HAL
open science

# Report on fully coupled data assimilation in simplified systems with implications for Earth system reanalysis

Arthur Vidard, Rémi Pellerej, Florian Lemarié

## To cite this version:

Arthur Vidard, Rémi Pellerej, Florian Lemarié. Report on fully coupled data assimilation in simplified systems with implications for Earth system reanalysis. [Research Report] Inria Grenoble Rhône-Alpes. 2017. hal-01667509

**HAL Id: hal-01667509**

**https://inria.hal.science/hal-01667509**

Submitted on 19 Dec 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Work Package 2**: Future coupling methods

# Deliverable 2.11:

## Report on fully coupled data assimilation in simplified systems with implications for Earth system reanalysis.

**Type**: Report

**Authors**: Arthur Vidard, Rémi Pellerej and Florian Lemarié

**Reviewer(s)**: Matthew Martin

**Delivered**: 15/11/2017

## Summary

This report describes the Inria contribution to WP2 regarding a study of coupled data assimilation algorithms applied to academic models. This task was twofold: 1. Propose and study new coupled variational data assimilation algorithms. 2. Create a stand-alone coupled single column model (SCM) that mimics the ocean atmosphere behaviour and that can be used to validate the algorithms proposed in 1.

Regarding subtask 1, a focus was made on ways to explicitly account for model coupling in the variational optimisation problem, either as a strong contraint or as a weak constraint, or as a combination of both. It is then applied to both linear and non linear coupled problems, and leads to the conclusion that it does bring some noticeable benefit, but at a cost, both in time of development and computing time. The cost-benefit ratio has therefore to be studied for each given application.

The coupled SCM has been developed in Fortran and interfaced with the OOPS data assimilation framework. It is documented in this report and a reference test case is given, so that it can be reused by partners of the project and beyond.

## 1 Introduction

In the context of operational meteorology and oceanography, forecast skills heavily rely on proper combination of model prediction and available observations via data assimilation techniques. Historically, numerical weather prediction is made separately for the ocean and the atmosphere in an uncoupled way. However, in recent years, fully coupled ocean-atmosphere models are increasingly used in operational centres to improve the reliability of seasonal forecasts and tropical cyclones predictions and to improve reanalyses. For coupled problems, the use of separated data assimilation schemes in each medium is not satisfactory since the result of such assimilation process is generally inconsistent across the interface, thus leading to unacceptable artefacts (Mulholland et al. 2015). Hence, there is a strong need for adapting existing data assimilation techniques to the coupled framework, as presented in Smith et al. (2015). In that respect, ERACLIM2 is an important milestone, with the implementation and major application of the CERA algorithm. Task 2.11 aims at using coupled data assimilation as an opportunity to improve the coupling mathematical consistency of the forecast coupled system. In this report, three classes of data assimilation algorithms, based on variational data assimilation techniques (Le Dimet & Talagrand 1986), are presented and applied to single column coupled problems. Reference of fully coupled solutions are obtained through an iterative Schwarz domain decomposition method (Gander 2008). The aim of the proposed methods is to properly take into account the coupling in the assimilation process in order to obtain a coupled solution close to the observations while satisfying the physical conditions across the air-sea interface. The paper is organised as follows. The model problem and coupling strategy are given in Sec. 2. In Sec. 3 we briefly recall some theoretical aspects of variational data assimilation techniques, and we introduce and discuss three algorithms to solve coupled constrained minimisation problems. Comparaison with the CERA system are also presented. The performance of the proposed schemes are illustrated by numerical experiments in Sec. 5 in a linear case and in Sec. 6 in a non-linear single column ocean-atmosphere model. We decided to move technical description of our single column model and discussion about convergence of the data assimilation schemes in appendices, but they are an important part of the report nonetheless.

## 2 Model problem and coupling strategy

Numerical difficulties induced by air-sea coupling mostly come from vertical interactions, so we restrict our study on single column models; however most of it can be extended to 3D systems without major theoretical difficulties. We consider a problem defined on a bounded set $\Omega \subset \mathbb{R}$. $\Omega$ is decomposed into two non-overlapping subdomains $\Omega_a$ and $\Omega_o$ with an interface $\Gamma = \{z = 0\}$. A model is defined on each space-time domain $\Omega_\beta \times [0, T]$ ($\beta = a$ or $o$) thanks to a differential operator $\mathscr{L}_\beta$ which acts on the variable $u_\beta$. The problem is to couple the two models at their interface $\Gamma$. To do so, we introduce the air-sea flux $\mathscr{F}_{oa}$ and interface operators $\mathscr{C}_\beta$. Those operators must be chosen to satisfy the required consistency on $\Gamma$.

Omitting the external boundary conditions, the equations driving the coupled column system can be summarised as:

$$\begin{cases} \mathscr{L}_a(u_a) = f_a & \text{on } \Omega_a \times T_W \\ u_a(z,0) = u_0(z) & z \in \Omega_a \\ \mathscr{C}_a(u_a) = \mathscr{F}_{oa}(u_a, u_o) & \text{on } \Gamma \times T_W \end{cases} \quad \begin{cases} \mathscr{L}_o(u_o) = f_o & \text{on } \Omega_o \times T_W \\ u_o(z,0) = u_0(z) & z \in \Omega_o \\ \mathscr{C}_o(u_o) = \mathscr{F}_{oa}(u_a, u_o) & \text{on } \Gamma \times T_W \end{cases} \tag{1}$$

Where $T_W = [0, T]$, and $f_\beta \in L^2(0, T; L^2(\Omega_\beta))$ are given right-hand sides.

In the vast majority of models, at least part of the vertical equations are solved using an implicit scheme, meaning that, in order to get consistency at the interface ($\mathscr{C}_a(u_a) = \mathscr{C}_o(u_o)$), one needs to solve the equations on the whole column at once, which is impracticable in an air-sea context. Moreover, time discretisation being significantly different in the atmosphere and the ocean the meaning of this sought equality may not be obvious.

Most of the time this difficulty is overcome by using so-called asynchronous coupling where the flux seen by the atmosphere model is computed using the (possibly averaged) ocean state from the previous time interval (see Figure 1), sacrificing the flux consistency on the altar of practicability.

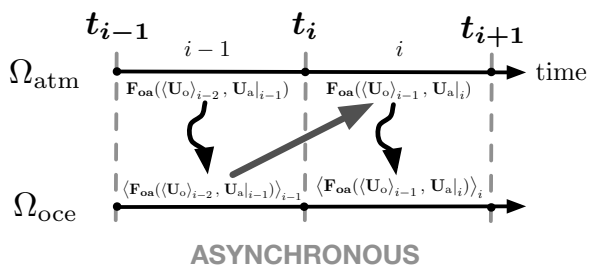



Figure 1: Schematics of asynchronous coupling

Instead, Lemarié et al. (2013*a*,*b*) propose to use a global-in-time Schwarz algorithm (a.k.a. Schwarz Waveform Relaxation, SWR see Gander (2008) for a review) to solve the corresponding coupling problem. This method consists in solving iteratively each model on their respective space-time subdomain using the interface conditions on $\Gamma$ computed during the previous iteration. It can be seen as iterations of asynchronous coupling until convergence.

For a given initial condition $u_0 \in H^1(\Omega_a \cup \Omega_o)$ and *first-guess* $u_o^0(0, t)$, the corresponding coupling algorithm reads

$$\begin{cases} \mathscr{L}_a(u_a^k) = f_a & \text{on } \Omega_a \times T_W \\ u_a^k(z,0) = u_0(z) & z \in \Omega_a \\ \mathscr{C}_a(u_a^k) = \mathscr{F}_{oa}(u_a^k, u_o^{k-1}) & \text{on } \Gamma \times T_W \end{cases} \quad \begin{cases} \mathscr{F}_o(u_o^k) = f_o & \text{on } \Omega_o \times T_W \\ u_o^k(z,0) = u_0(z) & z \in \Omega_o \\ \mathscr{C}_o(u_o^k) = \mathscr{F}_{oa}(u_a^k, u_o^k) & \text{on } \Gamma \times T_W \end{cases} \tag{2}$$

where $k$ is the iteration number. At convergence, this algorithm provides a mathematically strongly coupled solution which satisfies $\mathscr{C}_a(u_a) = \mathscr{C}_o(u_o)$ on $\Gamma \times T_W$. The convergence speed of the method greatly depends on the choice for $\mathscr{C}_\beta$ operators, and the choice of the *first-guess* $u_o^0(0,t)$. For $k = 1$ and if $u_o^0$ is the ocean state at previous time, then SWR is equivalent to asynchronous coupling.

Note that this is the sequential version of the algorithm, a parallel version can be obtained using $\mathscr{C}_o(u_o^k) = \mathscr{F}_{oa}(u_a^{k-1}, u_o^k)$ instead, as interface condition for the ocean. Both systems can then be solved in parallel, but it is likely to require more iterations to converge. Whether the increase in parallelism compensates the degraded convergence depends on application and hardware configuration.

It is difficult to advocate for SWR for operational ocean-atmosphere coupled systems, since it requires to run several instances of the model over the same time widow, significantly increasing the computational burden. However coupled data assimilation is an opportunity to improve the flux consistency and SWR methods can be used both as a reference and an inspiration to adapt data assimilation techniques.

## 3 Data assimilation

Let us now suppose that some discrete estimates $\mathbf{y}$ of the solution to problem (1) are available over an irregular set of points in the interval $\Omega \times T_W$. In this context we are interested in using a data assimilation (DA) procedure to account for this additional source of information. For the present study we use the variational methods of DA, based on optimal control theory. Our aim is to evaluate a set of parameter $\mathbf{x}_0$, including for instance the initial condition $u_0$ of problem (1), through the minimisation of a cost function $J(\mathbf{x}_0)$ ($\mathbf{x}_0$ is the control vector) which quantifies in some sense the misfit between the observations $\mathbf{y}$ and the model prediction. This minimisation requires the gradient of $J(\mathbf{x}_0)$, which can be computed using adjoint methods (Le Dimet & Talagrand 1986).

The cost function is generally composed of two terms: the *observation term* $J_o$, which penalises the misfit between model trajectory and observations $\mathbf{y}_{t_i}^o$ of the system, and the *background term* $J_b$ which penalises the departure to a given reference estimate $\mathbf{x}^b$.

$$J(\mathbf{x}_0) = \underbrace{\left(\mathbf{x}_0 - \mathbf{x}^b\right)^T \mathbf{B}^{-1} \left(\mathbf{x}_0 - \mathbf{x}^b\right)}_{J_b} + \underbrace{\sum_{i=0}^{N} \left(\mathscr{H}_{t_i}(\mathscr{M}_{t_i}(\mathbf{x}_0)) - \mathbf{y}_{t_i}^o\right)^T \mathbf{R}_{t_i}^{-1} \left(\mathscr{H}_{t_i}(\mathscr{M}_{t_i}(\mathbf{x}_0)) - \mathbf{y}_{t_i}^o\right)}_{J_o} \tag{3}$$

Where $\mathbf{R}_{t_i}$ is the observation error covariance matrix associated with observation at time $t_i$ and $\mathbf{B}$ is the background error covariance matrix. $\mathscr{M}_{t_i}(\mathbf{x}_0)$ is the solution of the numerical model equations (*e.g.* the one described in equation 2) integrated from time $t_0$ to time $t_i$, and $\mathscr{H}_{t_i}$ is the observation operator that maps the model state space onto the observation space at time $t_i$.

A very important aspect for coupled data assimilation is the explicit description or air-sea error correlations (*i.e.* the off-diagonal blocks of the $\mathbf{B}$ matrix), but this is not the topic of this report and is addressed in other tasks of WP2 (Smith et al. 2017, for instance). The main focus here is on coupling consistency, so, in the following we assume that:

- $\mathbf{B} = \begin{pmatrix} \mathbf{B}_a & 0 \\ 0 & \mathbf{B}_o \end{pmatrix}$ is the coupled system background error covariance and $\mathbf{B}_a$ and $\mathbf{B}_o$ are the atmosphere and ocean background error covariances matrices respectively (no explicit cross media error covariances are considered)

- $\mathscr{H}(.) = \begin{pmatrix} \mathscr{H}_a(.) & 0 \\ 0 & \mathscr{H}_o(.) \end{pmatrix}$ is the observation operator whose components are $\mathscr{H}_a(.)$ for atmospheric and $\mathscr{H}_o(.)$ for oceanic observations.

- The numerical coupled model can be written with a slight abuse of notation as

$$\mathscr{M}(.) = \begin{pmatrix} \mathscr{M}_a(.) & \mathscr{M}_{ao}(.) \\ \mathscr{M}_{oa}(.) & \mathscr{M}_o(.) \end{pmatrix}$$

  It includes both atmosphere ($\mathscr{M}_a(.)$) and ocean ($\mathscr{M}_o(.)$) as well as coupling ($\mathscr{M}_{ao}(.)$, $\mathscr{M}_{oa}(.)$) components

### 3.1 Uncoupled variational data assimilation

Historically the assimilation has been performed separately on the ocean and atmosphere. In a variational data assimilation context, it means that for $\beta = a$ and $\beta = o$, the control vector is restricted to subdomain $\Omega_\beta$ and is noted $\mathbf{x}_{0,\beta} = u_0|_{z\in\Omega_\beta}$. The optimal control problem amounts to find $\mathbf{x}^a_{0,\beta}$, the *analysed state*, which best fit observations $\mathbf{y}$ and a previous estimate of the initial state $\mathbf{x}^b_\beta$ called the *background*. Noting $\mathscr{H}_{t_i}(.)$ the observation operator that goes from model space to observations space at time $t_i$ and $\mathbf{x}_\beta = u_\beta$ the state vector, each cost function to minimise reads (for $\beta = a$ and $\beta = o$)

$$J_\beta(\mathbf{x}_{\beta,0}) = \left(\mathbf{x}_{\beta,0} - \mathbf{x}^b_\beta\right)^T \mathbf{B}^{-1}_\beta \left(\mathbf{x}_{\beta,0} - \mathbf{x}^b_\beta\right) + \sum_{i=0}^{N} \left(\mathscr{H}_{\beta,t_i}(\mathscr{M}_{\beta,t_i}(\mathbf{x}_{\beta,0})) - \mathbf{y}^o_{\beta,t_i}\right)^T \mathbf{R}^{-1}_{\beta,t_i} \left(\mathscr{H}_{\beta,t_i}(\mathscr{M}_{\beta,t_i}(\mathbf{x}_{\beta,0})) - \mathbf{y}^o_{\beta,t_i}\right) \quad (4)$$

where $\mathbf{R}_\beta$ is the covariance matrix associated to observation errors and $\mathbf{B}_\beta$ is the background error covariance matrix. $\mathscr{M}_{\beta,t_i}(\mathbf{x}_0))$ is the solution of the model integration from 0 to $t_i$ starting from initial condition $\mathbf{x}_{\beta,0}$ for $\beta = a$ and $\beta = o$ independently. Obviously, if the DA process is done separately on each subdomain (with prescribed boundary conditions on the interface $\Gamma$), the initial condition $u_0 = (\mathbf{x}^a_{a,0}, \mathbf{x}^a_{o,0})^T$ obtained on $\Omega$ does not satisfy the interface conditions, hence $u_0 \notin H^1(\Omega)$ and well-posedness of the coupled problem is no longer guaranteed. In practice this type of imbalance in the initial condition can severely damage the forecast skills of coupled models (Mulholland et al. 2015).

### 3.2 Toward a coupled variational data assimilation

Deriving data assimilation methods able to properly account for the coupling is therefore an important matter. This section aims at providing methods leading to a solution close to the observations while satisfying the interface conditions on $\Gamma$; or at least a weak form of it. The key properties of those algorithms are summarised in Tab. 1.

#### Fully Coupled Method (FCM)

A first possibility is to consider a monolithic view of the problem by ignoring the presence of an interface in the assimilation process. In this case the state vector is $\mathbf{x_0} = u_0(z),\ z \in \Omega$ and for each model integration we iterate the models on $\Omega_a$ and $\Omega_o$ either by asynchronous coupling, till convergence of the Schwarz algorithm or whatever coupling method is used.

The cost function for the FCM is

$$J_{FCM}(\mathbf{x}_0) = \left(\mathbf{x}_0 - \mathbf{x}^b\right)^T \mathbf{B}^{-1} \left(\mathbf{x}_0 - \mathbf{x}^b\right) + \sum_{i=0}^{N} \left(\mathscr{H}_{t_i}(\mathscr{M}_{t_i}(\mathbf{x}_0)) - \mathbf{y}^o_{t_i}\right)^T \mathbf{R}^{-1}_{t_i} \left(\mathscr{H}_{t_i}(\mathscr{M}_{t_i}(\mathbf{x}_0)) - \mathbf{y}^o_{t_i}\right) \quad (5)$$

where $\mathbf{x}(t) = (u_a(t), u_o(t))^T$. It can readily be seen that cost function (5) is identical to the cost function we would use for an uncoupled problem defined on $\Omega$. The solution provided by this approach is as coupled as the forecast system $\mathscr{M}(.)$. It also allows to relax the block-diagonal nature of $\mathbf{B}$ and $\mathscr{H}(.)$, but for the sake of comparison with other algorithms it is not considered here. If a SWR algorithm is used for the computation of $\mathbf{x}(t)$, first-guess $u_a^0$ in (2) is updated after each minimisation iteration with the converged solution obtained during the previous model integration. The Schwarz algorithm then converges more rapidly over the minimisation iteration. Note that the FCM requires the adjoint of the strongly coupled model (2) which can be tedious to derive. The main drawback of this method with SWR coupling is that it possibly requires a very large number of Schwarz iterations since it systematically iterates till convergence.

**Partially Coupled Method (PCM)**

In this variant only one part of the coupling process is accounted for when computing $\mathbf{x}(t)$ in the cost function. It is aimed at improving the coupling consistency of the solution through data assimilation by adding a penalty term in the cost function. For instance one can propose to truncate the Schwarz iterations in the direct and adjoint model after $k_{\max}$ iterations, with $k_{\max} < k_{\text{cvg}}$. Because we do not iterate till convergence, the coupled solution does not strictly satisfy the interface consistency. The equivalent approach in asynchronous coupling ($k_{\max} = 1$) would be to use one-way coupling to compute the trajectory and to aim at 'promoting' it to two way through a penalty term in the cost function.

As proposed by Gejadze & Monnier (2007) in the context of river hydraulics, a convenient way to propagate the information from one subdomain to the other during the minimisation iterations is to use an extended cost function which includes the misfit in the interface conditions. The idea behind this approach is to enforce a weak coupling within the minimisation iterations. The control vector $\mathbf{x}_0 = (u_0(z), u_o^0(0,t))^T$ now includes the first-guess on the interface and the cost function reads

$$J_{PCM}(\mathbf{x}_0) = J^b(\mathbf{x}_0) + \sum_{i=0}^{N} \left(\mathscr{H}_{t_i}(\mathscr{M}_{t_i}^{trunc}(\mathbf{x}_0)) - \mathbf{y}_{t_i}^o\right)^T \mathbf{R}_{t_i}^{-1} \left(\mathscr{H}_{t_i}(\mathscr{M}_{t_i}^{trunc}(\mathbf{x}_0)) - \mathbf{y}_{t_i}^o\right) + J^s(\mathbf{x}_0) \tag{6}$$

where $\mathscr{M}_{t_i}^{trunc}(\mathbf{x}_0)) = (u_a^{k_{\max}}(t_i), u_o^{k_{\max}}(t_i))^T$ and

$$J^s(\mathbf{x}_0) = \gamma \|\mathscr{C}_a(u_a(0,t)) - \mathscr{C}_o(u_o(0,t))\|_{[0,T]}^2. \tag{7}$$

Unlike FCM, the first-guess for the interface is part of the control vector here, but this method still requires part of the adjoint of the coupling. Note that since the first-guess $u_o^0$ is updated at the end of each minimisation iteration, we can expect that we will converge toward a good approximation of the SWR solution.

**Weakly Coupled Method (WCM)**

The last possibility we propose to investigate is to suppress the coupling operators in the model and rely solely on the minimisation iterations to weakly couple the two models. This approach only requires the adjoint of each individual model but not the adjoint of the coupling as for the previous algorithms. The control vector is $\mathbf{x}_0 = (\mathbf{x}_{a,0}, \mathbf{x}_{o,0})^T$ with $\mathbf{x}_{0,\beta} = (\, u_0|_{z\in\Omega_\beta}, u_\beta^0(0,t))$. The corresponding cost function is

$$J_{WCM}(\mathbf{x}_0) = J_a^b(\mathbf{x}_{a,0}) + J_o^b(\mathbf{x}_{o,0}) + J_a^o(\mathbf{x}_{a,0}) + J_o^o(\mathbf{x}_{o,0}) + J^s(\mathbf{x}_0)$$

It is straightforward to see that this algorithm provides only a weakly coupled solution. One model integration is performed (possibly in parallel for ocean and atmosphere) with boundary conditions on $\Gamma$ provided by the term $u^0_\beta(0,t)$ taken from the control vector.

Table 1 summarises different aspects of the three proposed algorithms. In the 'adjoint of the coupling' column, online means that it is required along with the adjoint model during the $\nabla J^o$ computation, while offline means that it is only required during $\nabla J^s$ computation, so that contributions to $\nabla J^o$ from ocean and atmosphere can be computed in parallel.

| Algo | Control vector | # of coupling iterations | extended cost function | Adjoint of the coupling | Coupling |
|---|---|---|---|---|---|
| FCM | $(u_0(z))$ | $k_{\text{cvg}}$ | no | online | strong |
| PCM | $(u_0(z), u^0_a)^T$ | $k_{\max}$ | yes | partially online/offline | ~strong |
| WCM | $(u_0(z), u^0_a, u^0_o)^T$ | 0 | yes | offline | weak |

Table 1: Overview of the properties of the coupled variational DA methods described in Sec. 3.2. Notations are consistent with those introduced in the text.

## 4 Incremental formulation and link to the CERA system

In classical 4D-Var, due to non linearities in $\mathscr{M}$ and $\mathscr{H}$, minimising efficiently (4) is not straightforward. Common practice is to use the so called incremental 4D-Var approach (a.k.a. Gauss-Newton in the optimisation community) where the original problem is solved through successive minimisations of quadratic cost functions (inner loops)

$$J^l(\delta\mathbf{x}^l) = \left(\delta\mathbf{x}^l + \sum_{k=1}^{l-1}\delta\mathbf{x}^k\right)^T \mathbf{B}^{-1}\left(\delta\mathbf{x}^l + \sum_{k=1}^{l-1}\delta\mathbf{x}^k\right) + \sum_{i=0}^{N}\left(\mathbf{H}^{l-1}_{t_i}\mathbf{M}^{l-1}_{t_i}\delta\mathbf{x}^l - \mathbf{d}^{l-1}_{t_i}\right)^T \mathbf{R}^{-1}_{t_i}\left(\mathbf{H}^{l-1}_{t_i}\mathbf{M}^{l-1}_{t_i}\delta\mathbf{x}^l - \mathbf{d}^{l-1}_{t_i}\right) \quad (8)$$

with $\mathbf{d}^0_{t_i} = \mathbf{y}^o_{t_i} - \mathscr{H}_{t_i}(\mathscr{M}_{t_i}(\mathbf{x}_0))$ and $\mathbf{d}^l_{t_i} = \mathbf{y}^o_{t_i} - \mathscr{H}_{t_i}\left(\mathscr{M}_{t_i}(\mathbf{x}_0 + \sum_{i=1}^l \delta\mathbf{x}^i)\right)$

and $\mathbf{M}^l_{t_i}$ (resp. $\mathbf{H}^l_{t_i}$) being the tangent linear operator of $\mathscr{M}_{t_i}$ (resp $\mathscr{H}_{t_i}$) differentiated around $\mathbf{x}_0 + \sum_{i=1}^l \delta\mathbf{x}^i$

Non linearities are therefore accounted for through the re-linearisation of the $\mathbf{M}^l$ and $\mathbf{H}^l_{t_i}$ operators and the computation of the innovation vectors $\mathbf{d}^l$. Under some regularity hypotheses, such algorithm is known to converge toward the solution of the original problem (see appendix C for more details).

FCM being a direct transposition of classical 4D-Var, its incremental variant uses the same inner loop formulation as equation 8. For PCM and WCM, both $J^l_o$ and $J^l_b$ can be obtained similarly as equation 8 as well, and the inner loop coupling penalty term reads

$$J^l_s = \gamma\|\mathbf{C}_a\delta u^l_a(0,t) - \mathbf{C}_o\delta u^l_o(0,t) + \mathscr{C}_a(u^{l-1}_a(0,t)) - \mathscr{C}_o(u^{l-1}_o(0,t))\|^2_{[0,T]}$$

$\mathbf{C}_a$ and $\mathbf{C}_o$ being the tangent linear operators of $\mathscr{C}_a$ and $\mathscr{C}_o$ respectively.

The CERA system uses a different formulation compared to regular incremental 4Dvar, indeed it aims at solving the FCM problem by successive inner loop approximations that are uncoupled, *i.e.* CERA uses

$\mathscr{M}(.) = \begin{pmatrix} \mathscr{M}_a(.) & \mathscr{M}_{ao}(.) \\ \mathscr{M}_{oa}(.) & \mathscr{M}_o(.) \end{pmatrix}$ in the outer iterations, but $\mathbf{M}^l = \begin{pmatrix} \mathbf{M}_a^l & 0 \\ 0 & \mathbf{M}_o^l \end{pmatrix}$ in the inner loops. Roughly speaking, CERA can be seen as a parallel Schwarz algorithm for solving FCM minimisation, with the important limitation that it cannot cope with air-sea correlation in $\mathbf{B}$. Likewise, PCM with no $J_s$ term ($\gamma = 0$) can be seen as a sequential Schwarz algorithm for solving FCM.

Convergence properties of algorithms presented in this report are discussed in appendix C.

## 5 Application to a 1D linear diffusion problem

In this section, previous algorithms are applied on a 1D diffusion problem. We, thus, consider for both part of the system $\mathscr{L}_\beta = \partial_t + \nu_\beta \partial_z^2$ in (2), with $\beta = a$ and $\beta = o$ and with $\nu_a \neq \nu_o$ the diffusion coefficients in each subdomain. The computational domain is $\Omega = ]-L_a, L_o[$ with $L_a, L_o \in \mathbb{R}^{+*}$. We choose the interface operators on $\Gamma$ to obtain Dirichlet-Neumann condition, i.e. $\mathscr{C}_a = \begin{pmatrix} \nu_a \partial_z & 0 \\ 0 & 1 \end{pmatrix}$ and $\mathscr{C}_o = \begin{pmatrix} 1 & 0 \\ 0 & \nu_o \partial_z \end{pmatrix}$. We consider the analytical solution $u_\beta^\star$, and the corresponding right hand side $f_\beta = \mathscr{L}_\beta u_\beta^\star$, of the coupled problem on each subdomain as :

$$u_\beta^\star(z,t) = \frac{U_0}{4} e^{-\frac{|z|}{\varepsilon_\beta}} \left\{ 3 + \cos^2\left(\frac{3\pi t}{\tau}\right) \right\} \qquad \text{on } \Omega_\beta \times T_W. \tag{9}$$

where $U_0 = 20$ °C and $\tau = 22$ h. Note $\varepsilon_a \nu_o = \varepsilon_o \nu_a$ is required to ensure the proper regularity of the coupled solution across the interface $\Gamma$. To satisfy this constraint we choose $\varepsilon_a = 4$ km, $\varepsilon_o = 0.4$ km, $\nu_a = 1\ \mathrm{m}^2/\mathrm{s}$, $\nu_o = 0.1\ \mathrm{m}^2/\mathrm{s}$. The model problem (2) is discretized using a backward Euler scheme in time and a second-order scheme in space. The resolution in each subdomain is $\Delta z = 20$ m with $L_a = L_o = 1$ km and the time-step is $\Delta t = 180$ s. The total simulation time is $T = 12$ h, which is also the size of the SWR window ($T_W$ in equation 2). The latter implies that at least 2 iterations of SWR are necessary to get some coupling in the models integration.

For the assimilation experiments, we consider true-state $\mathbf{x}^t$ to be the analytical solution while background $\mathbf{x}^b$ corresponds to the solution obtained with a biased initial condition. In general, the Schwarz algorithm converges in $k_{\mathrm{cvg}} = 50$ iterations with a tolerance $\epsilon = 10^{-6}$. Some observations $\mathbf{y}$ of the true-state are generated such that $\mathbf{y} = H(\mathbf{x}^t)$, with $H$ the observation operator. The observation and background errors covariance matrices are considered diagonal such that $\mathbf{R} = 10$ Id and $\mathbf{B} = 100$ Id. For the extended cost function we consider different values of $\gamma$. All the minimisation are done until convergence of a conjugate gradient algorithm with a stopping criterion $\| \nabla J(\mathbf{x}_0) \|_\infty < 10^{-5}$.

### Single column observation experiment

For our experiments, we consider that observations are available in $\Omega \setminus \{\Gamma\}$ only at the end of the time-window (i.e. at $t = T$). In this case, the results obtained for different assimilation schemes are reported in table 2 where the performance of each scheme is presented both in terms of number of minimisation and computational cost. The latter being given relative to that of uncoupled data assimilation. Possible parallel aspects are not accounted for in thes measure of computational cost. To evaluate the strength of the coupling we define an *interface imbalance indicator* which corresponds to the value of $J_s$ at the end of the DA process, with $\gamma = 1$. Values of $J^s$ close to zero indicate a strongly coupled analysed state. In table 2, a root mean square error (RMSE) defined as $\sqrt{\mathbb{E}\left((\mathbf{x}^a - \mathbf{x}^t)^2\right)}$ on $\Omega \times T_W$ is also used to evaluate how much the analysed state is close to the true-state.

Experiment names start with the algorithm name, then, if relevant, the number of Schwarz iteration (F stands for full convergence) and finally, if relevant, Js$\gamma$ value. From table 2, we can first note that the

| Algo | $\gamma$ | $k_{\max}$ | # of minimisation iterations | Computing cost (relative to uncoupled) | Interface imbalance indicator | RMSE in °C |
|---|---|---|---|---|---|---|
| FCM-F | - | $k_{\text{cvg}}$ | 14 | 28.5 | $3.\,10^{-12}$ | 0.221 |
| FCM-2 | - | 2 | 14 | 1.3 | 9.866 | 0.415 |
| PCM-F | 0 | $k_{\text{cvg}}$ | 14 | 28.5 | $3.\,10^{-12}$ | 0.221 |
| PCM-5 | 0 | 5 | 14 | 3.1 | $4.\,10^{-2}$ | 0.220 |
| PCM-2 | 0 | 2 | 14 | 1.3 | 9.87 | 0.415 |
| PCM-2-Js0.1 | 0.1 | 2 | 184 | 15.8 | $2.\,10^{-4}$ | 0.216 |
| PCM-1-Js0.1 | 0.1 | 1 | 117 | 5 | $6.\,10^{-9}$ | 0.217 |
| WCM-Js1. | 1.0 | 0 | 365 | 16.3 | $1.\,10^{-8}$ | 0.286 |
| WCM-Js0.1 | 0.1 | 0 | 396 | 16.9 | $1.\,10^{-6}$ | 0.228 |
| WCM-Js0.01 | 0.01 | 0 | 390 | 16.6 | $1.\,10^{-4}$ | 0.226 |
| Uncoupled | 0 | 0 | 22 | 1 | 13700 | 8.338 |

Table 2: Results obtained for the three coupled variational DA methods described in Sec. 3.2 with several settings. Observations are available in $\Omega \setminus \{\Gamma\}$ at the end of the time-window.

FCM algorithm requires few minimisation iterations to obtain a low RMSE value and a strongly coupled analysed state ($J^s \sim 10^{-12}$). A drawback of this approach is a high computational cost (almost 30 times that of uncoupled). PCM-F (resp -2) only differs from FCM-F (resp -2) by the first guess of the interface within the control vector, and gives pretty much the same results as FCM-F (resp -2), for the same computing cost. This shows that without $J^s$ adding the interface in the control has little effect.

Since in the other PCM approaches, coupling iterations are truncated and first-guess $u_a^0$ is part of the control vector, we expect a reduced computational cost compared to FCM-F. If $\gamma = 0$ (no $J_s$ term, PCM-2 and 5) the same number of iterations are required for the minimisation to converge reducing dramatically the cost over FCM, but at the expense of a lower quality analysis. On the one hand decreasing the value of $k_{\max}$ increases the number of minimisation iterations. Indeed, going to Schwarz convergence ($k_{\max} = k_{\text{cvg}}$) procures the best model solution, it then needs few minimisation iterations. However, for the next iteration, the background interface is given by the control vector rather than the previous converged estimate; therefore it requires again numerous Schwarz iterations. On the other hand, by reducing the $k_{\max}$ value, the number of Schwarz iterations is reduced and the update of first-guess more significant. However the coupling quality is affected and this leads to a slower minimisation convergence. Here, a good compromise is to choose $k_{\max} = 5$.

When $J_s$ term is activated (PCM-2-Js0.1 and -1-Js0.1) the number of iterations required for the minimisation to converge rises significantly as it is usually the case when one adds constraint to the cost function. However, computing cost is still significantly lower than that of FCM-F while reaching similar quality analysis. Smaller values of $k_{\max}$ provide a faster convergence of the algorithm. With $k_{\max} = 1$, which corresponds to a one-way coupling, it requires only 5 times the cost of the uncoupled DA to provide a good approximate of the strongly coupled solution ($J^s = 6.87\ 10^{-9}$, RMSE = 0.217 °C). For $k_{max} > 1$ two ways interactions are accounted for in the model integration, making interactions between the con-

trol vector components more intricate and therefore damaging the convergence properties.

By considering uncoupled models in the WCM algorithm, the convergence property of the minimisation is severly damaged, as can be seen in figure 2 This is true whatever the choice of $\gamma$ to balance $J^s$ and $J^o$ in the cost function and leads to a significant increase of computing cost compared to PCM. The analysed state shows a larger interface imbalance indicator compared to FCM and the best PCMs, which confirms that WCM provides a weakly coupled solution, but is significantly better than the uncoupled DA in that respect. The RMSE level is similar to that of FCM and most PCMs though.

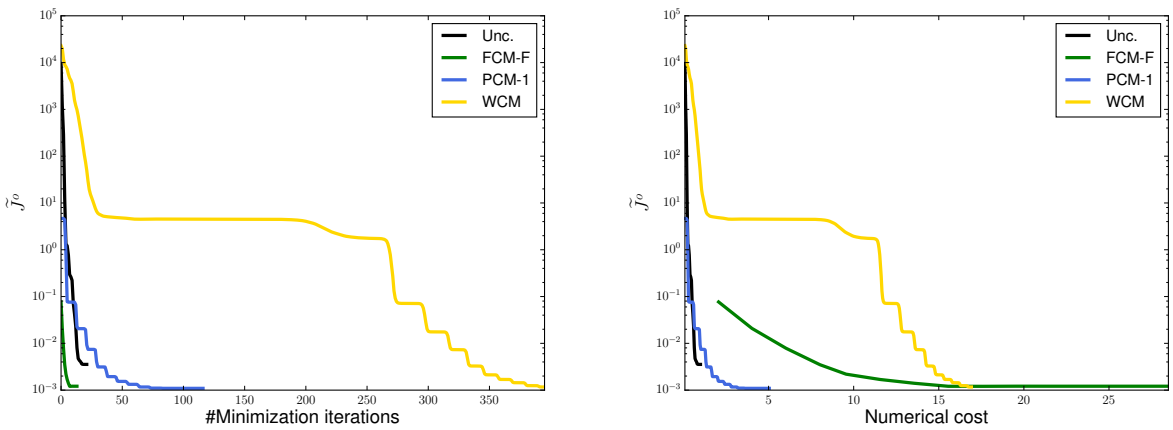



Figure 2: $J^o$ evolution for the different schemes, along minimisation iterations (left panel) and corresponding computing cost (right panel, unit: uncoupled minimisation cost)

Table 2 presents results at convergence of the minimisation. In realistic ocean-atmosphere coupling such convergence is generally not affordable so rate of convergence may be more important than the actual minimum. Table 3 shows the same results as 2 for the best PCM and WCM limiting their cost to that of uncoupled minimisation.

| Algo | $\gamma$ | $k_{\max}$ | # of minimisation iterations | Computing cost (relative to uncoupled) | Interface imbalance indicator | RMSE in °C |
|---|---|---|---|---|---|---|
| **PCM-1-Js0.1** | **0.1** | **1** | **22** | **1** | **5. 10$^{-4}$** | **0.353** |
| WCM-Js0.1 | 0.1 | 0 | 22 | 1 | 2980 | 2.51 |
| Uncoupled | 0 | 1 | 22 | 1 | 13700 | 8.338 |

Table 3: Same as previous table but limiting the number of minimisation iterations to 24

Both approaches outperform the uncoupled DA by far, PCM-1-Js0.1 being even quite close to its converged state, with less than a third of its original cost. For WCM to be a viable option, in that context, some substantial work has to be performed on preconditioning, in order to significantly improve convergence.

In the linear case, a CERA equivalent can be defined using outer and inner loops, with only the outer loop being coupled. Here the incremental approach is used to account for the coupling instead of the

non linearities. Any strength of coupling can be used in the outer iterations. Results are given in table 4 for fully converged coupling and 2 iterations of Schwarz (corresponding FCM results are recalled). Both CERA-F and CERA-2 give similar results to their FCM counterparts with a much higher iteration count, but a smaller computational cost. In CERA-2 the coupling Schwarz iterative scheme restart from the same initial interface condition at each outer iteration. CERA-2-save on the other hand, reuses the output of the Schwartz algorithm from the previous outer iteration, which allows to really improve the strength of the coupling for very little extra-cost. This improvement is really related to the way the coupling is done in this case, though. Indeed since the SWR window is the same as the assimilation window, the background of the interface has a strong impact when using low SWR number of iteration.

| Algo | $\gamma$ | $k_{\max}$ | # of minimisation iterations | Computing cost (relative to uncoupled) | Interface imbalance indicator | RMSE in °C |
|---|---|---|---|---|---|---|
| FCM-F | - | $k_{\text{cvg}}$ | 14 | 28.5 | $3.\,10^{-12}$ | 0.221 |
| FCM-2 | - | 2 | 14 | 1.3 | 9.866 | 0.415 |
| CERA-F | – | $k_{\text{cvg}}$ | 188 | 15.5 | $4.\,10^{-12}$ | 0.271 |
| CERA-2 | – | 2 | 190 | 8.95 | 10.37 | 0.49 |
| CERA-2-save | – | 2 | 194 | 9.13 | $8.\,10^{-5}$ | 0.270 |

Table 4: Same as previous table but for FCM and CERA experiments

Fig 3 presents the evolution of outer and inner cost functions during minimisation for CERA-F, CERA-2 and CERA-2-save. It shows that CERA-2-save benefits from previous outer iteration to improve its coupling convergence and even outperform FCM-2. On the downside, figure 3 right panel shows erratic inner minimisation behaviour with some oscillation around the optimum making the choice of stopping criterion complicated. It also shows a degradation of the convergence properties of CERA-2-save successive inner loops (growing offset between vertical dashed lines).

It is difficult to draw a definitive conclusion from such a simple test case. The next section goes a step further toward realistic applications and present preliminary results on a more complex ocean-atmosphere single column model.

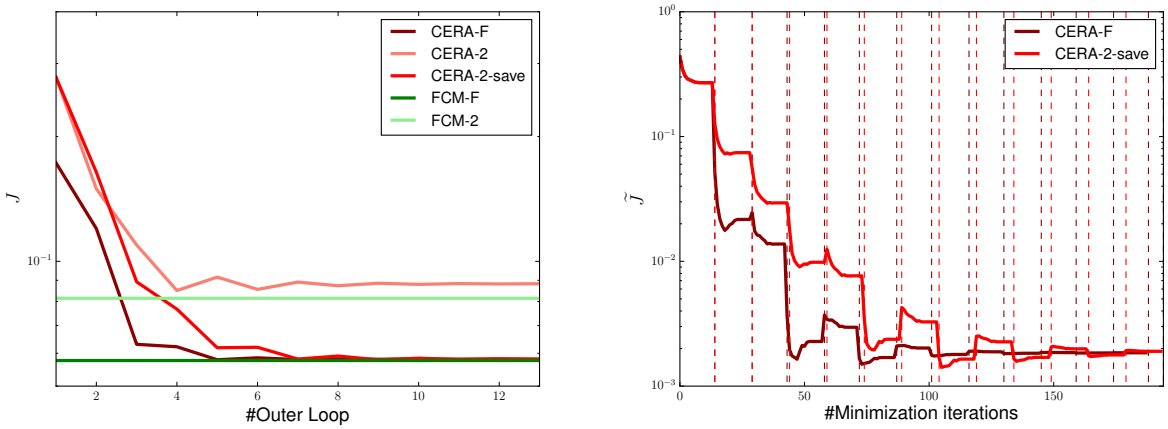
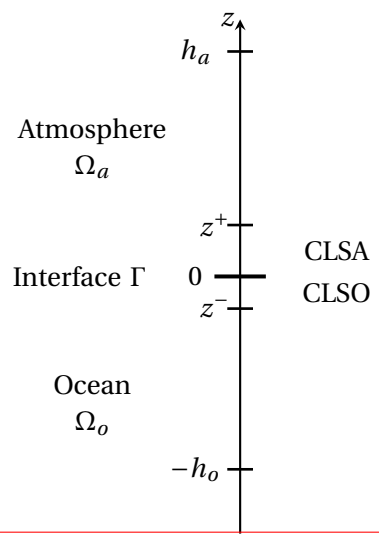



Figure 3: left panel: outer cost function values for different flavours of CERA (in red) and corresponding FCM optimum value (horizontal green lines). Right panel: inner cost function values for CERA-F and CERA-2-save vertical dashed lines mark outer iterations.

## 6 Application to a single column ocean-atmospheric boundary layer model

In order to perform further tests of our algorithmic developments, we set up a more realistic single column model of a column of the ocean coupled with a boundary layer of the atmosphere. The atmosphere is defined, as presented opposite, on domain $\Omega_a = [0, h_a]$ and the ocean on $\Omega_o = [-h_o, 0]$. The surface boundary layer is located between $z^+$ in the atmosphere and $z^-$ in the ocean. Ocean atmosphere interface being represented by $\Gamma$. The time domain is $[0, T]$ with $T > 0$s. We can then express evolution equations for velocities $\mathbf{u}$ and tracers $\mathbf{t}$ (temperature and humidity for the atmosphere, temperature and salinity for the ocean) of our 1D single column model :




$$
\begin{aligned}
\frac{\partial \mathbf{u}_\beta(z,t)}{\partial t} &= -f\mathbf{k}\times\mathbf{u}_\beta(z,t) + \frac{\partial}{\partial z}\left(K_m^\beta(z)\frac{\partial \mathbf{u}_\beta(z,t)}{\partial z}\right) + F_{\mathbf{u}_\beta}(z,t) && \text{sur } \Omega_\beta\times[0,T]\\
\frac{\partial \mathbf{t}_\beta(z,t)}{\partial t} &= \frac{\partial}{\partial z}\left(K_s^\beta(z)\frac{\partial \mathbf{t}_\beta(z,t)}{\partial z}\right) + F_{\mathbf{t}_\beta}(z,t) && \text{sur } \Omega_\beta\times[0,T]\\
\rho_a K_m^a(z^+)\left.\frac{\partial \mathbf{u}_a}{\partial z}\right|_\Gamma &= \rho_o K_m^o(z^-)\left.\frac{\partial \mathbf{u}_o}{\partial z}\right|_\Gamma = \mathscr{F}_{oa}^m(\mathbf{u}_a,\mathbf{u}_o,\mathbf{t}_a,\mathbf{t}_o) && \text{sur } \Gamma\times[0,T]\\
\rho_a K_s^a(z^+)\left.\frac{\partial \mathbf{t}_a}{\partial z}\right|_\Gamma &= \rho_o K_s^o(z^-)\left.\frac{\partial \mathbf{t}_o}{\partial z}\right|_\Gamma = \mathscr{F}_{oa}^s(\mathbf{u}_a,\mathbf{u}_o,\mathbf{t}_a,\mathbf{t}_o) && \text{sur } \Gamma\times[0,T]
\end{aligned}
$$

where $\beta = a, o$ refer to atmosphere and ocean variables respectively. Both models use the same structure and differ from their forcing terms $F_*$, their interface conditions and the computation of their turbulent

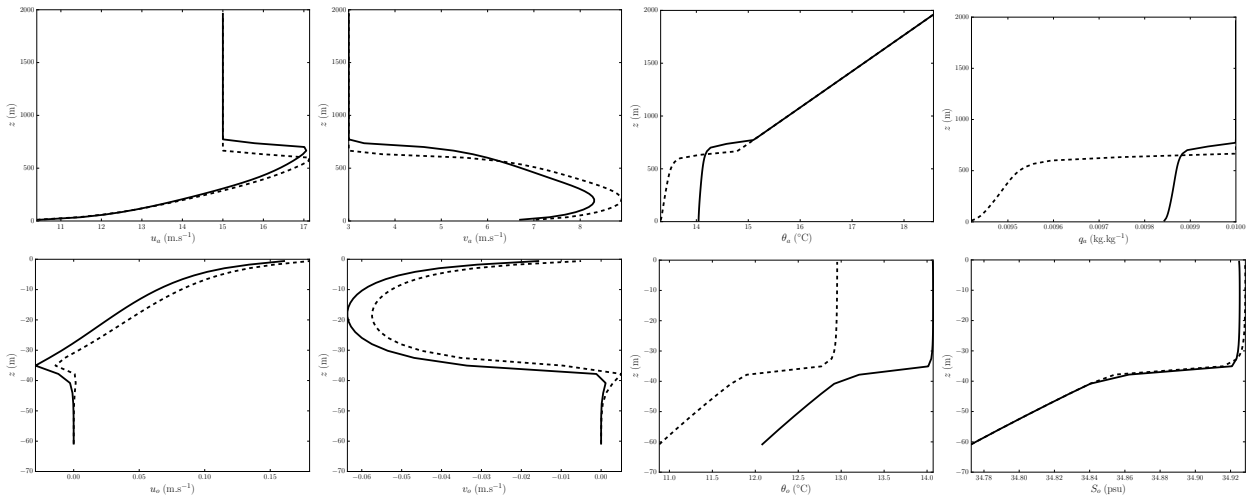


Figure 4: Truth (plain line) and background (dashed lines) initial conditions for atmospheric (top) and oceanic (bottom) quantities

viscosity and diffusivity coefficients $K_m^\beta$ and $K_s^\beta$. The construction of this model from the 3D equations, along with the definition of air-sea fluxes $\mathscr{F}_{oa}^*$ are described extensively in appendix A

### 6.1 Model and data assimilation settings

The model settings used in this study are described in appendix B. The assimilation window is chosen to be 12h, starting from day 2, so that it includes a switch from stable to unstable regime. The SWR window is set to 1h, so that one-iteration only will yield a classical asynchronous coupling with exchange of fluxes every hour. Observation are taken every 3h and all the vertical column is observed and perturbed with a white noise, corresponding to specified observation errors variances in **R**, namely

$$\begin{aligned} \mathbf{r}_{u_a} &= 0.4 & \mathbf{r}_{u_o} &= 0.0075 \\ \mathbf{r}_{v_a} &= 0.4 & \mathbf{r}_{v_o} &= 0.0075 \\ \mathbf{r}_{\theta_a} &= 0.35 & \mathbf{r}_{\theta_o} &= 0.8 \\ \mathbf{r}_{q_a} &= 0.0002 & \mathbf{r}_{S_o} &= 0.055 \end{aligned} \tag{10}$$

The background is obtained by perturbing the true state 3 days prior to the start of the assimilation window and propagated forward in time so that it is physically balanced. Both truth and background initial quantities are shown in figure 4.

The background error correlation matrix is considered block diagonal (no cross variable covariances), each block representing Gaussian covariances.

$$
\begin{aligned}
\mathbf{B}_{\mathbf{x}_{a,0}} &= \begin{pmatrix} \mathbf{B}_{u_a} & 0 & 0 & 0 \\ 0 & \mathbf{B}_{v_a} & 0 & 0 \\ 0 & 0 & \mathbf{B}_{\theta_a} & 0 \\ 0 & 0 & 0 & \mathbf{B}_{q_a} \end{pmatrix} \mathbf{B}_{\mathbf{x}_{o,0}} = \begin{pmatrix} \mathbf{B}_{u_o} & 0 & 0 & 0 \\ 0 & \mathbf{B}_{v_o} & 0 & 0 \\ 0 & 0 & \mathbf{B}_{\theta_o} & 0 \\ 0 & 0 & 0 & \mathbf{B}_{S_o} \end{pmatrix} \\
\mathbf{B}_{\mathbf{x}_{a,\Gamma}} &= \begin{pmatrix} \mathbf{B}_{u_{a,\Gamma}} & 0 & 0 & 0 \\ 0 & \mathbf{B}_{v_{a,\Gamma}} & 0 & 0 \\ 0 & 0 & \mathbf{B}_{\theta_{a,\Gamma}} & 0 \\ 0 & 0 & 0 & \mathbf{B}_{q_{a,\Gamma}} \end{pmatrix} \mathbf{B}_{\mathbf{x}_{o,\Gamma}} = \begin{pmatrix} \mathbf{B}_{\text{ssu}} & 0 & 0 \\ 0 & \mathbf{B}_{\text{ssv}} & 0 \\ 0 & 0 & \mathbf{B}_{\text{sst}} \end{pmatrix}
\end{aligned}
\tag{11}
$$

Let $\mathbf{b}_{i,j}$ being the $\mathbf{B}$ coefficient at line $i$ and column $j$. It is set to:

$$
\begin{aligned}
k &= \min(|j-i|, N-j+i) \qquad \text{where } N \text{ is the number of vertical levels} \\
\mathbf{b}_{i,j} &= \sigma^2 \exp\left(-\frac{(k\Delta_t)^2}{2L^2}\right) \\
\mathbf{b}_{j,i} &= \mathbf{b}_{i,j}
\end{aligned}
\tag{12}
$$

In this experiment, we also set:

$$
\begin{aligned}
&\sigma_{u_a} = 1 && L_{u_a} = 50 && \sigma_{u_o} = 0.02 && L_{u_o} = 15 \\
&\sigma_{v_a} = 1 && L_{v_a} = 50 && \sigma_{v_o} = 0.02 && L_{v_o} = 15 \\
&\sigma_{\theta_a} = 1 && L_{\theta_a} = 40 && \sigma_{\theta_o} = 1 && L_{\theta_o} = 50 \\
&\sigma_{q_a} = 0.0007 && L_{q_a} = 20 && \sigma_{S_o} = 0.2 && L_{S_o} = 30 \\
\\
&\sigma_{u_{a,\Gamma}} = 1 && L_{u_{a,\Gamma}} = 0 && \sigma_{\text{ssu}} = 0.1 && L_{\text{ssu}} = 0 \\
&\sigma_{v_{a,\Gamma}} = 1 && L_{v_{a,\Gamma}} = 0 && \sigma_{\text{ssv}} = 0.1 && L_{\text{ssv}} = 0 \\
&\sigma_{\theta_{a,\Gamma}} = 10 && L_{\theta_{a,\Gamma}} = 0 && \sigma_{\text{sst}} = 1.5 && L_{\text{sst}} = 0 \\
&\sigma_{q_{a,\Gamma}} = 0.001 && L_{q_{a,\Gamma}} = 0
\end{aligned}
\tag{13}
$$

Finally, the coupling penalty term is defined for PCM and WCM respectively as

$$
J^s_{PCM}(\mathbf{x}_0) = \alpha \left(F_{oa}(\mathbf{x}^k_a, \mathbf{x}^k_o, \mathscr{R}) - F_{oa}(\mathbf{x}^k_a, \mathbf{x}^{k-1}_o, \mathscr{R})\right)^T \mathbf{W}^{-1} \left(F_{oa}(\mathbf{x}^k_a, \mathbf{x}^k_o, \mathscr{R}) - F_{oa}(\mathbf{x}^k_a, \mathbf{x}^{k-1}_o, \mathscr{R})\right) \tag{14}
$$

and

$$
J^s_{WCM}(\mathbf{x}_0) = \alpha \left(F_{oa}(\mathbf{x}^0_a, \mathbf{x}_o, \mathscr{R}) - F_{oa}(\mathbf{x}_a, \mathbf{x}^0_o, \mathscr{R})\right)^T \mathbf{W}^{-1} \left(F_{oa}(\mathbf{x}^0_a, \mathbf{x}_o, \mathscr{R}) - F_{oa}(\mathbf{x}_a, \mathbf{x}^0_o, \mathscr{R})\right) \tag{15}
$$

with $\alpha = 0.001$ and $\mathbf{W}$ such that:

$$
\mathbf{W} = \begin{pmatrix} \mathbf{W}_{\tau_i} & 0 & 0 & 0 \\ 0 & \mathbf{W}_{\tau_j} & 0 & 0 \\ 0 & 0 & \mathbf{W}_{Q_{\text{net}}} & 0 \\ 0 & 0 & 0 & \mathbf{W}_F \end{pmatrix} \tag{16}
$$

where $\tau_*$ are wind stresses, $Q_{\text{net}}$ is the net heat flux and $F$ is the fresh water flux:

$$
\begin{aligned}
\mathbf{w}_{\tau_i} &= 0.001 \\
\mathbf{w}_{\tau_j} &= 0.001 \\
\mathbf{w}_{Q_{\text{net}}} &= 1 \\
\mathbf{w}_F &= 3\,10^{-6}
\end{aligned}
\tag{17}
$$

In order to mimic realistic settings, only two outer iterations will be performed. The inner minimisation will be allowed to go to convergence since they are quite efficient with these settings.

### 6.2 Numerical results

Table 5 presents a similar summary as of previous section, with the notable difference that the analysis error is represented as improvement over background, meaning the higher the better (background being at 0% and truth at 100%). It is computed as the mean over physical quantities of

$$\sqrt{\frac{\|\varepsilon^a\|^2 - \|\varepsilon^b\|^2}{\|\varepsilon^b\|^2}} \tag{18}$$

where $\varepsilon^b$ and $\varepsilon^a$ are background and analysis error respectively.

In this more demanding setup, CERA does not yield the same result as its FCM counterpart. Increasing the number of outer iterations (not shown) actually degrades the CERA results, hinting that the convergence requirement are not met and that the model used in the inner loop is too different from the outer one. This is aggravated in CERA-F where the outer model goes to SWR-convergence. This lack of outer convergence precludes CERA from fully benefiting from outer coupling

On the other hand, accounting for the coupling during the assimilation process (through PCM or WCM) allows to get a reasonably good analysis, both in term of coupling and RMSE.

| Algorithm | $\gamma$ | $k_{\max}$ | # of minimisation iterations | Computing cost (relative to CERA) | Interface imbalance indicator | RMSE improvement (in %) |
|---|---|---|---|---|---|---|
| FCM-F | – | $k_{\text{cvg}}$ | 26 | 3.8 | $2.10^{-12}$ | 74 |
| FCM-1 | – | 1 | 28 | 1.06 | 5 | 65 |
| CERA-F | – | $k_{\text{cvg}}$ | 24 | 1.1 | $5.810^{-12}$ | 24 |
| CERA-1 | – | 1 | 26 | 1 | 1.6 | 40 |
| PCM-1 | 0.1 | 1 | 25 | 0.96 | $4.10^{-3}$ | 60 |
| WCM | 0.1 | 0 | 31 | 1.2 | $6.10^{-3}$ | 57 |

Table 5: Result summary for the SCM system

Differences are largely located at the limit of the boundary layers that CERA and, to a lesser extent, WCM tend to misplace. Figure 5 shows the difference in the analysed state between FCM-F and CERA-1, FCM-F and PCM, and FCM-F and WCM for atmospheric temperature and v-velocities along the assimilation window. For temperatures, even if CERA benefits from its less constrained optimisation problem to get a better temperature at the beginning of the time window, not accounting enough for the coupling processes quickly degrades the analysis. PCM and WCM manage a better estimation of the ABL height and are close to the FCM one throughout the assimilation window. One can notice that as for CERA, PCM benefit from more degrees of freedom and improve the analysis compared to FCM at the beginning of the window, and thanks to its stronger coupling, it manages to retain a good analysis.

Differences are less striking on ocean quantities (not shown) and actually CERA-1, with a 40% improvement over background is doing a reasonable job. Longer term forecast can also be affected by differences in the initial conditions. Figure 6 shows the evolution of sea surface velocities for four days, starting from the end of the assimilation window, using the same model coupling (asynchronous). All

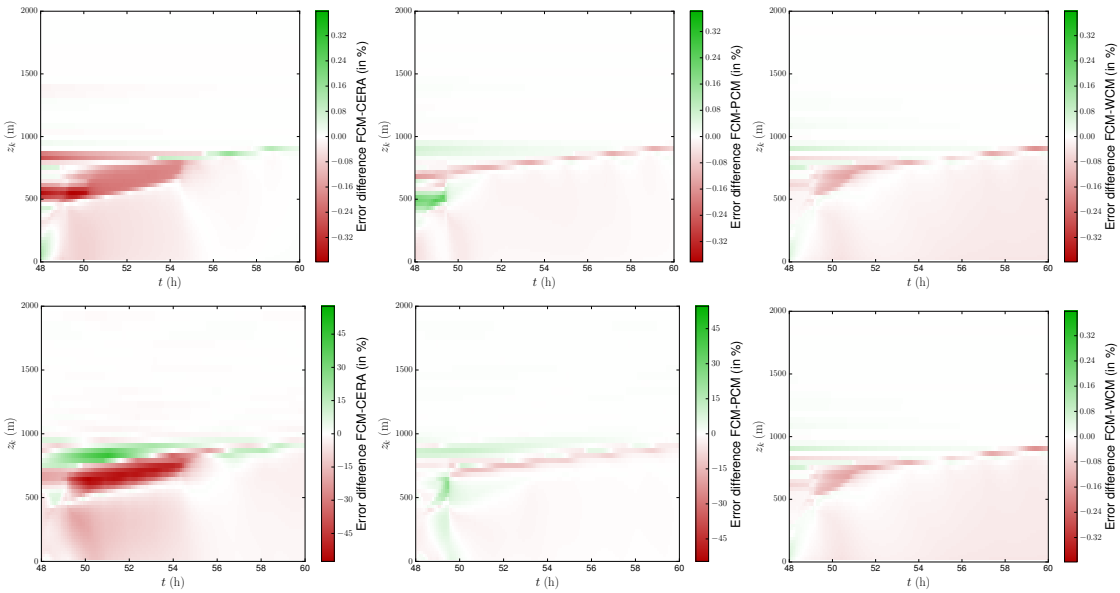



Figure 5: Differences in the analysed state between FCM-F and CERA-1 (left panels), FCM-F and PCM (middle panels), and FCM-F and WCM (right panels) for atmospheric temperature (top) and v-velocities (bottom) along the assimilation window

three systems improve significantly from the background and get the right variability even though forecast from CERA analysis has a tendency to overshoot.

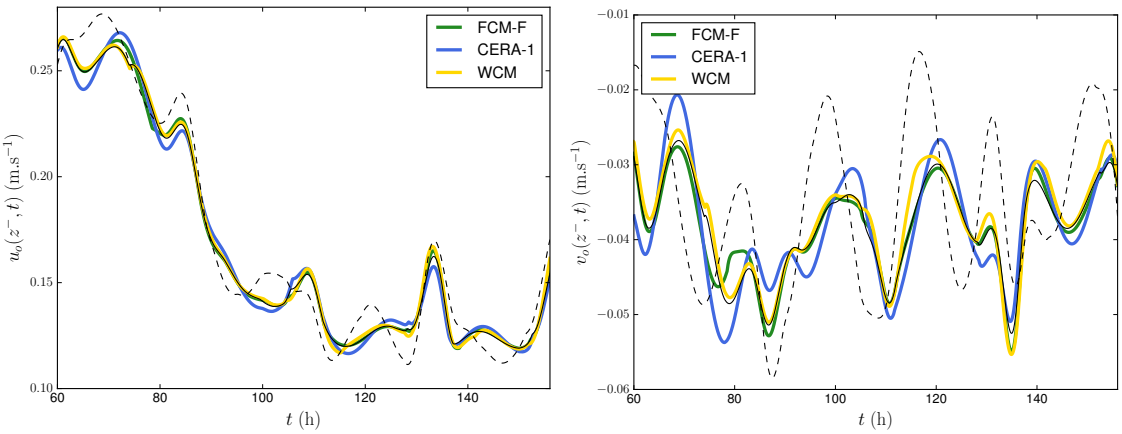



Figure 6: Forecast of SSU and SSV from FCM-F, CERA-1 and WCM analysis. Dashed and plain black lines are background and truth evolutions respectively

With a 20% increase in computing cost WCM brings a noticeable improvement over CERA, without most of the complexity of an FCM or PCM scheme. It therefore and may be a good candidate for an intermediary step toward fully coupled data assimilation. However this optimistic conclusion needs to be moderated. First Note that in this study, inner problems in CERA are solved as a single optimisation problem, unlike realistic application where ocean and atmosphere are solved separately. Although it does not change the result at convergence, doing so may affect the convergence speed of the inner prob-

lems and therefore increase artificially CERA's computing cost. Also experience showed that WCM was more difficult to properly tune and required much more trial and error to get it working efficiently than the other algorithms. On a simplified model it is doable, in order to go toward more complex problems, one will need to devise an efficient tuning strategy.

## 7 Software developments

Both linear and non linear coupled models used in this report have been developed and interfaced within the OOPS environment. The coupled diffusion model is made of 9 000 lines of C++ including hand-written tangent and adjoint models and interfaces while the coupled SCM is made of 70 000 lines of Fortran/C++ including tangent and adjoint and interfaces. SCM tangent and adjoint models were generated using the automatic differentiation tool TAPENADE[1]. It will then be made available to the community and be proposed as a reference test case to be included in OOPS.

## 8 Conclusion and perspectives

Ocean-Atmosphere coupling is a complicated matter and is still a somewhat open question. In this report we mostly mentioned the time inconsistency in the fluxes exchanged by ocean and atmosphere, one can also add the highly parameterised nature of such exchanges and the uncertainties that are associated. In both cases coupled data assimilation is a challenge but also an opportunity. First by improving the coupling consistency of the analysis, but also, going beyond the point of this report, by providing materials for improving coupled modelling. The latter can be achieve by studying the coupled increment statistics, for instance, provided that the coupling is accounted for in the data assimilation scheme.

In this task we compare several possibilities to improve the coupling quality through variational data assimilation. Three algorithms are presented and their convergence properties discussed in appendix C. They are applied first to a simple linear diffusion coupled problem and then to a more realistic coupled single column model. The coupled SCM is thoroughly described in appendices A and B and is itself a result of the task, since it can be used as a testbed for future research in coupled data assimilation.

One can draw a few conclusions from this preliminary study. First, CERA was probably a reasonable choice, with a good trade-off between low complexity and efficiency. Indeed, if both ocean and atmosphere data assimilation systems are available, its implementation is relatively easy (as operational implementation permits) and it is able, thanks to its outer loops, to account for part of the coupling. However, this is limited by its lack of global convergence. The latter being also observed on CERA applied to operational OA system, where nothing is gained beyond a couple of outer iterations (but much is gained by the second outer iteration). Second, accounting explicitly for the coupling, either as a strong or a weak constraint of the variational data assimilation problem, can be of benefit. Indeed, it improves the overall quality of the analysis and the global (outer loop) convergence of the system. As a side product it gives a feedback on the coupling processes, that can be highly valuable to improve the system, as mentioned above. As a downside, it increases the complexity of the system, hence its development time. Moreover it potentially damages the local (inner loop) convergence speed, hence its computing time.

The real conclusion of this report is that ocean and atmosphere coupled modelling and coupled data assimilation are both a complicated matter, and one should look into both problems at the same time in a consistent manner.

[1] http://www-sop.inria.fr/tropics/tapenade.html

# A Single column coupled model description

This appendix describes a single column coupled ocean-atmosphere model. It is a quite valuable tool since it mimics the complexity of model used in a forecasting center without the tremendous computational burden. Single column models are also widely used to study sub-mesh parameterisations in the ocean-atmosphere surface boundary layer without the cumbersome 3D model. These sub-mesh parameterisations introduce non-differentiabilities in addition to strong non-linearities. These non-differentiability and non-linearities can cause several problems in a process of data assimilation and the development of an adjoint model (Janisková, Veersé, Thépaut, Desroziers & Pouponneau 1999, Janisková, Thépaut & Geleyn 1999), which is why it will be very interesting to use such a model to test the robustness of our algorithms introduced in section 3.2.

## A.1 Model description

### A.1.1 From 3D primitive equations to single column

Equations of our simple column model derive from 3D primitive equations used in atmosphere or ocean operational forecasting models. These are the Navier-Stokes equations describing the motion of a Newtonian fluid simplified by various hypotheses themselves specific to the case of atmosphere or ocean:

- Fluid is in hydrostatic equilibrium, which consists in neglecting the vertical acceleration of the fluid in equations of vertical motion.
- One neglects the vertical component of Coriolis acceleration.
- The thin-layer hypothesis is carried out, that is to say that the fluid is contained in a very thin layer relative to the radius of the sphere. In practice this amounts to neglect altitude with respect to the earth radius.

These equations are thoroughly studied in the literature (see McWilliams (2006) for example)

#### A.1.1.a Single-column hypothesis and Reynolds decomposition

In order to study the behaviour of the models at the air-sea interface with sub-mesh parameterisations, it is possible to make some additional simplifying hypotheses. These hypotheses allow us to reduce to a simpler case study, in 1D, while keeping the complexity of modelling parametrisation at the interface. We will therefore have a somewhat representative approximation of a *realistic* model based on primitive 3D equations with the simplicity and low numerical cost of a 1D model. Such a model is obviously not meant for weather forecasting, but still behaves in a similar way than that of a 3D realistic model. Therefore, we can hope that the results obtained will be transposable to 3D models.

The first hypothesis is to consider horizontal homogeneity:

- $\frac{\partial \cdot}{\partial x} = \frac{\partial \cdot}{\partial y} = 0$

From there we can simplify the 3D primitive equations, leading to:

$$
\begin{aligned}
&\text{(Momentum equation)} && \frac{\partial \mathbf{u}}{\partial t} = -f\mathbf{k}\times\mathbf{u} + \frac{\partial w\mathbf{u}}{\partial z} + \nu_m \nabla^2 \mathbf{u} + \mathscr{F}_{\mathbf{u}} \\
&\text{(Conservation equation)} && \frac{\partial w}{\partial z} = 0 \\
&\text{(Tracers equation)} && \frac{\partial \mathscr{T}}{\partial t} = \frac{\partial w\mathscr{T}}{\partial z} + \nu_s \nabla^2 \mathscr{T} + \mathscr{F}_{\mathscr{T}}
\end{aligned}
\tag{19}
$$

Where $\mathbf{u} = (u, v)^T$ and $w$ are zonal, meridional and vertical wind or sea current velocity components, $f$ is the Coriolis factor and , $\nu_m$ and $\nu_s$ are molecular viscosity and diffusion coefficients respectively. $\mathscr{T}$ represents the various tracers considered in the atmosphere and the ocean (air humidity $q_a$, air and sea potential temperature $\theta_\beta$ ($\beta = a, o$) and sea salinity $S_o$). Finally $\mathscr{F}_{\mathbf{u},\mathscr{T}}$ terms represent the sources and relaxation terms.

Applying the Reynolds decomposition to the above equations, one gets:

$$
\begin{aligned}
\frac{\partial \langle \mathbf{u} \rangle}{\partial t} &= -f\mathbf{k}\times\langle \mathbf{u} \rangle + \frac{\partial \langle w \rangle \langle \mathbf{u} \rangle}{\partial z} + \frac{\partial \langle w'\mathbf{u}' \rangle}{\partial z} + \nu_m \nabla^2 \langle \mathbf{u} \rangle + \mathscr{F}_{\mathbf{u}} \\
\frac{\partial \langle \mathscr{T} \rangle}{\partial t} &= \frac{\partial \langle w \rangle \langle \mathscr{T} \rangle}{\partial z} + \frac{\partial \langle w'\mathscr{T}' \rangle}{\partial z} + \nu_s \nabla^2 \langle \mathscr{T} \rangle + \mathscr{F}_{\mathscr{T}} \\
\frac{\partial \langle w \rangle}{\partial z} &= 0
\end{aligned}
\tag{20}
$$

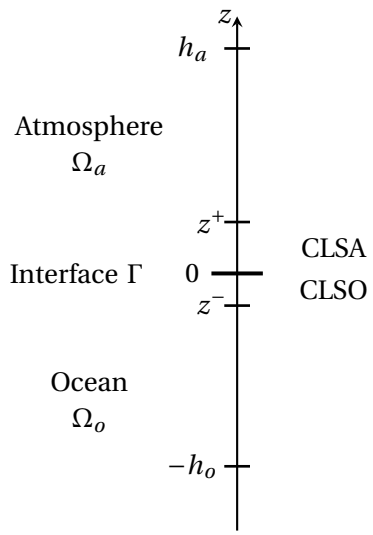
A second hypothesis is to neglect the vertical advection, *i.e.* $\langle w \rangle \approx 0$ allowing for a further simplification. The unresolved sub-grid variations $\langle w'\mathbf{u}' \rangle$ and $\langle w'\mathscr{T}' \rangle$ are represented through the turbulent closure scheme:

$$
\begin{aligned}
\langle w'\mathbf{u}' \rangle &= K_m \frac{\partial \langle \mathbf{u} \rangle}{\partial z} \\
\langle w'\mathscr{T}' \rangle &= K_s \frac{\partial \langle \mathscr{T} \rangle}{\partial z}
\end{aligned}
\tag{21}
$$

where $K_m$ is the turbulent viscosity coefficient and $K_s$ the turbulent diffusivity coefficient.

#### A.1.1.b Single column model equations

A 1D single column model can then be derived from the above mentioned hypotheses. First by defining the vertical domain of our model (figure opposite). The atmosphere is defined on domain $\Omega_a = [0, h_a]$ and the ocean on $\Omega_o = [-h_o, 0]$. The surface boundary layer is located between $z^+$ in the atmosphere and $z^-$ in the ocean. Ocean atmosphere interface being represented by $\Gamma$. The time domain is $[0, T]$ with $T > 0$s. We can then express the equations of our 1D single column model:



$$
\boxed{\begin{aligned}
\frac{\partial \mathbf{u}_\beta(z,t)}{\partial t} &= -f\mathbf{k}\times\mathbf{u}_\beta(z,t) + \frac{\partial}{\partial z}\left(K_m^\beta(z)\frac{\partial \mathbf{u}_\beta(z,t)}{\partial z}\right) + \mathscr{F}_{\mathbf{u}_\beta}(z,t) && \text{on } \Omega_\beta\times[0,T]\\
\frac{\partial \mathscr{T}_\beta(z,t)}{\partial t} &= \frac{\partial}{\partial z}\left(K_s^\beta(z)\frac{\partial \mathscr{T}_\beta(z,t)}{\partial z}\right) + \mathscr{F}_{\mathscr{T}_\beta}(z,t) && \text{on } \Omega_\beta\times[0,T]
\end{aligned}} \tag{22}
$$

with $\beta = a$ or $o$ designating atmospheric or oceanic variables. Thus, both atmosphere and ocean model have the same structure and differences will come from RHS choices, interface conditions and computation of viscosity and turbulent diffusivity coefficients.

Molecular viscosity and diffusivity terms vanished because they are negligible compared to turbulent terms within the surface boundary layer. Outside the surface layer, turbulent coefficients can be considered as being constant and equal to molecular coefficients and representing the free atmosphere or the deep ocean properly. Note also that only resolved terms $\langle \mathbf{u}_{a,o}\rangle$ and $\langle \mathscr{T}_{a,o}\rangle$ are present in the equation, so $\langle . \rangle$ are omitted in 22 for the sake of clarity.

In our model, we use the following forcings $\mathscr{F}_{\mathbf{u}_{a,o}}(z,t)$ and $\mathscr{F}_{\mathscr{T}_{a,o}}(z,t)$:

$$
\begin{aligned}
&\forall (z,t)\in\Omega_{a,o}\times[0,T]\ : \\
&\mathscr{F}_{\mathbf{u}_a}(z,t) = -f\mathbf{k}\times\mathbf{u}_{aG}(z,t) && \text{relaxation toward geostrophic winds } \mathbf{u}_{aG}(z,t)\\
&\mathscr{F}_{\mathbf{u}_o}(z,t) = 0 \\
&\mathscr{F}_{\theta_a}(z,t) = \lambda_S(z)\left(\theta_a(z,t) - \theta_{LS}(z,t)\right) && \text{large scale relaxation toward } \theta_{LS}(z,t)\\
&\mathscr{F}_{\theta_o}(z,t) = \partial_z\left(\frac{Q_S(z,t)}{\rho_o C_p^o}\right) && \text{penetrating solar fluxes } Q_S(z,t)\\
&\mathscr{F}_{q_a}(z,t) = \lambda_S(z)\left(q_a(z,t) - q_{LS}(z,t)\right) && \text{large scale relaxation toward } q_{LS}(z,t)\\
&\mathscr{F}_{S_o}(z,t) = 0
\end{aligned} \tag{23}
$$

with $1/\lambda_S(z)$ a altitude dependent relaxation time allowing to define an atmospheric large scale nudging term and $C_p^o$ water thermal capacity. $Q_S(z,t)$ is the solar flux penetrating into the ocean, computed from the surface solar flux $Q_{S_0} = Q_S(0,t)$. Underscore $_{LS}$ stands for Large Scales, these quantities being defined externally.

The model also require conditions at $z = h_{\max}$ and $z = -p_{\max}$, external boundaries of the domain, denoted $\partial\Omega_{a,o}^{\text{ext}}$ :

$$
\begin{aligned}
\frac{\partial \mathbf{u}_\beta}{\partial z}(t) &= 0 && \text{on } \partial\Omega_\beta^{\text{ext}}\\
\frac{\partial \mathscr{T}_\beta}{\partial z}(t) &= F_{\mathscr{T}_\beta} && \text{on } \partial\Omega_\beta^{\text{ext}}
\end{aligned} \tag{24}
$$

where tracers fluxes $F_{\mathscr{T}_\beta}$ are chosen as:

$$
\begin{aligned}
F_{\theta_a} &= 0 & F_{\theta_o} &= \nu_s\cdot\left.\frac{\partial\theta_o(z,0)}{\partial z}\right|_{z=-ho}\\
F_{q_a} &= 0 & F_{S_o} &= \nu_s\cdot\left.\frac{\partial S_o(z,0)}{\partial z}\right|_{z=-ho}
\end{aligned} \tag{25}
$$

## A.2 Interface boundary conditions and air-sea coupling

These conditions at ocean-atmosphere interface must make it possible to verify the continuity of momentum and tracers fluxes. Both models are therefore forced at interface $\Gamma$ by different fluxes calculated from atmospheric and oceanic components, which allows for a coupling between the two media.

### A.2.1 Momentum flux continuity

Recall that atmosphere and ocean velocities follow:

$$\begin{aligned} \frac{\partial \mathbf{u}_a}{\partial t} &= -f\mathbf{k}\times(\mathbf{u}_a - \mathbf{u}_{aG}) + \frac{\partial}{\partial z}\left(K_m^a \partial_z \mathbf{u}_a\right) \\ \frac{\partial \mathbf{u}_o}{\partial t} &= -f\mathbf{k}\times\mathbf{u}_o + \frac{\partial}{\partial z}\left(K_m^o \partial_z \mathbf{u}_o\right) \end{aligned} \tag{26}$$

Interface conditions on $\mathbf{u}_a$ and $\mathbf{u}_o$ are given by:

$$\rho_a K_m^a \frac{\partial \mathbf{u}_a}{\partial z} = \rho_o K_m^o \frac{\partial \mathbf{u}_o}{\partial z} = \boldsymbol{\tau} \quad \text{on } \Gamma\times[0,T] \tag{27}$$

where $\boldsymbol{\tau}$ is the momentum flux, defined by the bulk formula:

$$\boldsymbol{\tau} = \rho_a C_d \parallel \mathbf{u}_a(z^+) - \mathbf{u}_o(z^-) \parallel \left(\mathbf{u}_a(z^+) - \mathbf{u}_o(z^-)\right) \tag{28}$$

with $C_d$ the bulk drag coefficient.

### A.2.2 Heat fluxes

In the ocean, potential temperature $\theta_o$ in (22) and (23) is defined by:

$$\frac{\partial \theta_o}{\partial t} = \frac{\partial}{\partial z}\left(K_s^o \frac{\partial \theta_o}{\partial z} + \frac{Q_S}{\rho_o C_p}\right) \tag{29}$$

At ocean surface, flux continuity imposes:

$$K_s^o \frac{\partial \theta_o}{\partial z} + \frac{Q_{S_0}}{\rho_o C_p^o} = \frac{Q_{\text{net}}}{\rho_o C_p^o} \quad \text{on } \Gamma\times[0,T] \tag{30}$$

where $Q_{\text{net}}$ is the net heat flux, defined as:

$$Q_{\text{net}} = Q_{S_0} + Q_{L\downarrow} + Q_{L\uparrow} + Q_E + Q_H \tag{31}$$

with:

- $Q_{S_0}$ surface solar flux (prescribed)
- $Q_{L\downarrow}(t)$ long wave radiative input normally coming from the atmosphere model, but prescribed here.
- $Q_{L\uparrow}$ long wave radiative output. Considering the ocean as a black body, $Q_{L\uparrow}$ is given by the Stephen-Boltzmann law: $Q_{L\uparrow} = -\epsilon\sigma\theta_o(z^-)^4$ with $\epsilon = 1$ and $\sigma = 5.67\cdot 10^{-8}\text{W.m}^{-2}.\text{K}^{-4}$.
- $Q_H$ and $Q_E$ sensible and latent heat fluxes (defined later)

Regarding the atmosphere, temperature satisfies:

$$\frac{\partial \theta_a}{\partial t} = \frac{\partial}{\partial z}\left(K_s^a \frac{\partial \theta_a}{\partial z}\right) + \lambda_S \left(\theta_a - \theta_{LS}\right) \tag{32}$$

At the interface:

$$K_s^a \frac{\partial \theta_a}{\partial z} = \frac{Q_H}{\rho_a C_p^a} \quad \text{on } \Gamma \times [0, T] \tag{33}$$

with $C_p^a$ the thermal capacity and $Q_H$ the sensible heat flux defined by:

$$Q_H = \rho_a C_p^a C_H \parallel \mathbf{u}_a(z^+)) - \mathbf{u}_o(z^-)) \parallel \left(\theta_a(z^+)) - \theta_o(z^-))\right) \tag{34}$$

with $C_H$ the bulk sensible heat transfer coefficient.

Note that we made an additional simplification: the atmosphere is supposed to be transparent, *i.e* it does not absorb radiative fluxes, so they do not directly impact air temperature.

### A.2.3 Evaporation

Air humidity $q_a$ is defined by

$$\frac{\partial q_a}{\partial t} = \frac{\partial}{\partial z}\left(K_s^a \frac{\partial q_a}{\partial z}\right) + \lambda_S \left(q_a - q_{LS}\right) \tag{35}$$

The interface condition is:

$$K_s^a \frac{\partial q_a}{\partial z} = \frac{-E}{\rho_a} \quad \text{on } \Gamma \times [0, T] \tag{36}$$

where $E$ is the evaporation flux defined by bulk formula:

$$E = \rho_a C_E \parallel \mathbf{u}_a(z^+) - \mathbf{u}_o(z^-) \parallel \left(q_a(z^+) - q_o(z^-)\right) \tag{37}$$

with $C_E$ an exchange coefficient and $q_o(z^-)$ the humidity at ocean surface. Assuming the air is saturated with humidity at ocean surface, $q_o(z^-)$ is estimated from sea surface temperature by:

$$q_o(z^-) = q_{\text{sat}} = \frac{0.98}{\rho_o} 640380 \exp\left(-\frac{5107.4}{\theta_o(z^-)}\right) \qquad \text{(Large 2006)} \tag{38}$$

Evaporation flux $E$ allows the definition of latent heat flux $Q_E$ in (31) as:

$$Q_E = \Lambda E \tag{39}$$

with $\Lambda = 2.5 \times 10^6$ J.kg$^{-1}$.

### A.2.4 Fresh water fluxes

Lastly, one needs to define interface conditions for salinity $S_o$, the latter being described by:

$$\frac{\partial S_o}{\partial t} = \frac{\partial}{\partial z}\left(K_s^o \frac{\partial S_o}{\partial z}\right) \tag{40}$$

The interface condition reads:

$$K_s^o \partial_z S_o = F \tag{41}$$

where $F$ is the fresh water flux, so that:

$$F = 0.001(E-P)S_o(z^-) \tag{42}$$

with $P$ the precipitation and $E$ the evaporation (given by equation 37). Constant 0.001 being needed to get the right unit.

#### A.2.5 Summary of interface fluxes

$$\begin{aligned}
\text{Momentum flux} \quad & \boldsymbol{\tau} = \rho_a C_d \parallel \mathbf{u}_a(z^+) - \mathbf{u}_o(z^-) \parallel \left(\mathbf{u}_a(z^+) - \mathbf{u}_o(z^-)\right) \\
\text{Sensible heat flux} \quad & Q_H = \rho_a C_p^a \parallel \mathbf{u}_a(z^+)) - \mathbf{u}_o(z^-)) \parallel \left(\theta_a(z^+)) - \theta_o(z^-))\right) \\
\text{Evaporation} \quad & E = \rho_a C_E \parallel \mathbf{u}_a(z^+) - \mathbf{u}_o(z^-) \parallel \left(q_a(z^+) - q_{\text{sat}}\right) \\
\text{Latent heat flux} \quad & Q_E = \Lambda E \\
\text{Fresh water flux} \quad & F = 0.001(E-P)S_o(z^-) \\
\text{Solar radiative flux} \quad & Q_{S_0} \\
\text{radiative fluxes} \quad & Q_{L\uparrow}, Q_{L\downarrow} \\
\text{Net heat flux} \quad & Q_{\text{net}} = (1-\beta)Q_{S_0} + Q_{L\downarrow} + Q_{L\uparrow} + Q_E + Q_H
\end{aligned}$$

### A.3 Summary of continuous equations

The single column model equations can be summarised as:

$$\begin{aligned}
&\underbrace{\begin{aligned}
\frac{\partial \mathbf{u}_a}{\partial t} &= -f\mathbf{k}\times(\mathbf{u}_a - \mathbf{u}_{aG}) + \frac{\partial}{\partial z}\left(K_m^a \frac{\partial \mathbf{u}_a}{\partial z}\right) \\
\frac{\partial \theta_a}{\partial t} &= \frac{\partial}{\partial z}\left(K_s^a \frac{\partial \theta_a}{\partial z}\right) + \lambda_S\left(\theta_a - \theta_{LS}\right) \\
\frac{\partial q_a}{\partial t} &= \frac{\partial}{\partial z}\left(K_s^a \frac{\partial q_a}{\partial z}\right) + \lambda_S\left(q_a - q_{LS}\right) \\
K_m^a \frac{\partial \mathbf{u}_a}{\partial z} &= \frac{\boldsymbol{\tau}}{\rho_a} \\
K_s^a \frac{\partial \theta_a}{\partial z} &= \frac{Q_H}{\rho_a C_p^a} \\
K_s^a \frac{\partial q_a}{\partial z} &= -\frac{E}{\rho_a} \\
\frac{\partial \mathbf{u}_a}{\partial z} &= 0 \\
\frac{\partial \theta_a}{\partial z} &= 0 \\
\frac{\partial q_a}{\partial z} &= 0
\end{aligned}}_{\text{Atmosphere model}}
\quad
\underbrace{\begin{aligned}
\frac{\partial \mathbf{u}_o}{\partial t} &= -f\mathbf{k}\times\mathbf{u}_o + \frac{\partial}{\partial z}\left(K_m^o \frac{\partial \mathbf{u}_o}{\partial z}\right) && \text{on } \Omega_\beta \times [0,T] \\
\frac{\partial \theta_o}{\partial t} &= \frac{\partial}{\partial z}\left(K_s^o \frac{\partial \theta_o}{\partial z} + \frac{Q_S}{\rho_o C_p^o}\right) && \text{on } \Omega_\beta \times [0,T] \\
\frac{\partial S_o}{\partial t} &= \frac{\partial}{\partial z}\left(K_s^o \frac{\partial S_o}{\partial z}\right) && \text{on } \Omega_\beta \times [0,T] \\
K_m^o \frac{\partial \mathbf{u}_o}{\partial z} &= \frac{\boldsymbol{\tau}}{\rho_o} && \text{on } \Gamma \times [0,T] \\
K_s^o \frac{\partial \theta_o}{\partial z} + \frac{Q_s}{\rho_o C_p^o} &= \frac{Q_{\text{net}}}{\rho_o C_p^o} && \text{on } \Gamma \times [0,T] \\
K_s^o \frac{\partial S_o}{\partial z} &= 0.001\,(E-P)\,S_o(z^-) && \text{on } \Gamma \times [0,T] \\
\frac{\partial \mathbf{u}_o}{\partial z} &= 0 && \text{on } \partial\Omega_\beta^{\text{ext}} \times [0,T] \\
\frac{\partial \theta_o}{\partial z} &= \nu_s \cdot \left.\frac{\partial \theta_o(z,0)}{\partial z}\right|_{z=-ho} && \text{on } \partial\Omega_\beta^{\text{ext}} \times [0,T] \\
\frac{\partial S_o}{\partial z} &= \nu_s \cdot \left.\frac{\partial S_o(z,0)}{\partial z}\right|_{z=-ho} && \text{on } \partial\Omega_\beta^{\text{ext}} \times [0,T]
\end{aligned}}_{\text{Ocean model}}
\end{aligned} \tag{43}$$

- $K_s^{a,o}$ and $K_m^{a,o}$ are estimated through turbulent closure schemes TKE (Cuxart et al. 2000) and KPP (Large et al. 1994) for tracers and momentum respectively
- $\boldsymbol{\tau}$, $Q_{\text{net}}$ and $F = E - P$ are estimated from bulk formulae

- Ocean atmosphere coupling at the interface is done through turbulent fluxes

Even though this is simplified model compared to the complete 3D primitive equations, it allows to study two major problems of the operational models, namely the computation of the interface flows via the bulk formulations and the evaluation of the turbulence coefficient in the surface boundary layer via complex parametrizations. It can be a valuable tool for both the study of coupling processes and for evaluation of coupled data assimilation schemes

## A.4 Discretisation

Equations 43 are solved numerically through a second order finite difference scheme in space and a Euler backward scheme in time. One exception being the Coriolis term in the atmospheric moment equation (first equation in system 43), where a forward backward scheme is used.

### A.4.1 Vertical grid

Vertical discretisation is similar for ocean and atmosphere with higher resolution near the interface. For both domain $\Omega_\beta = [0, h_\beta]$ where $\beta = a$ or $\beta = o$. Both vertical grids are defined through 4 parameters:

- $h_\beta$ altitude of top grid level
- $N_{L\beta}$ number of vertical levels
- $\theta_{s\beta}$ stretching coefficient: the more $\theta_{s\beta}$ the more the grid will be refined close to the interface $\Gamma$.
- $h_{c\beta}$ height of transition between uniform (for $z < h_{ca}$ and $z > h_{co}$) and non uniform grid (for $z > h_{ca}$, $z < h_{co}$)

from these parameters one can define the altitude/depth of cells centres $z_k$ and interfaces $z_{k+\frac{1}{2}}$ for each level $k$.

$$\begin{aligned} &\text{(Centres)} && z_k = h_{c\beta}\sigma_k^\beta + (h_\beta - h_{c\beta})\frac{\sinh \sigma_k \theta_{s\beta}}{\sinh \theta_{sa}} && \text{with } k \in [\![1, N_{L\beta}]\!] \\ &\text{(Interfaces)} && z_{k+1/2} = h_{c\beta}\sigma_{k+1/2}^\beta + (h_\beta - h_{c\beta})\frac{\sinh \sigma_{k+1/2} \theta_{s\beta}}{\sinh \theta_{s\beta}} && \text{with } k \in [\![0, N_{L\beta}]\!] \end{aligned} \tag{44}$$

where $\sigma_k^a = \frac{k-1/2}{N_{La}}$ and $\sigma_k^o = \frac{k-1/2-N_{Lo}}{N_{Lo}}$. Numerical values used in this report are summarised in table 6.

| Description | Atmosphere coefficients | Ocean coefficients |
|---|---|---|
| Number of vertical levels | $N_{La} = 51$ | $N_{Lo} = 50$ |
| top/bottom altitude/depth | $h_a = 2000\,\mathrm{m}$ | $h_o = 500\,\mathrm{m}$ |
| Transition altitude/depth | $h_{ca} = 200\,\mathrm{m}$ | $h_{co} = 50\,\mathrm{m}$ |
| Stretching coefficients | $\theta_{sa} = 2$ | $\theta_{so} = 6.5$ |

Table 6: vertical grids coefficients for both models

### A.4.2 Time discretisation and coupling strategy

For the sake of simplicity, both models use the same time steps ($\Delta t_a = \Delta t_o = \Delta t = 30$ s). This is probably not an optimal choice for a testbed, since it avoid one important difficulty in air-sea coupling. It should be revised in future versions of the code.

In order to describe the coupling between ocean and atmosphere, let $\mathbf{x}_a = \left(\mathbf{u}_a, \theta_a, q_a\right)^T$ and $\mathbf{x}_o = (\mathbf{u}_o, \theta_o, S_o)^T$ the atmosphere and ocean states. Equations (43) can be represented by:

$$\begin{array}{lll} \mathscr{M}_a(\mathbf{x}_a) & \mathscr{M}_o(\mathbf{x}_o) & \text{on } \Omega_\beta \times [0,T] \\ \mathscr{C}_a(\mathbf{x}_a) = \mathscr{F}_{oa} & \mathscr{C}_o(\mathbf{x}_o) = \mathscr{F}_{oa} & \text{on } \Gamma \times [0,T] \\ \dfrac{\partial \mathbf{x}_a}{\partial z} = \mathbf{B}_a & \dfrac{\partial \mathbf{x}_o}{\partial z} = \mathbf{B}_o & \text{on } \partial\Omega_\beta^{\text{ext}} \times [0,T] \end{array} \tag{45}$$

Implicit formulation in time implies that a simple exchange of fluxes between domain is not enough to find a consistent solution to equation 45. So unless the implicit problem is solved all at once for both model altogether, which is impracticable for realistic applications, one needs to use an iterative algorithme such as the Schwarz Waveform Relaxation described in section 2. Here we chose the multiplicative form this algorithm:

$$\begin{array}{lll} \mathscr{M}_a(\mathbf{x}_a^k) & \mathscr{M}_o(\mathbf{x}_o^k) & \text{on } \Omega_{a,o} \times [0,T] \\ \mathscr{C}_a(\mathbf{x}_a^k) = \mathscr{F}_{oa}^{k-1} & \mathscr{C}_o(\mathbf{x}_o^k) = \mathscr{F}_{oa}^{k} & \text{on } \Gamma \times [0,T] \\ \dfrac{\partial \mathbf{x}_a^k}{\partial z} = \mathbf{B}_a & \dfrac{\partial \mathbf{x}_o^k}{\partial z} = \mathbf{B}_o & \text{on } \partial\Omega_{a,o}^{\text{ext}} \times [0,T] \end{array} \tag{46}$$

where $k$ represents the SWR iterations, that runs until the convergence criterion is met:

$$\| \mathscr{F}_{oa}^{k+1} - \mathscr{F}_{oa}^{k} \|_2 < \epsilon$$

It can be transformed into a more classical asynchronous coupling by performing only one iteration.

## A.5 Numerical implementation

The coupled SCM is made of 70 000 lines of Fortran/C++ including tangent and adjoint and interfaces. SCM tangent and adjoint models were generated using the automatic differentiation tool TAPENADE[2]. It has been interfaced with OOPS for data assimilation use and will then be made available to the community and be proposed as a reference test case to be included in OOPS. For now, it is considered as a single model by OOPS, which is suboptimal and will be revised. However it has allowed to implement all assimilation algorithm presented in this report.

# B Single column model reference test case

This appendix describes experimental settings that are used in section 6. The single column model presented in appendix A is set up so that it mimics the behaviour of a mid-latitude air-sea column (coriolis parameter $f = 10^{-4}$ s$^{-1}$).

[2] http://www-sop.inria.fr/tropics/tapenade.html

## B.1 Atmosphere initial conditions

For the atmosphere, initial conditions are

$$\begin{aligned}
&u_a(z,t=0) = u_{a0} = 15\,\mathrm{m\cdot s^{-1}} && z\in\Omega_a \\
&v_a(z,t=0) = v_{a0} = 3\,\mathrm{m\cdot s^{-1}} && z\in\Omega_a \\
&\theta_a(z,t=0) = \theta_{a0} = \theta^{\mathrm{ref}} + \frac{N^2\theta^{\mathrm{ref}}}{g} z\ \mathrm{K} && z\in\Omega_a, \quad \text{with } N = 0.01\mathrm{s}^{-1} \text{ and } \theta^{\mathrm{ref}} = 286\,\mathrm{K} \\
&q_a(z,t=0) = q_{a0} = 0.01\,\mathrm{kg/kg} && z\in\Omega_a
\end{aligned} \tag{47}$$

where $N$ is the Brunt-Väisälä frequency and $g$ is the acceleration of gravity constant.

While for the ocean column, one sets up

$$\begin{aligned}
&u_o(z,t=0) = 0\,\mathrm{m\cdot s^{-1}} && z\in\Omega_o \\
&v_o(z,t=0) = 0\,\mathrm{m\cdot s^{-1}} && z\in\Omega_o \\
&\theta_o(z,t=0) = 5 + 1.475\cdot 10^{-3}\cdot z + 5\cdot\exp\left(\frac{z}{100}\right) + 5\cdot\exp\left(\frac{z}{500}\right)\ \mathrm{K} && z\in\Omega_o \\
&S_o(z,t=0) = S = 35 - \left(1.75\cdot 10^{-4}\cdot z + 0.8\cdot\exp\left(\frac{z}{200}\right) + 0.2\cdot\exp\left(\frac{z}{400}\right)\right)\ \mathrm{psu} && z\in\Omega_a
\end{aligned} \tag{48}$$

The system then undergoes a 10-day spin-up, which leads to reference initial vertical profiles showed in figure (7) for both atmosphere (top) and ocean (bottom) variables.

## B.2 Large scale relaxation

The atmosphere model includes some large scale relaxation terms $\mathbf{u}_{aG}$, $\theta_{LS}$ and $q_{LS}$ (see equations 43). They are defined so that the system shows enough variability, in particular so that it switches from stable to unstable regimes on a short enough time window. In addition, one would like to observe a diurnal cycle for the temperature.

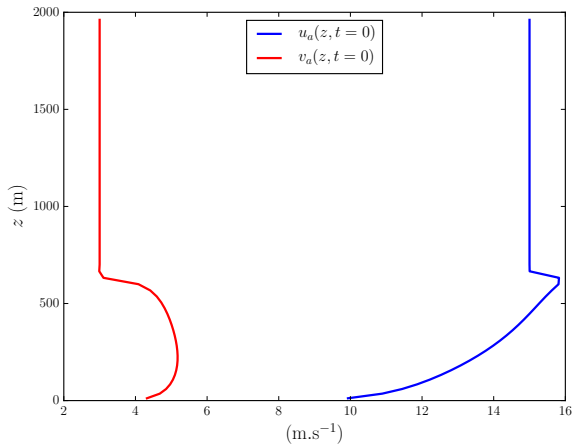
This is achieved setting:

$$\begin{aligned}
&u_{aG}(z,t) = \begin{cases} u_{a0} + U_{up}\cdot l(z,t) & \text{si } z \le h_{\mathrm{lim}} \\ u_{a0} & \text{si } z > h_{\mathrm{lim}} \end{cases} && z\in\Omega_a,\ t\in[0,T] \\
&v_{aG}(z,t) = \begin{cases} v_{a0} + V_{up}\cdot l(z,t) & \text{si } z \le h_{\mathrm{lim}} \\ v_{a0} & \text{si } z > h_{\mathrm{lim}} \end{cases} && z\in\Omega_a,\ t\in[0,T] \\
&\theta_{LS}(z,t) = \theta_{a0} - \frac{5}{2}\cos\left(\frac{2\pi t}{T_p}\right)\cdot\left(\frac{Q_S(0,t)}{\max\limits_t(Q_S(0,t))} + 1.\right)\cdot\exp\left(\frac{-15z}{h_{\max}}\right) && z\in\Omega_a,\ t\in[0,T] \\
&q_{LS}(z,t) = q_{a0} && z\in\Omega_a,\ t\in[0,T]
\end{aligned} \tag{49}$$

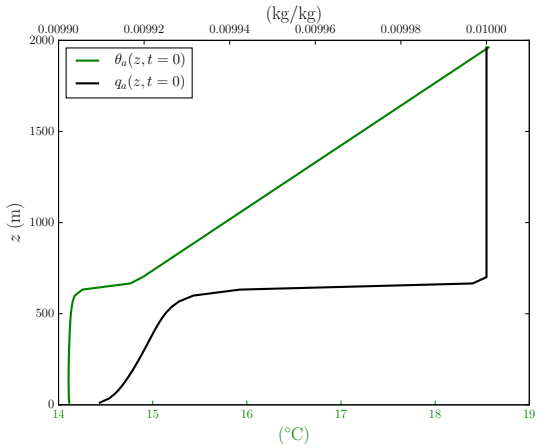
where $T_p = 86400$ s, $U_{up} = 15\,\mathrm{m.s^{-1}}$, $V_{up} = 13\,\mathrm{m.s^{-1}}$ and :

$$l(z,t) = \exp\left(-\frac{(t-\mu)^2}{2\cdot\sigma^2}\right)\cdot\sin\left(\frac{\pi z}{h_{\mathrm{lim}}}\right)$$

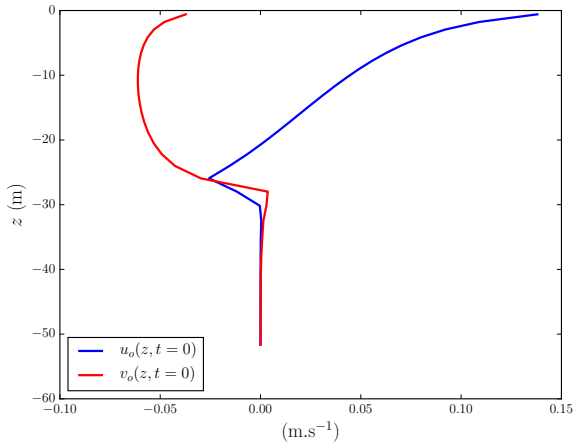
with $\mu = 25\cdot 10^4$ s, $h_{\mathrm{lim}} = 400$ m and $\sigma = 6\cdot 10^4$ s. Function $l$ along with $U_{up}$ and $V_{up}$ allows to define a wind variation for $0 \le z \le h_{\mathrm{lim}}$, which looks like a Gaussian, centred around the 3$^{\mathrm{rd}}$ day of the time window. Large scale relaxation term $\theta_a$ is defined so that it mimic the diurnal cycle. Figure (8) shows these large scale forcings on a 7 day window.



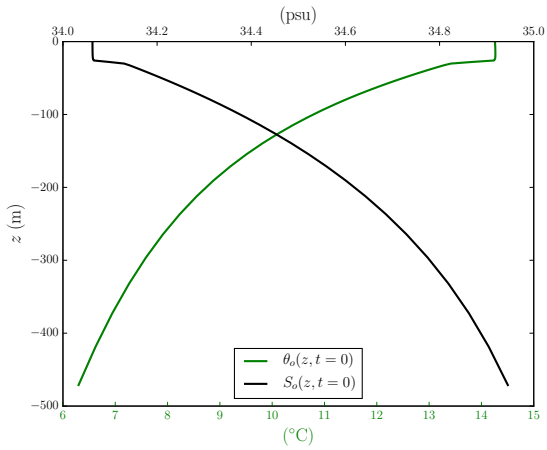

(a) Atmosphere wind




(b) Atmosphere temperature and humidity




(c) Ocean current velocities




(d) Ocean temperature and salinity

Figure 7: Initial vertical profiles for atmosphere (top) and ocean (bottom) reference test case. (a) shows wind $u_a(z,0)$ and $v_a(z,0)$ profiles and (b) atmosphere temperature and humidity $\theta_a(z,0)$ and humidity $q_a(z,0)$ profiles. (a) shows ocean current velocity profiles $u_o(z,0)$ and $v_o(z,0)$ for the first 50m and (b) shows ocean temperature $\theta_o(z,0)$ and salinity $S_o(z,0)$ profiles.

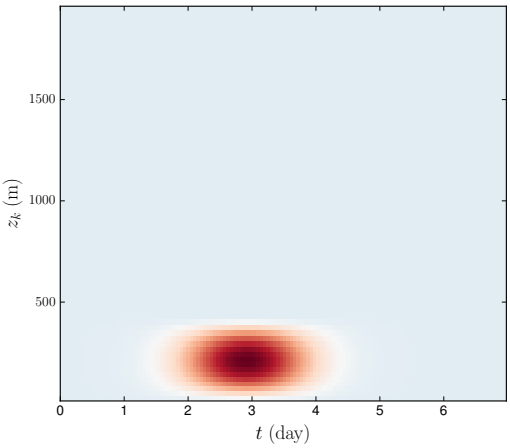
### B.3 Prescribed ocean-atmosphere Fluxes

As mentioned before, $Q_{S_0}$, $Q_{L\downarrow}$ and $P$ fluxes are exogenous of our system and have to be prescribed

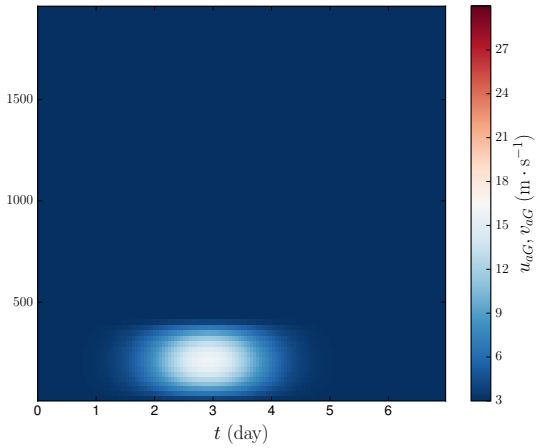
To do so, one first needs to define the surface pressure

$$p(0,t) = 101320\,\text{Pa} \quad \forall\, t \in [0,T] \tag{50}$$

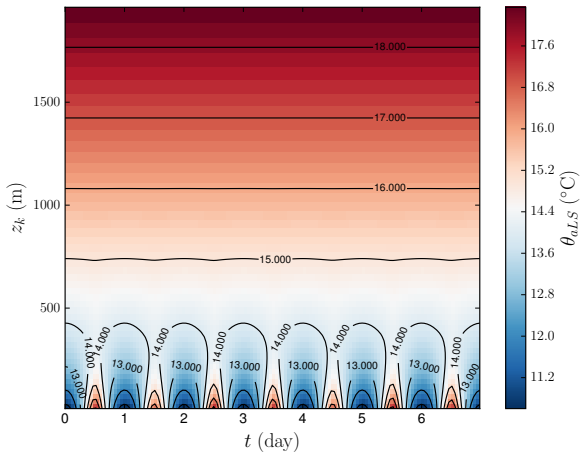
Solar flux $Q_S(0,t)$ is then defined so that both a diurnal cycle and a day-to-day variability are present (see figure 9a).




(a) Zonal geostrophic wind $u_{aG}$




(b) Meridional geostrophic wind $v_{aG}$




(c) Large scale potential temperature relaxation $\theta_{aLS}$

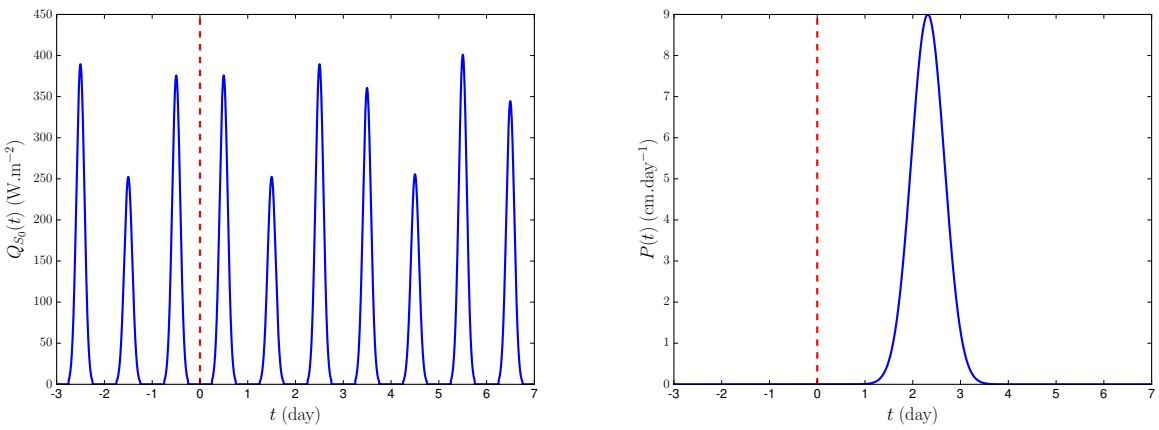
Figure 8: Large scale atmospheric forcings. (a) and (b) show the zonal and meridional geostrophic winds and (c) shows the large scale potential temperature.

$$
\begin{aligned}
Q_{S_0} &= \begin{cases} \mathrm{Var}(t) \cdot Q_{\max} \cdot \exp\left(-\dfrac{(t-\mu)^2}{2\sigma^2}\right) & \forall t \in [21600 + k \cdot J, 64800 + j \cdot J] \\ & \text{with } J = 86400\,\mathrm{s} \text{ and } j \in [\![0,6]\!] \\ 0 & \text{otherwise} \end{cases} \\
\text{with} \quad \mathrm{Var}(t) &= 0.8 + 0.2 \cdot \cos\left(\frac{t}{40000}\right) \\
Q_{\max} &= 420\,\mathrm{W.m}^{-2} \\
\sigma &= 7200\,\mathrm{s} \\
\mu &= 43200 + j \cdot J\,\mathrm{s}
\end{aligned}
\tag{51}
$$

There is then no solar flux at night (between 18 h and 6 h) and it follows a Gaussian-type evolution between 6 h and 18 h. Daily maximum is set via $\mathrm{Var}(t)$ which adds a day-to-day variability throughout the time window.

Precipitation flux $P(t)$ is set so that strong precipitations occurs at the same time as the strong geostrophic wind event mentioned in the previous section (around 3rd day, see figure 9b). This can be seen as the signature of a depression passing through our atmosphere column.

$$\begin{aligned} P(t) &= \frac{1}{0.001} 9 \cdot \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) \text{kg.m}^{-2}.\text{s}^{-1} \quad \forall t \in [0, T] \\ \text{with} \quad \mu &= 200000 \text{ s} \\ \sigma &= 30000 \text{ s} \end{aligned} \tag{52}$$




(a) Prescribed solar flux $Q_S(0, t)$

(b) Prescribed precipitation $P(t)$

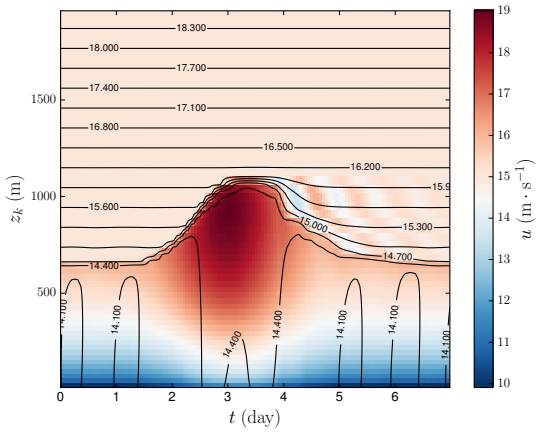
Figure 9: Reference prescribed solar flux at ocean surface $Q_S(0, t)$ (a) and precipitation $P(t)$ (b) over time

Lastly waveband radiation flux $Q_{L\downarrow}$ is set constant throughout the time window.

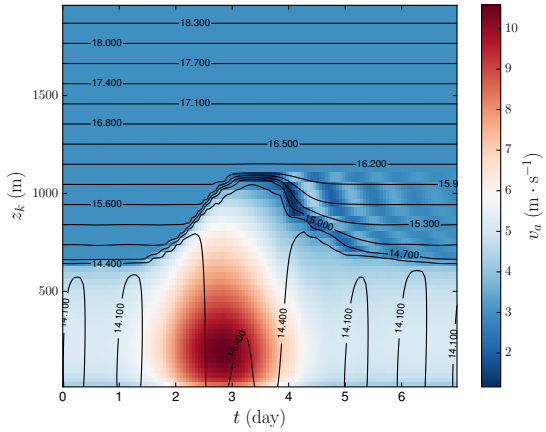
$$Q_{L\downarrow}(t) = 350 \text{ W.m}^{-2} \quad \forall t \in [0, T] \tag{53}$$

## B.4 Model simulations

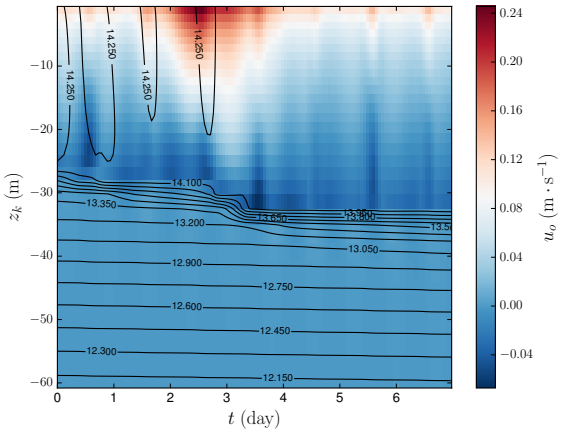
Figures (10) show profiles of wind and current velocities over all the time window. We can notice an increase responding to that of geostrophic winds. This increase had the effect of raising the level of the atmospheric boundary layer around 1000 m before returning to its previous altitude. We can also observe that the air temperature varies according to the diurnal cycle. This comes from changes in ocean surface temperature directly impacted by solar flux $Q_S$. Figure (12) represents temperature differences at the air-sea interface between both media. This diagnosis allows to determine if the system is in a stable ($\theta_a > \theta_o$) or unstable ($\theta_a < \theta_o$) regime.



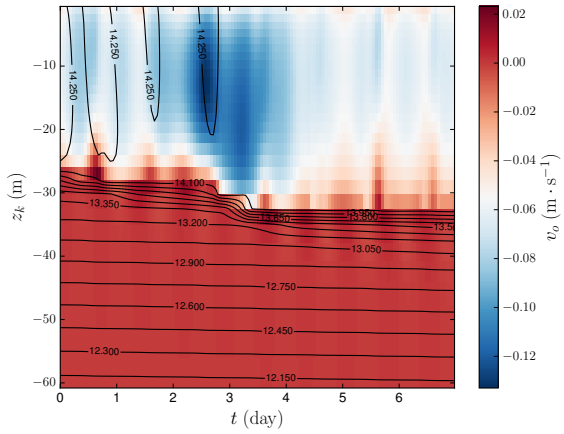

(a) Zonal wind $u_a(z,t)$




(b) Meridional wind $v_a(z,t)$

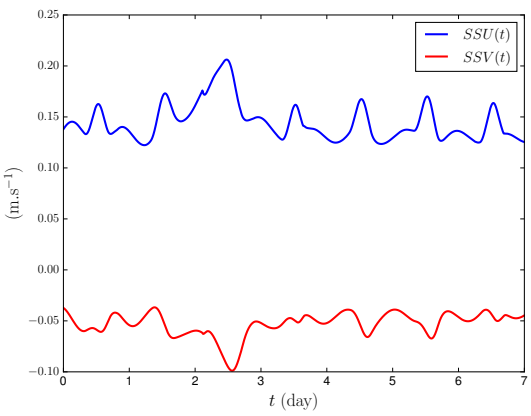



(c) Zonal ocean current velocity $u_o(z,t)$

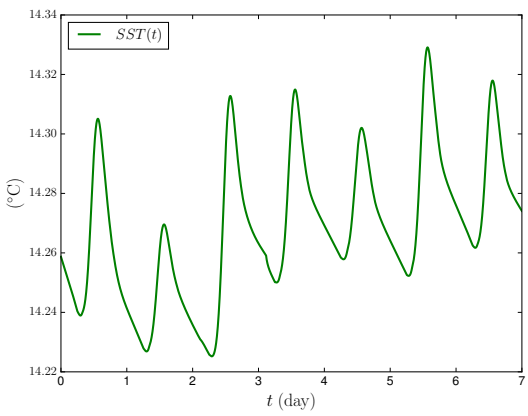



(d) Meridional ocean current velocity $v_o(z,t)$

Figure 10: Colors represent zonal (a,c) and meridional (b,d) atmosphere wind and ocean current velocities components and black isolines represent temperatures.




(a) First-guess SSU and SSV




(b) First-guess SST

Figure 11: First guess for ocean interface conditions for SSU and SSV (a) and SST b

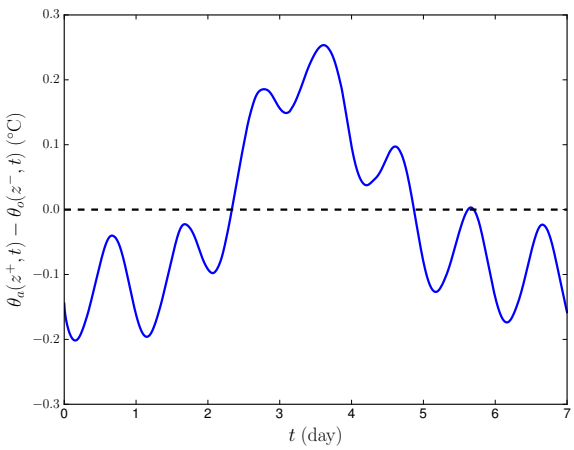



Figure 12: Differences between atmosphere and ocean surface temperature $\theta_a(z^+, t) - \theta_o(z^+, t)$

# C Elements of convergence study

Let us recall that the aim of any data assimilation methods mentioned in this report is to minimise, albeit approximatively, the following cost function:

$$J(\mathbf{x}) = \frac{1}{2}\left(\mathbf{x}-\mathbf{x}^b\right)^T \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^b\right) + \frac{1}{2}\left(\mathcal{H}(\mathcal{M}(\mathbf{x}))-\mathbf{y}^o\right)^T \mathbf{R}^{-1}\left(\mathcal{H}(\mathcal{M}(\mathbf{x}))-\mathbf{y}^o\right) \tag{54}$$

where $\mathcal{M}$ is the coupled model.

Since iterative schemes are used for solving the optimisation problem it is important to study their convergence. To that matter, we use a different classification than that of the core of the report. On the one hand we consider incremental and approximate incremental 4DVar schemes and on the other hand splitting-based 4DVar. Following this classification for the algorithm presented in this report, FCM can be seen as a direct application of incremental 4D-Var, while PCM is an approximate incremental 4DVar. WCM is a splitting based 4DVar and CERA can either be seen as a splitting 4DVar or an approximate incremental 4DVar.

## C.1 Incremental and approximate incremental 4DVar

Incremental 4DVar is a well known and thoroughly studied variational data assimilation scheme. In particular Gratton et al. (2007) (GLN2007 hereafter) studies convergence properties for approximate incremental 4D-Var, where the tangent linear model used in the inner loop is an approximation of that of the original problem. In this section we shamelessly exploit results from GLN2007 that are directly applicable to our problems.

First let us reformulate the 4Dvar problem in a form compatible with GLN2007 and recall their main results. If $n$ is the size of the control vector and $p$ the size of the observation vector, let $F : \mathbb{R}^n \longrightarrow \mathbb{R}^{n+p}$ such that

$$F(\mathbf{x}) = \begin{pmatrix} \mathbf{B}^{-1/2}\left(\mathbf{x}-\mathbf{x}^b\right) \\ \mathbf{R}^{-1/2}\left(\mathcal{H}(\mathcal{M}(\mathbf{x}))-\mathbf{y}^o\right) \end{pmatrix} \tag{55}$$

Equation 54 can then be rewritten

$$J(\mathbf{x}) = \frac{1}{2}\|F(\mathbf{x})\|_2^2 \tag{56}$$

Denoting $\mathbf{F_x} = \begin{pmatrix} \mathbf{B}^{-1/2} \\ \mathbf{R}^{-1/2}\mathbf{H_x}\mathbf{M_x} \end{pmatrix}$ the jacobian (tangent linear) of $F$ differentiated around $\mathbf{x}$, gradient and Hessian of $J$ read

$$\nabla_{\mathbf{x}} J = \mathbf{F}_{\mathbf{x}}^T F(\mathbf{x}) \quad \in \quad \mathbb{R}^n \tag{57}$$

$$\nabla\nabla_{\mathbf{x}} J = \mathbf{F}_{\mathbf{x}}^T \mathbf{F_x} + Q(\mathbf{x}) \quad \in \quad \mathbb{R}^{n\times n} \tag{58}$$

where $Q(\mathbf{x})$ denotes the second order terms

Incremental 4DVar, which can be seen as a Gauss-Newton algorithm, solves problem 55 by successive approximations neglecting the second order terms.

Theorem 4 of GLN2007 states that a sufficient condition for this algorithm to converge to the minimum of the original problem is that there exist a sequence $\eta^{(k)} < 1$ such that

$$\left\| Q(\mathbf{x}^{(k)})\left(\mathbf{F}_{\mathbf{x}^{(k)}}^T \mathbf{F}_{\mathbf{x}^{(k)}}\right)^{-1} \right\|_2 \leq \eta^{(k)} \tag{59}$$

**Algorithm 1:** Incremental 4D-Var

**while** *not converged* **do**
  compute $F(\mathbf{x}^{(k)})$ and differentiate;
  solve inner problem $\mathbf{F}_{\mathbf{x}^{(k)}}^T \mathbf{F}_{\mathbf{x}^{(k)}} \delta\mathbf{x}^{(k)} = -\mathbf{F}_{\mathbf{x}^{(k)}}^T F(\mathbf{x}^{(k)})$;
  update reference $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \delta\mathbf{x}^{(k)}$
**end**

Meaning that the change of curvature of J must not be too strong. Additionally, for convergence to be ensured the first guess of the algorithm $\mathbf{x}^b$ has to be 'close enough' to the global optimum (depending on the aforementioned curvature).

In general, due to non differentiabilities in the model and in order to save computing time, the jacobian used in the inner loop is only an approximation of $\mathbf{F}_{\mathbf{x}^{(k)}}$. For instance PCM, with only part of the coupling being accounted for, is a good example for such an approximation. CERA goes a little bit further, neglecting the coupling processes in the inner loop altogether. In order to describe this approximate incremental 4DVar (a.k.a Perturbed Gauss-Newton in GLN2007), let us denote $\widetilde{\mathbf{F}}_{\mathbf{x}^{(k)}}$ this approximate tangent operator.

**Algorithm 2:** Approx-Incremental 4D-Var

**while** *not converged* **do**
  compute $F(\mathbf{x}^{(k)})$ and differentiate;
  solve inner problem $\widetilde{\mathbf{F}}_{\mathbf{x}^{(k)}}^T \widetilde{\mathbf{F}}_{\mathbf{x}^{(k)}} \delta\mathbf{x}^{(k)} = -\widetilde{\mathbf{F}}_{\mathbf{x}^{(k)}}^T F(\mathbf{x}^{(k)})$;
  update reference $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \delta\mathbf{x}^{(k)}$
**end**

Then GLN2007's theorem 6 tells us that the sufficient condition becomes

$$\left\| \mathbf{I} - \left( \widetilde{\mathbf{F}}_{\mathbf{x}^{(k)}}^T \mathbf{F}_{\mathbf{x}^{(k)}} + \tilde{Q}(\mathbf{x}^{(k)}) \right) \left( \widetilde{\mathbf{F}}_{\mathbf{x}^{(k)}}^T \widetilde{\mathbf{F}}_{\mathbf{x}^{(k)}} \right)^{-1} \right\|_2 \leq \eta^{(k)} \tag{60}$$

Note that if $\widetilde{\mathbf{F}}_{\mathbf{x}^{(k)}} = \mathbf{F}_{\mathbf{x}^{(k)}}$ this is equivalent to condition 59. Condition 60 is a bit more difficult to interpret in a geometrical way. However it states that for convergence to be guaranteed, the requirement is that $\widetilde{\mathbf{F}}_{\mathbf{x}}^T \widetilde{\mathbf{F}}_{\mathbf{x}}$ has to be a good approximation of $\widetilde{\mathbf{F}}_{\mathbf{x}}^T \mathbf{F}_{\mathbf{x}} + \tilde{Q}(\mathbf{x})$ (the Hessian of the approximate inner problem).

Even if the Approximate 4DVar converges, there is no guarantee that it will converge toward the same minimum as the original problem. In general it is not the case, and theorem 7 from GLN2007 gives a superior bound of this error at optimum

$$\|\tilde{\mathbf{x}}^* - \mathbf{x}^*\|_2 \leq \frac{1}{1-\nu} \left\| \left( \widetilde{\mathbf{F}}_{\tilde{\mathbf{x}}^*}^+ - \mathbf{F}_{\tilde{\mathbf{x}}^*}^+ \right) F(\tilde{\mathbf{x}}^*) \right\|_2 = \frac{1}{1-\nu} \left\| \mathbf{F}_{\tilde{\mathbf{x}}^*}^+ F(\tilde{\mathbf{x}}^*) \right\|_2 \tag{61}$$

where $\mathbf{F}^+ = \left(\mathbf{F}^T\mathbf{F}\right)^{-1}\mathbf{F}^T$ denotes the Moore Penrose inverse and $0 \leq \nu < 1$ is the upper bound of the convergence speed of the non approximated problem and depends on second order terms $Q$ ($\nu = 0$ in the linear case). In vernacular words, this can be interpreted as the less regular $J$ is, the less the inner model can be approximated.

At the fixed point, the perturbed Jacobian $\widetilde{\mathbf{F}}_{\mathbf{x}^*}$ must, be such that $\widetilde{\mathbf{F}}_{\mathbf{x}^*}^T F(\mathbf{x}^*)$

**Remark:** This section only focuses on convergence (or not) of the outer problem. The inner problem is quadratic, so it always converges. Regarding the speed of convergence, for outer problems it is directly driven by $\eta_k$ values (the smaller the better), while for inner problems it depends on the conditioning of $\widetilde{\mathbf{F}}^T_{\mathbf{x}^{(k)}}\widetilde{\mathbf{F}}_{\mathbf{x}^{(k)}}$. However CERA can perform the minimisation separately on each media, so the inner problem is likely to converge faster, depending on the conditioning of the sub problems.

## C.2 Splitting methods

WCM, and CERA to a certain extent, can be seen as a splitting algorithm. Indeed, the original problem is split into two somewhat independent problems. In particular WCM can be seen as resembling a parallel version of the antique Alternating Direction Method of Multiplier (Fortin & Glowinski 1985, ADMM), which is used in machine learning.

ADMM is meant for linear problems where the minimisation problem can be rewritten into 2 independent constrained minimisation problems *e.g.* finding

$$\arg\min J(\mathbf{x})$$

is equivalent to finding

$$\begin{gathered}\arg\min\left(J_a(\mathbf{x}_a)+J_o(\mathbf{x}_o)\right)\\ \text{under constraint } \mathbf{A}\mathbf{x}_a=\mathbf{B}\mathbf{x}_o\end{gathered}$$

The ADMM algorithm is then:

**Algorithm 3:** ADMM

initialise $\mathbf{x}_a^{(0)}=\mathbf{x}_a^b$, $\mathbf{x}_o^{(0)}=\mathbf{x}_o^b$ and $\mathbf{c}^{(0)}=0$;
**while** *not converged* **do**
  $\mathbf{x}_a^{(k+1)}=\arg\min_{\mathbf{x}_a}\left(J_a(\mathbf{x}_a)+\gamma\|\mathbf{A}\mathbf{x}_a+\mathbf{B}\mathbf{x}_o^{(k)}+\mathbf{c}^{(k)}\|^2\right)$;
  $\mathbf{x}_o^{(k+1)}=\arg\min_{\mathbf{x}_o}\left(J_o(\mathbf{x}_o)+\gamma\|\mathbf{A}\mathbf{x}_a^{(k+1)}+\mathbf{B}\mathbf{x}_o+\mathbf{c}^{(k)}\|^2\right)$;
  $\mathbf{c}^{(k+1)}=\mathbf{c}^{(k)}+\mathbf{A}\mathbf{x}_a^{(k+1)}-\mathbf{B}\mathbf{x}_o^{(k+1)}$
**end**

ADMM is known to converge if $J$, $J_a$ and $J_o$ are proper, convex and lower semi-continuous functions, which is generally the case in data assimilation if $\mathscr{H}$ and $\mathscr{M}$ are linear.

In CERA, which is not really an ADMM, $\gamma=0$ but $J_a$ and $J_o$ are updated at each iteration. WCM on the other hand is quite close, $\gamma\|\mathbf{A}\mathbf{x}_a+\mathbf{B}\mathbf{x}_o+\mathbf{c}\|^2$ resembling the $J^s$ term.

In the non linear case, in order to ensure convergence, one could imagine a 3-level algorithm where the inner problem of a non approximated incremental 4DVar would be solved through an ADMM algorithm. Such an algorithm would require the same conditions as for incremental 4DVar to converge. However, it would be quite expensive since it would multiply the number of minimisations. WCM proposes to squeeze it into a two-level minimisation, just like incremental 4DVar, where both non linearity and $J^s$ update are done at the same time (during the outer iteration). Convergence study of this variant remains to be done.

### C.3 Discussion

Considering coupled data assimilation schemes as (perturbed) Gauss-Newton algorithms allows for reusing the theoretical convergence results. However in the non linear case, theory only provide sufficient conditions (i.e. the algorithm may converge even if they are violated). Moreover these conditions are difficult to evaluate due to the second order terms. On the other hand, in the linear case these sufficient conditions become also necessary and are far easier to evaluate if tangent and adjoint models are available. Consequently, if full convergence of the algorithm were the main concern, the best option would probably be to use three level strategy where the inner problem would itself be solved through either Perturbed Gauss-Newton or splitting. In practice full convergence is seldom sought for, and the number of iteration being limited, speed of (partial) convergence at the early stage of the algorithm is the most important. Theory also gives a measure of inner convergence speed (the $\eta_k$s), but yet again, this can only be really exploited in the linear case. Evidence actually shows that ECMWF operational uncoupled incremental 4D-Var does not converge as outer loops is concerned but it is very efficient at providing a good analysis with a limited number of outer iterations (Trémolet 2007).

## References


Cuxart, J., Bougeault, P. & Redelsperger, J.-L. (2000), 'A turbulence scheme allowing for mesoscale and large-eddy simulations', *Quarterly Journal of the Royal Meteorological Society* **126**(562), 1–30.
**URL:** *http://dx.doi.org/10.1002/qj.49712656202*

Fortin, M. & Glowinski, R. (1985), 'Augmented lagrangian methods: Applications to the numerical solution of boundary-value problems.', *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik* **65**(12), 622–622.
**URL:** *http://dx.doi.org/10.1002/zamm.19850651211*

Gander, M. J. (2008), 'Schwarz methods over the course of time', *Electron. Trans. Numer. Anal.* **31**, 228–255. http://etna.math.kent.edu/vol.31.2008/pp228-255.dir.

Gejadze, I. Y. & Monnier, J. (2007), 'On a 2D 'zoom' for the 1D shallow water model: Coupling and data assimilation', *Comput. Methods Appl. Mech. and Engrg.* **196**(45–48), 4628 – 4643.

Gratton, S., Lawless, A. S. & Nichols, N. K. (2007), 'Approximate Gauss–Newton Methods for Nonlinear Least Squares Problems', *SIAM J. Optim.* **18**(1), 106–132.

Janisková, M., Thépaut, J.-N. & Geleyn, J.-F. (1999), 'Simplified and regular physical parameterizations for incremental four-dimensional variational assimilation', *Monthly Weather Review* **127**(1), 26–45.
**URL:** *https://doi.org/10.1175/1520-0493(1999)127<0026:SARPPF>2.0.CO;2*

Janisková, M., Veersé, F., Thépaut, J.-N., Desroziers, G. & Pouponneau, B. (1999), 'Impact of a simplified physical package in 4d-var analyses of fastex situations', *Quarterly Journal of the Royal Meteorological Society* **125**(559), 2465–2485.
**URL:** *http://dx.doi.org/10.1002/qj.49712555907*

Large, W. B. (2006), *Surface Fluxes for Practitioners of Global Ocean Data Assimilation*, Springer Netherlands, Dordrecht, pp. 229–270.
**URL:** *https://doi.org/10.1007/1-4020-4028-8_9*

Large, W. G., McWilliams, J. C. & Doney, S. C. (1994), 'Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization', *Reviews of Geophysics* **32**(4), 363–403.
**URL:** *http://dx.doi.org/10.1029/94RG01872*

Le Dimet, F. X. & Talagrand, O. (1986), 'Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects', *Tellus A* **38A**(2), 97–110.
**URL:** *http://dx.doi.org/10.1111/j.1600-0870.1986.tb00459.x*

Lemarié, F., Debreu, L. & Blayo, E. (2013*a*), 'Toward an optimized global-in-time Schwarz Algorithm for diffusion equations with discontinuous and spatially variable coefficients. Part 1: The constant coefficients case', *Electron. Trans. Numer. Anal.* **40**, 148–169.

Lemarié, F., Debreu, L. & Blayo, E. (2013*b*), 'Toward an optimized global-in-time Schwarz Algorithm for diffusion equations with discontinuous and spatially variable coefficients. Part 2: The variable coefficients case', *Electron. Trans. Numer. Anal.* **40**, 170–186.

McWilliams, J. (2006), *Fundamentals of Geophysical Fluid Dynamics*, Cambridge University Press.

Mulholland, D. P., Laloyaux, P., Haines, K. & Balmaseda, M. A. (2015), 'Origin and Impact of Initialization Shocks in Coupled Atmosphere-Ocean Forecasts', *Mon. Wea. Rev.* **143**, 4631–4644.

Smith, P., Fowler, A. & Lawless, A. (2015), 'Exploring strategies for coupled 4D-Var data assimilation using an idealised atmosphere-ocean model', *Tellus A* **67**(0).
**URL:** *http://dx.doi.org/10.3402/tellusa.v67.27025*

Smith, P. J., Lawless, A. S. & Nichols, N. K. (2017), 'Estimating forecast error covariances for strongly coupled atmosphere–ocean 4d-var data assimilation', *Monthly Weather Review* **145**(10), 4011–4035.
**URL:** *https://doi.org/10.1175/MWR-D-16-0284.1*

Trémolet, Y. (2007), 'Incremental 4D-Var convergence study', *Tellus A* **59**(5), 706–718.
**URL:** *http://doi.wiley.com/10.1111/j.1600-0870.2007.00271.x*