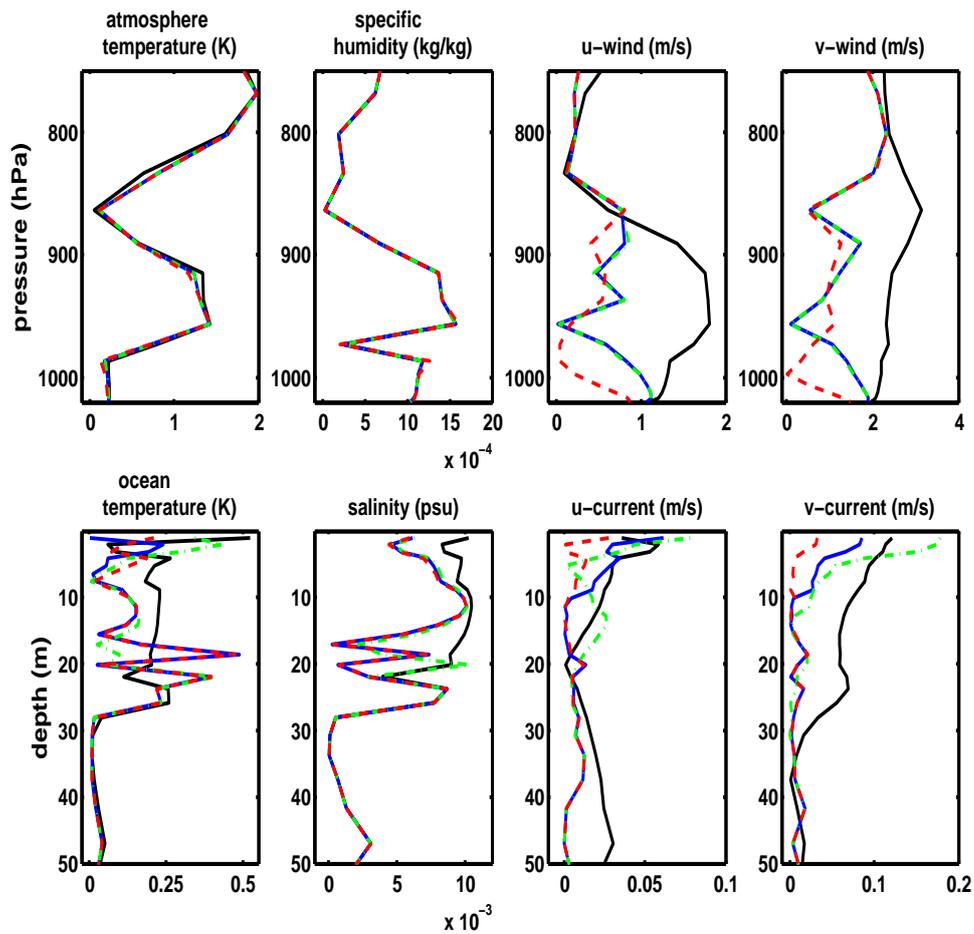
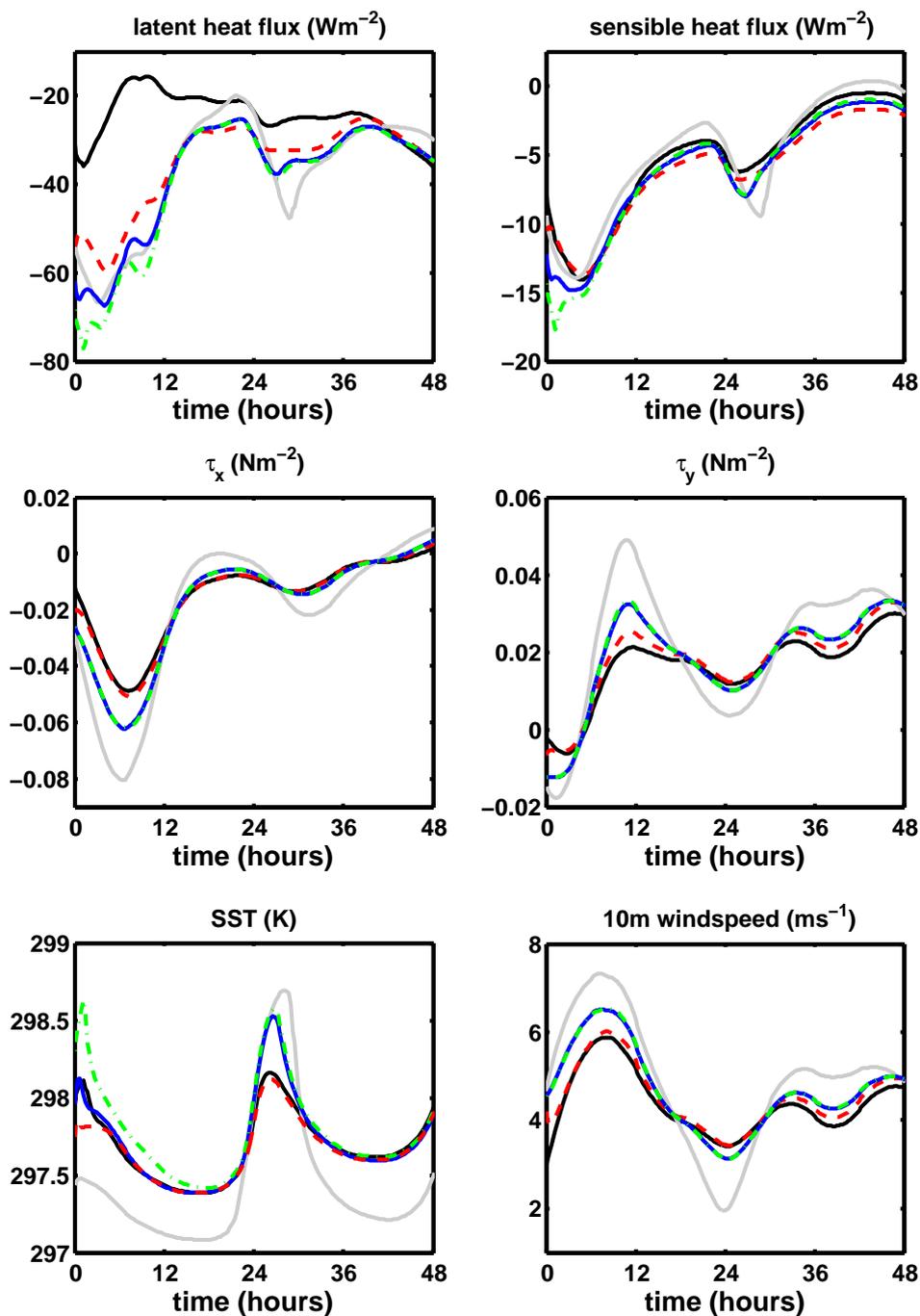


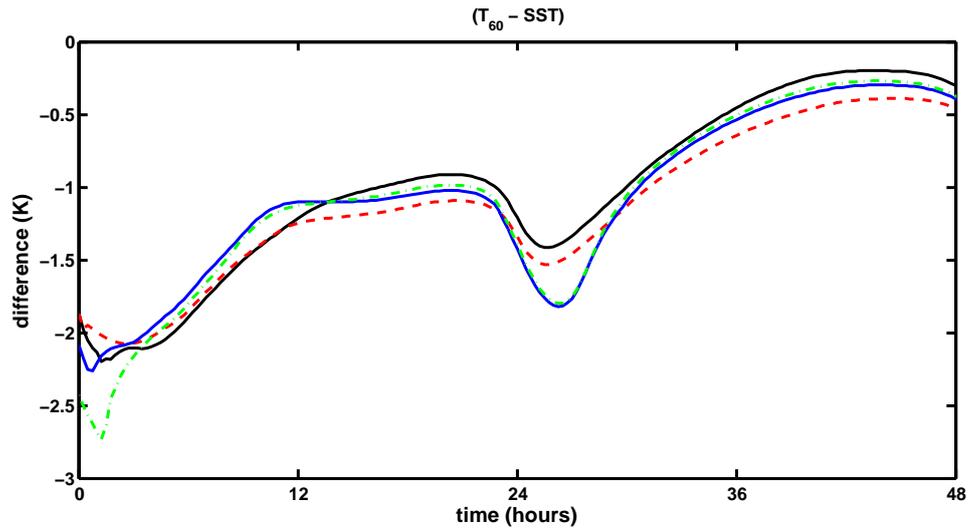
Fig. 2. Initial background (solid grey line) and observation (vertical dashed grey line) error standard deviations.



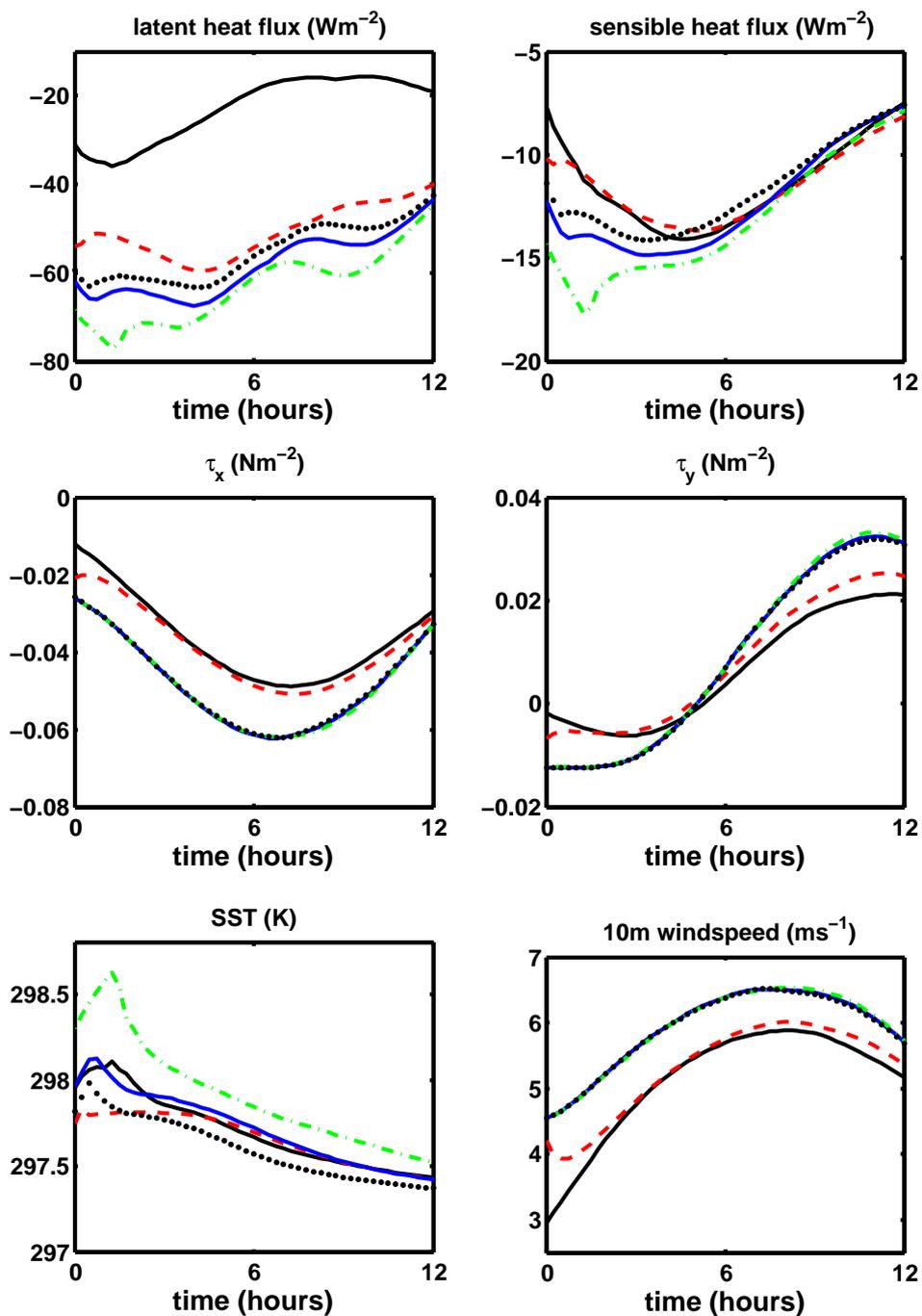
*Fig. 3.* Absolute errors at initial time: background error (*solid black line*), strongly coupled analysis error (*dashed red line*), weakly coupled analysis error (*solid blue line*), and uncoupled analysis error (*dot-dash green line*). Experiments using a 12 hour assimilation window with 3 hourly atmosphere & 6 hourly ocean observations. Observation locations are given in tables 2 and 3.



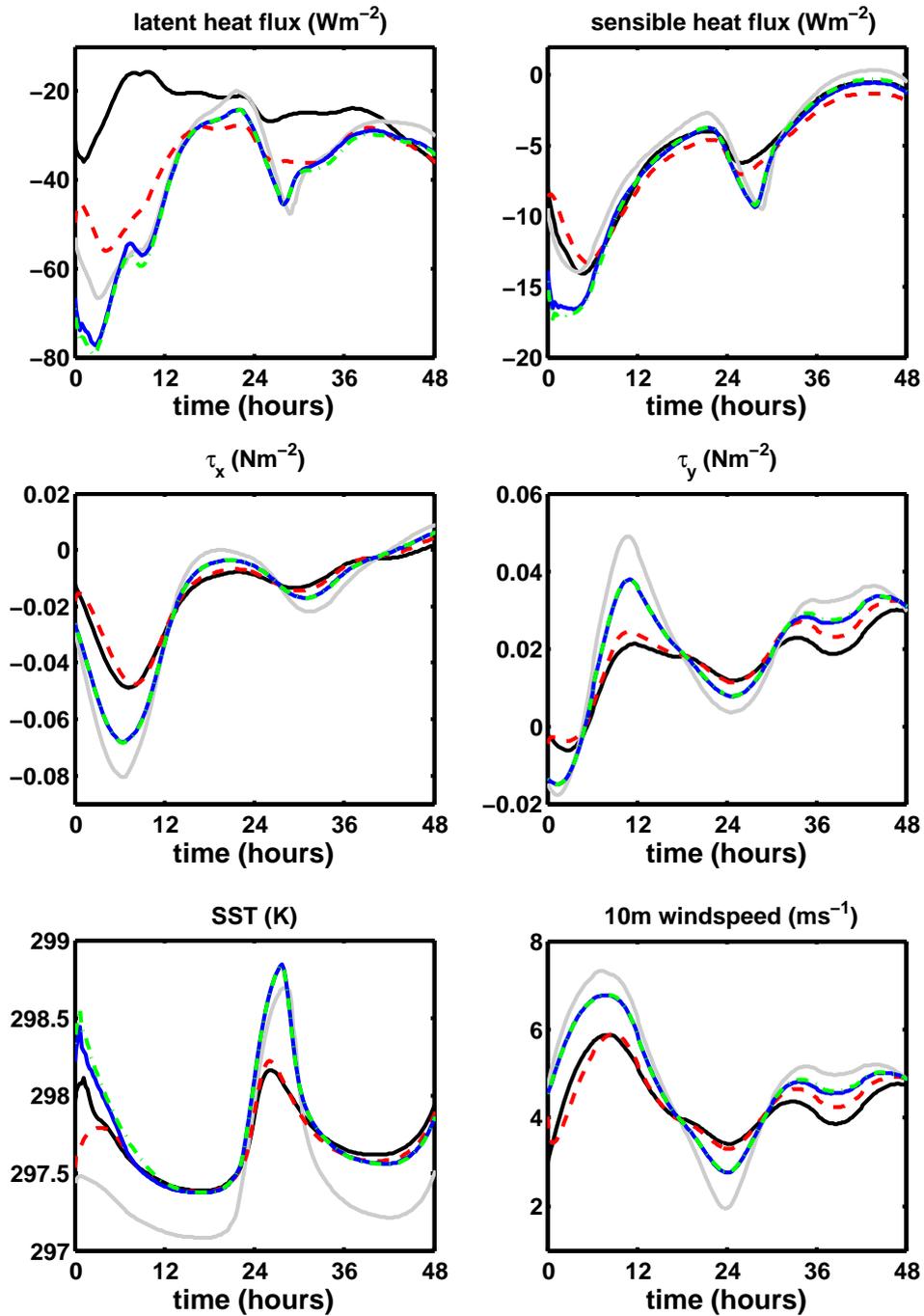
*Fig. 4.* Coupled model SST & surface fluxes for coupled model forecast initialised from  $t_0$  analyses, first 48 hours of forecast: truth (*solid black line*), forecast initialised from initial background state (*solid grey line*), forecast initialised from strongly coupled analysis (*dashed red line*), forecast initialised from weakly coupled analysis (*solid blue line*), and forecast initialised from uncoupled analyses (*dot-dash green line*). Experiments using a 12 hour assimilation window with 3 hourly atmosphere & 6 hourly ocean observations.



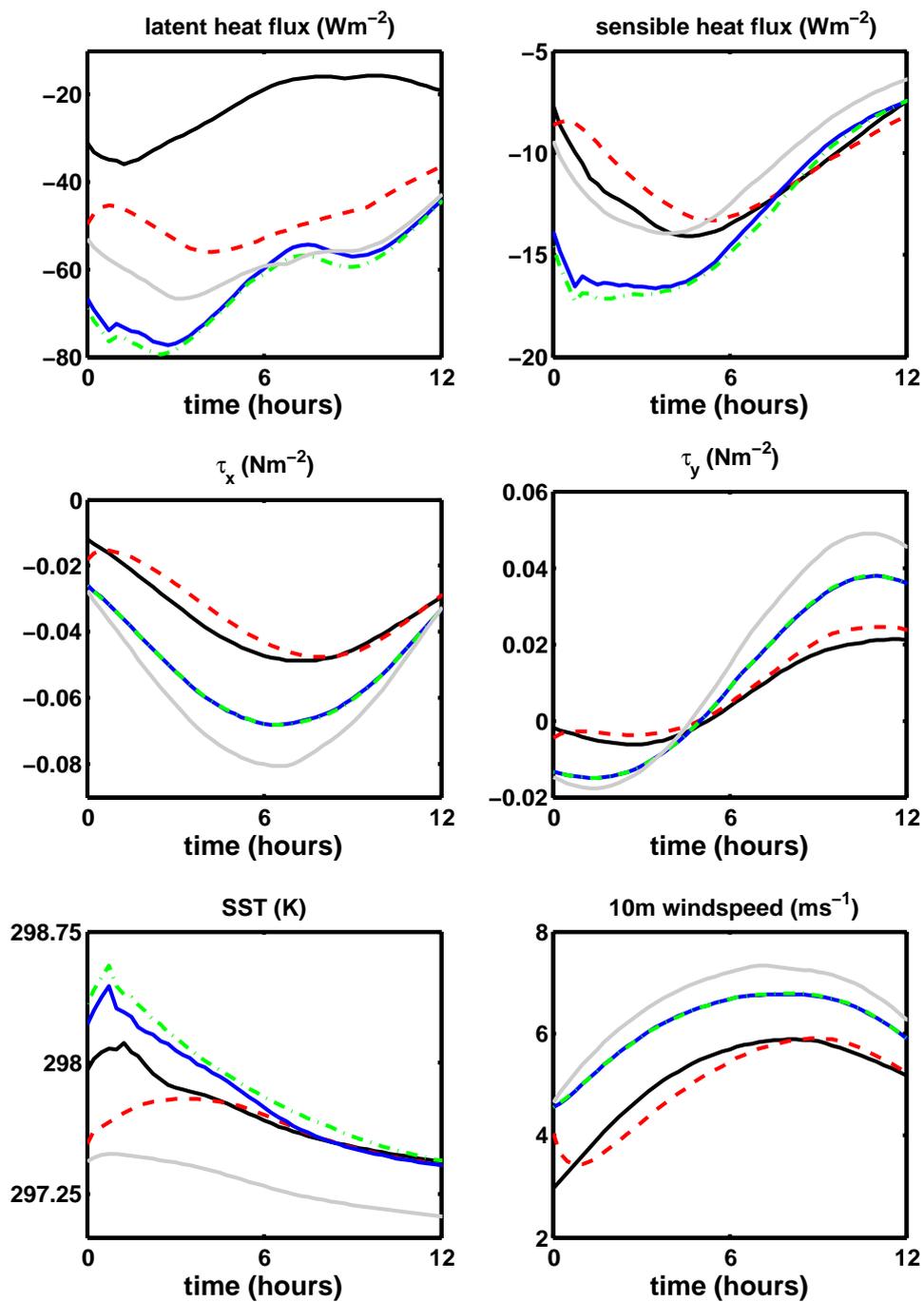
*Fig. 5.* Atmosphere-Ocean temperature difference ( $T_{60} - SST$ ) for coupled model forecast initialised from  $t_0$  analyses: truth (*solid black line*), forecast initialised from strongly coupled analysis (*dashed red line*), forecast initialised from weakly coupled analysis (*solid blue line*), and forecast initialised from uncoupled analyses (*dot-dash green line*). Experiments using 12 hour assimilation window with 3 hourly atmosphere & 6 hourly ocean observations



*Fig. 6.* Coupled model SST & surface fluxes for coupled model forecast initialised from  $t_0$  analyses, first 12 hours of forecast: truth (*solid black line*), forecast initialised from strongly coupled analysis (*dashed red line*), forecast initialised from weakly coupled analysis (*solid blue line*), forecast initialised from uncoupled analyses (*dot-dash green line*), and forecast initialised from uncoupled analyses using ‘true’ forcing (*black dots*). Experiments using 12 hour assimilation window with 3 hourly atmosphere & 6 hourly ocean observations.



*Fig. 7.* Coupled model SST & surface fluxes for coupled model forecast initialised from  $t_0$  analyses, first 48 hours of forecast: truth (*solid black line*), forecast initialised from initial background state (*solid grey line*), forecast initialised from strongly coupled analysis (*dashed red line*), forecast initialised from weakly coupled analysis (*solid blue line*), and forecast initialised from uncoupled analyses (*dot-dash green line*). Experiments using 12 hour assimilation window with 6 hourly atmosphere & 6 hourly ocean observations.



*Fig. 8.* Coupled model SST & surface fluxes for coupled model forecast initialised from  $t_0$  analyses, first 12 hours of forecast: truth (*solid black line*), forecast initialised from initial background state (*solid grey line*), forecast initialised from strongly coupled analysis (*dashed red line*), forecast initialised from weakly coupled analysis (*solid blue line*), and forecast initialised from uncoupled analyses (*dot-dash green line*). Experiments using 12 hour assimilation window with 6 hourly atmosphere & 6 hourly ocean observations.

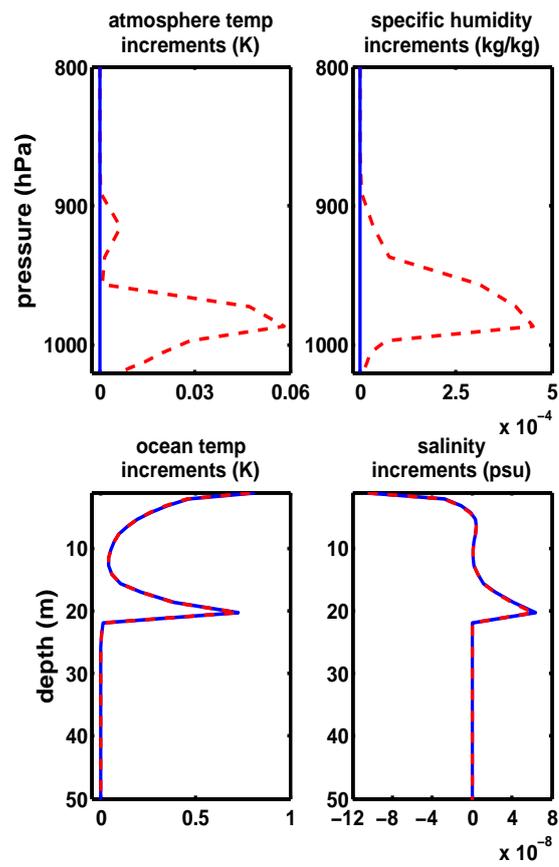


Fig. 9. Analysis increments at  $t = 0$ : strongly coupled (*dashed red line*) and weakly coupled (*solid blue line*) assimilations with single SST observation at end of 12 hour assimilation window.

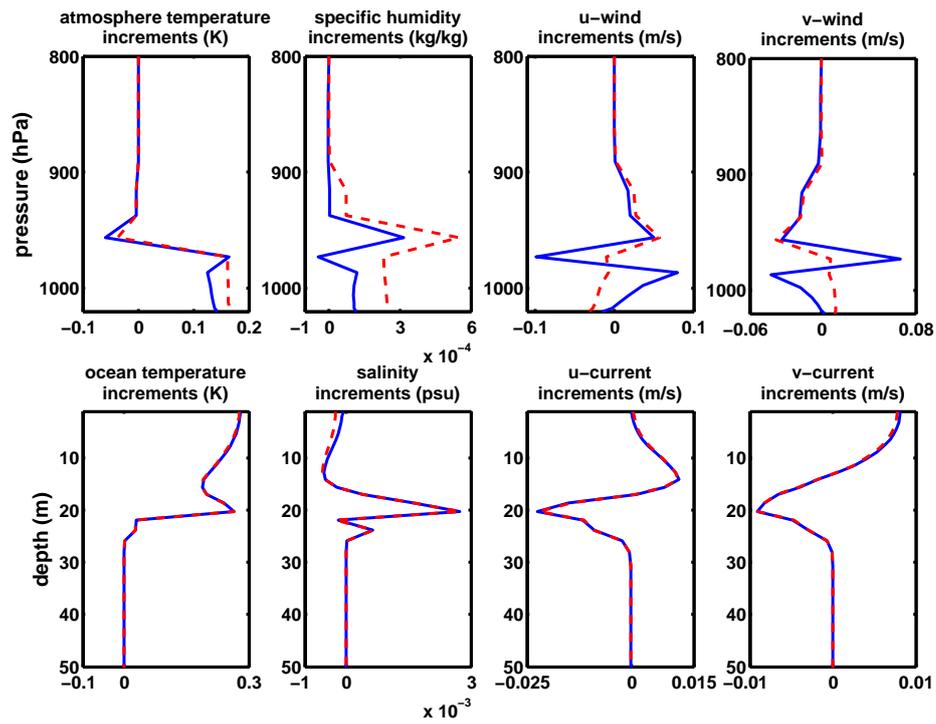


Fig. 10. Analysis increments at  $t = 6$  hr: strongly coupled (dashed red line) and weakly coupled (solid blue line) assimilations with single SST observation at end of 12 hour assimilation window.

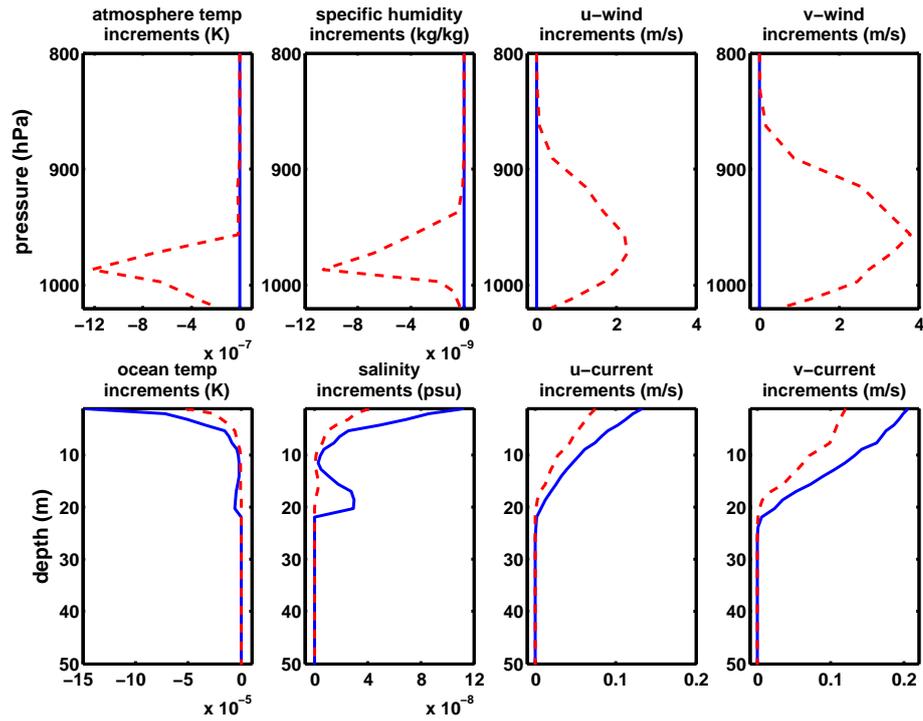


Fig. 11. Analysis increments at  $t = 0$ : strongly coupled (*dashed red line*) and weakly coupled (*solid blue line*) assimilations with single ocean surface current observation at end of 12 hour assimilation window.

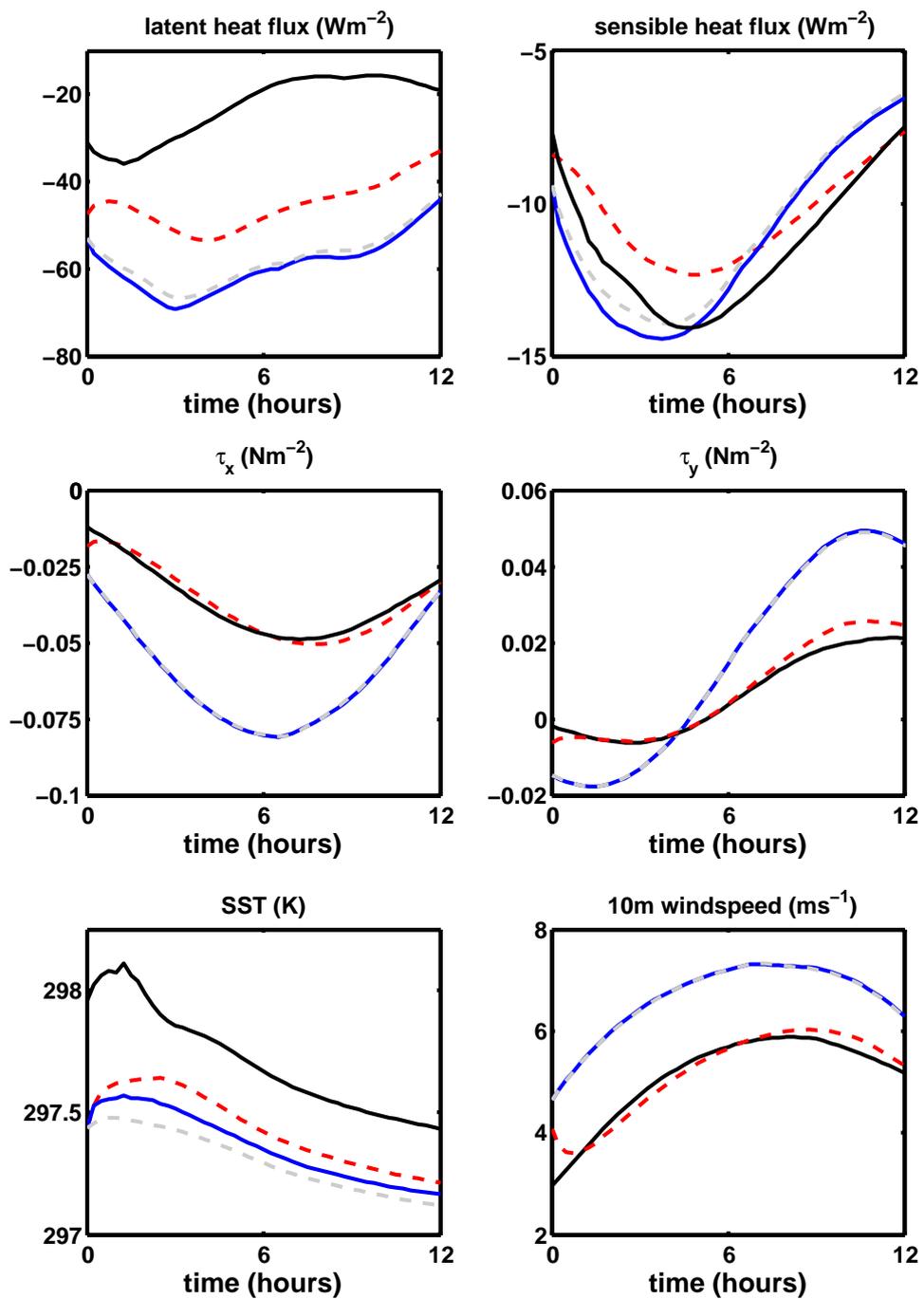
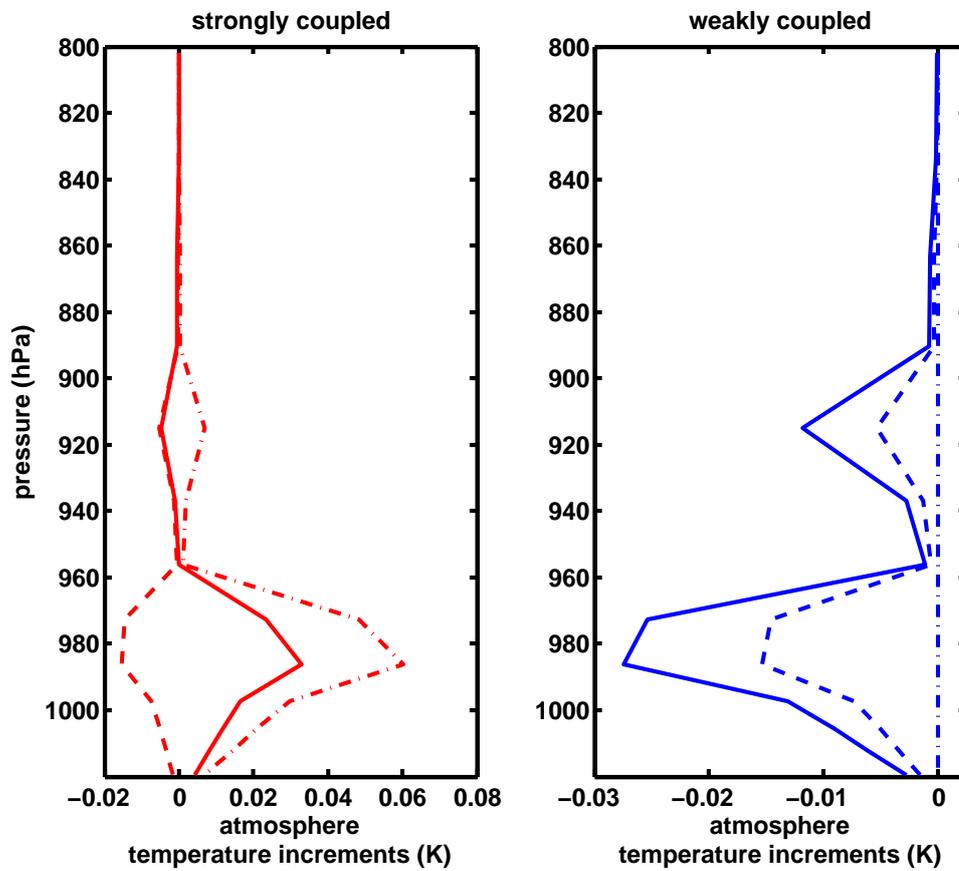


Fig. 12. SST & surface fluxes: truth (solid black line), initial background trajectory (dashed grey line), strongly coupled analysis (dashed red line) and weakly coupled analysis (solid blue line) from experiment with single ocean surface current observation at end of 12 hour assimilation window.



*Fig. 13.* Atmospheric temperature analysis increments at  $t = 0$ : strongly coupled (*left*) and weakly coupled (*right*) assimilations with two observations of temperature, corresponding to the lowest model level in the atmosphere and the top model level in the ocean, at the end of a 12 hour assimilation window. The solid lines indicate the analysis increment when both temperature observations are assimilated; the dashed lines correspond to assimilation of the atmosphere temperature observation only; and dot-dash lines correspond to assimilation of the ocean temperature observation only.