Geoscientific
Model Development
Discussions

Open Access

NEMOTAM: tangent and adjoint models for the ocean modelling platform NEMO

A. Vidard[1], P.-A. Bouttier[2], and F. Vigilant[1]

[1]Inria, Univ. Grenoble Alpes, CNRS, LJK, 38000 Grenoble, France
[2]Univ. Grenoble Alpes, CNRS, LGGE, 38000 Grenoble, France

Correspondence to: A. Vidard (vidard@imag.fr)

GMDD

7, 6705–6732, 2014

NEMOTAM

A. Vidard et al.

Abstract

The tangent linear and adjoint model (TAM) are efficient tools to analyse and to control dynamical systems such as NEMO. They can be involved in a large range of applications such as sensitivity analysis, parameter estimation or the computation of characteristics vectors. TAM is also required by the 4-D-VAR algorithm which is one of the major method in Data Assimilation. This paper describes the development and the validation of the Tangent linear and Adjoint Model for the NEMO ocean modelling platform (NEMOTAM). The diagnostic tools that are available alongside NEMOTAM are detailed and discussed and several applications are also presented.

1 Introduction and history

Tangent linear and adjoint models (TAM in the following) are powerful modelling tool. The tangent linear model (TLM) provides the directional derivatives respect to a trajectory of the corresponding non linear system. The adjoint of the TLM gives information about the response of the system to variations of its input. TAM are therefore widely used for variational assimilation applications, but also for the analysis of physical processes, since they can be used for sensitivity analysis, parameter identification and for the computation of characteristic vectors (singular vectors, Liapunov vectors, etc. see Moore et al., 2004, for an extended review).

This is particularly true for geophysical applications such as meteorology and oceanography where many data assimilation systems rely on the availability of such models. However, only few ocean general circulation models are routinely provided with their TAM. Among them, one can cite MITgcm (MIT general circulation model) and ROMS (Regional Ocean Modelling System). They are quite symbolic of the two possible routes for deriving tangent and adjoint model, either using automatic differentiation tool (MITgcm uses automatic differentiation Marotzke et al., 1999) or hand-coded (ROMS' TAM was developed by a pool of researchers Moore et al., 2004).

Automatic tools are now mature enough, and they show some significant advantages over the hand-coded route. Differentiation of a numerical code is a tedious and error-prone task, the use of an automatic tool could alleviate this difficulty. Moreover when an updated version of the non linear model is provided the corresponding TAM can in general be obtained effortlessly. Additionally automatic tools offer an important flexibility when one wants to change the variable to differentiate around. Indeed, typically, for data assimilation purpose, the TAM is differentiated around the initial condition, while for parameters estimation it is differentiated around the sought parameter set. Most of the time the obtained TAM are very similar, but in some cases (e.g. for grid related parameter) going from one to another can lead to a significant amount of code changes.

On the other hand, automatic differentiation suffer from some limitations compared to hand-coding. First some newer (or archaic) language features may not be supported at first by the automatic tools, second the numerical performance of automatic derived TAM is still relatively poor compared to that of hand-coded one, in particular, the handling of the parallelization is still an open issue. Third non differentiable part of the original code still require to be dealt with by specialists in the field. All these issues can however be overcame by mixing automatic and manual coding.

The NEMO ocean engine (Madec, 2008) was previously known as the OPA model (Madec et al., 1998). It used to have a TAM (called OPATAM), fully hand-coded. OPATAM was initially developed for a Pacific ocean configuration, and targeted at variational data assimilation applications in the framework of OPAVAR (Weaver et al., 2003, 2005). OPATAM/OPAVAR were extended to other regional basins (Mediterranean sea (Rémy, 1999), North Atlantic 1/3° (Forget et al., 2008), South Atlantic 1°), to the global ocean (ORCA 2° Daget et al., 2009), and were used for methodological studies such as control of the 3-D model error (Vidard, 2001), control of the surface forcing and open boundary conditions (Deltel, 2002; Vossepoel et al., 2003). OPATAM was also used for sensitivity studies (Sévellec et al., 2008), singular vectors (Moore et al., 2003; Sévellec et al., 2009), etc.

Back   Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

For several reasons, mainly because of lack of workforce, OPATAM, OPAVAR and related developments were not included in the standard release of OPA. As a consequence, synchronisation of OPATAM with OPA's releases could not be achieved on a regular basis, and all developments were on individual branches, without feedback to the OPATAM/OPAVAR system. The pool of potential users was therefore reduced significantly. The complete rewriting of the model during the transition to NEMO rendered OPATAM and OPAVAR obsolete. As part of the NEMOVAR initiative (a variational data assimilation with NEMO, Mogensen et al., 2009) a first prototype of NEMOTAM was obtained by (Tber et al., 2007) using the TAPENADE automatic differentiation tool (Hascoët and Pascual, 2004) for a fixed and somewhat simplified configuration (ORCA 2° with all non-differentiable options switched off Tber et al., 2007). Even for this simplified configuration, however, substantial human intervention and additional work was required to obtain a useable product from the raw generated code. Three main drawbacks were identified for this application. First, the memory management and CPU performance of the raw code were rather poor. Second, the version of TAPENADE at that time generated single-processor code only and could not handle directives from the C-PreProcessor (CPP keys), which are widespread in NEMO. Third, the technique of binomial checkpointing that is used to handle nonlinearities (see Tber et al., 2007) is not compatible, at least in its present implementation, with the incremental algorithm of NEMOVAR, which employs separate executables and (possibly) different resolutions for the outer and inner loops. Improved memory management and extensions to support MPP and CPP keys are planned in future versions of TAPENADE so the first two deficiencies are not fundamental. The third deficiency, however, is more problematic and the trajectory management for nonlinearities in NEMOTAM is done differently from TAPENADE, along the lines of the simpler strategy implemented in OPAVAR. The modifications required to make any AD tool, compatible with the multi-incremental approach are really substantial and cannot be done in a short or medium term. Moreover the numerical performances of the generated TAM do not allow yet their use for "big" configurations and for operational applications. The writing of an adjoint code is not

a simple technical task, this explains why Autodiff tools can seldom be used as a black-box. Numerous potential problems can arise and in general they should be dealt with on a case by case basis by someone who masters both the model and differentiation techniques.

5   From that experience it has been decided to go toward the hand-coding approach, the use of AD tool being left aside for the time being. We may reconsider this in a medium or long term though. An optimal mix of both approaches is likely to be the preferred choice in the short term future and lead to a semi-automatic way of generating NEMOTAM.

10   This paper will first discuss the methodology used for the NEMOTAM development and explain some of the particular choices that have been made. Then the validation tools that are available alongside NEMOTAM will be detailed. Finally some application examples will be presented.

2   Methodology and choices for NEMOTAM

15   The approach used in developing NEMOTAM is based on the well known differentiation rules as described in Giering and Kaminski (1998). It relies in particular on the definition of active and passive variables. The former depends on the control variables (variables one differentiates around) variations, while the latter are independent of the control (typically model or grid parameters). Active variables have tangent and adjoint
20   counterpart while passive variables do not.

In the current version (3.4.1), only the general circulation component of NEMO is supported by NEMOTAM and a few key components are still missing, namely the variable volume (VVL), open boundary conditions (OBC) and the nesting capabilities (AGRIF). There is no fundamental reason for non supporting these options (support
25   for OBC is actually on its way), but it was not flagged as priority. In particular VVL would require a tremendous amount of coding since it makes the surface grid cells size

depending on the flow (they become then active variables). This is a typical case where a semi-automatic route would be beneficial.

When coding tangent and adjoint models, either automatically or manually, one has to face several difficulties. The main ones are listed in the remaining of this section and discussed in the particular case of NEMOTAM.

2.1 Non-differentiabilities: problem and solutions

In realistic models like those included in NEMO, non-differentiable physics are quite frequent. They can be non-differentiable by essence (e.g., step processes), due to the way they have been programmed or the chosen numerical methods (e.g. non-oscillatory schemes). Coding-wise they can be represented by an IF statement with a condition on an active variable (or a max, abs, etc. statement but it is equivalent). In order to manage these non-differentiable parts several options are available.

- Regularisation: either mathematical regularisation (thanks to the introduction of a differentiable connection), or a physical regularisation whenever it is possible (by rewriting the processes in a differentiable manner).

- Approximation: the direct model physics is approximated in TAM in order to transform the non-differentiable part. One should obviously be careful that the approximation is not too strong.

- Non linear branching: the non-differentiable part is kept in TAM but the same branching (same side of the IF) than in the direct model is used (i.e. one left- and right-differentiates separately). This is in general the preferred choice, but in some cases it can lead to pathological behaviour and should be done very carefully.

The choice of treatment heavily relies on the type of the non-differentiability and it is more a matter of an educated choice. Indeed it requires an important knowledge of both the direct model and differentiation techniques. A typical example of this in NEMO is the vertical mixing schemes included in OPA (TKE in particular, but it is true for the

other options as well). The current version is strongly non-differentiable. In NEMOTAM this has been sorted out by simplification of the physics, i.e. some active variables are computed by the direct model and treated as passive variable in the TAM. Adopting this strategy has the advantage to keep the full physics in the non linear model, but is

5 at the expense of an approximation in the TAM. However, most of the physical process it models being quite regular, it may be possible to rewrite the direct version to make it differentiable[1].

Another important source of non differentiability is the non-oscillatory part of one of the most popular tracer's advection term in NEMO (TVD). A non linear branching of

10 such scheme can lead to an unstable TAM. Following Thuburn (2001) two viable approximations are at hand: either one differentiate at continuous level and then apply the same non oscillatory discretisation scheme as for the direct model or the non oscillatory correcting term is removed altogether in the tangent and adjoint schemes. The former is generally a better approximation for the tangent model, but it degrades the

15 exactness of the adjoint model respect to the tangent model (see next section). For this reason, in NEMOTAM the second solution has been adopted.

2.2 Checkpointing

One of the issue one has to tackle when dealing with TAM is the storage and/or re-computation of the non-linear trajectory that is required for the differentiation of the

20 non-linear terms. This is particularly important for the adjoint model which need the trajectory in reverse order as it is produced by the non linear model. A common practice is the so-called checkpointing where well-chosen snapshots of the trajectory (the check-points) are stored and intermediate variables are re-computed between checkpoints. This is the strategy used in NEMOTAM, a subset of direct variables are stored ev-

25 ery time step and only required intermediate variables are recomputed. The number of non-linearities in NEMO being relatively small, the need for storage and re-computation

---

[1]This was the strategy adopted in ROMS for the KPP vertical mixing scheme.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◀ | ▶|

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

is limited. However, for long period of time the amount of required storage may become too important; in that case a possible approximation is to subsample the output of the trajectory (say one per day) and NEMOTAM will interpolate linearly between check-points, additionally it can be stored in single precision. The validity of these approxima-

5   tions is discussed in Sect. 3.

2.3   Numerical issues

The numerical characteristics of the direct model are not always conserved in the adjoint model. In particular the most common difficulty is related to the convergence speed of an iterative algorithm that may be different for the direct and its tangent/adjoint

10   counterparts.

In that case, a specific solution should be provided, it can go as far as replacing the problematic scheme by a more TAM-friendly one. Once again, in OPATAM this kind of problem arose, and one had to replace the conjugate gradient solver used for the computation of the surface pressure gradient (while self-adjoint, in theory) by a red-

15   black SOR solver. However, this scheme being on the verge of depreciation in NEMO, NEMOTAM will soon adopt the time-splitting-based direct solver.

Stability is not the only numerical issue: the TAM are, by essence, more expensive numerically than the direct model because of the additional computations they require. Moreover, and it is especially true for the adjoint model, direct code optimisations may

20   not be so optimal for the TAM. Therefore it is important to specifically optimise the TAM code, and this can only be done through a careful performance analysis. As of today, such an optimisation may not be achieved automatically. The tendency toward application at high and very high resolution make the computing cost aspect a crucial issue. NEMOTAM adopted the same domain decomposition strategy as for NEMO and

25   is therefore fully parallel, and some specific targeted optimisation have been performed. Additionally some irrelevant (as far as TAM are concerned) on-line diagnostics have been removed. Thanks to all these effort the additional computing burden has been

contained to around twice the non-linear integration cost (more or less, depending on the application).

3 Validation

Due to the numerous caveats mentioned in Sect. 2 validation of TAM is a crucial aspect. Fortunately there are several numerical tests that can be performed to ensure this validity, moreover these tests can be applied both to the whole model and to individual modules separately.

In NEMOTAM, modules include both the tangent and adjoint parts, as well as the *validation interface*, for numerical verification of the TAM source codes. The *adjoint test* would always be present, while the *tangent test* could be optional and reserved to specific and problematic routines.

In the following $\mathcal{M}(x)$ stands for the full non-linear NEMO model with initial state vector x, $\mathbf{L}(x) \equiv (\partial \mathcal{M}/\partial x)|_{x=x}$ its tangent linear model (possibly simplified), and $\mathbf{L}^*(x)$ the adjoint of $\mathbf{L}(x)$.

3.1 Adjoint validation

The adjoint part is actually relatively easy to check, indeed, by definition of the adjoint one gets:

$$(\mathbf{L}(x)\delta x, \delta y) = \langle \delta x, \mathbf{L}^*(x)\delta y \rangle \tag{1}$$

where $\langle .,. \rangle$ and $(.,.)$ denote the appropriate dot product. Equation (1) being exact, the relative error between the two computed scalar products must be close to zero barring rounding errors. In NEMOTAM, the actual test performed is

$$(\mathbf{L}(x)\delta x)^* \mathbf{W}\delta y = \delta x^* \mathbf{L}^*(x)\mathbf{W}\delta y \tag{2}$$

where δx is a random vector, **W** is a diagonal matrix of scale factors and $\delta y = \mathbf{L}(x)\delta x$. This test only ensure that the adjoint is indeed the adjoint of the tangent linear, which in turn has to be validated.

3.2 Tangent validation

Validation of tangent modules are more tricky since there is generally no affordable exact tests available. A classical method of testing a numerical tangent linear model **L** is to compare the evolution of a perturbation by **L** with the difference of two evolutions, with and without the perturbation, by the full nonlinear model $\mathcal{M}$.

Considering a fixed small perturbation vector $\gamma\delta x$, where γ is a scale parameter, the Taylor expansion of $\mathcal{M}$ reads:

$$\mathcal{M}(x_0 + \gamma\,\delta x_0, t) = \mathcal{M}(x_0, t) + \gamma\,\mathbf{L}(x, t)\,\delta x_0 + O(\gamma^2) \tag{3}$$

If $\mathcal{N}(x_0, \gamma\delta x_0, t_0, t)$ denotes the non linear evolution of a perturbation and we use the simplified notation $\mathcal{N}(\gamma\delta x_0, t)$,

$$\mathcal{N}(\gamma\delta x_0, t) = \mathcal{M}(x_0 + \gamma\delta x_0, t) - \mathcal{M}(x_0, t) \tag{4}$$

The linearization error $\mathcal{E}(\gamma\delta x_0, t)$ is defined by:

$$\mathcal{E}(\gamma\delta x_0, t) = \mathcal{N}(\gamma\delta x_0, t) - \gamma\mathbf{L}(x, t)\delta x_0 \tag{5}$$

From Eq. (3), $\mathcal{E}(\gamma\delta x_0, t)$ behaves in $O(\gamma^2)$.

And the first order accuracy index e_γ:

$$e_\gamma = \frac{\|\mathcal{N}(\gamma\delta x_0, t)\|}{\|\mathbf{L}(x, t)\gamma\delta x_0\|} \tag{6}$$

e_γ tends to one when γ tends to zero, validates $\mathbf{L}(x, t)$, moreover when γ is small enough, $\mathcal{N}$ enters a linear regime and e_γ converges toward 1 with a rate γ.

Table 1 shows an example of such tests on a single routine. It illustrates nicely the expected behaviour of ϵ_γ which gains one digit in precision when γ is divided by ten. This diagnostics gives information of both first order (tends to 1) and second order (at rate γ).

3.3 Estimation of the approximations error

As mentioned above, when differentiating realistic models, approximations have to be done. To estimate the effect of these approximations on the numerical tangent linear model **L**, one must first estimates the truncated part, $O((\gamma\delta x)^2)$, of the Taylor expansion (see Eq. 3). In order to do this, following Lawless et al. (2003), one can write the Taylor expansion of $\mathcal{E}(\gamma\delta x, t)$ whose each individual component l follows

$$\mathcal{E}_l(\gamma\delta \boldsymbol{x}_0, t) = \frac{\gamma^2}{2}\partial^2 \mathcal{M}_l \delta \boldsymbol{x}_0^{-2} + \frac{\gamma^3}{6}\partial^3 \mathcal{M}_l \delta \boldsymbol{x}_0^3 + O(\gamma^4) \tag{7}$$

On another hand, from two nonlinear perturbations:

$$\mathcal{N}(\delta \boldsymbol{x}_0, t) = \mathcal{M}(\boldsymbol{x}_0 + \delta \boldsymbol{x}_0, t) + O((\delta \boldsymbol{x}_0)^2, t)$$
$$\mathcal{N}(\gamma\delta \boldsymbol{x}_0, t) = \mathcal{M}(\boldsymbol{x}_0 + \gamma\delta \boldsymbol{x}_0, t) + O((\gamma\delta \boldsymbol{x}_0)^2, t) \tag{8}$$

one can compute

$$E(\gamma\delta \boldsymbol{x}_0, t) = \frac{\mathcal{N}(\gamma\delta \boldsymbol{x}_0, t) - \gamma\mathcal{N}(\delta \boldsymbol{x}_0, t)}{\gamma^2 - \gamma} \tag{9}$$

whose Taylor expansion reads (for each individual component l)

$$E_l(\gamma\delta \boldsymbol{x}_0, t) = \frac{1}{2}\partial^2 \mathcal{M}_l \delta \boldsymbol{x}_0^2 + \frac{1+\gamma}{6}\partial^3 \mathcal{M}_l \delta \boldsymbol{x}_0^3 + O(\gamma^4) \tag{10}$$

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

For small values of γ and $\delta \boldsymbol{x}_0$, one can compare E and $\mathcal{E}$. That way one builds-up an estimator of the numerical tangent linear model error:

$$\hat{\mathcal{E}} = 100 \left(1 - \frac{\|E\|}{\|\mathcal{E}\|} \right) \tag{11}$$

Moreover, in NEMO, the vast majority of the nonlinearities are quadratic ones, meaning the third order and above derivatives vanish from the Taylor expansion and one gets $E = \mathcal{E}$.

This diagnostic is very valuable when comparing different simplifications made to the tangent linear model. Table 2 shows $\hat{\mathcal{E}}$ for two configurations of NEMO: an academic test case that is fully differentiable (SEABASS, see Appendix A) and 2°-resolution global realistic configuration (ORCA2). This allows to measure the effect of approximations mentioned in Sects. 2.2 and 2.1

4 Application examples

As stated by Errico (1997): "the principal application of adjoint models is sensitivity analysis, and all its other applications may be considered as derived from it". Performing Sensitivity analysis means evaluating how variations on the input of the system will affect the output. This can be of use for understanding the behaviour of the system (sensitivity analysis, propagation of incertitude), for optimising it (through data assimilation for instance) and for performing stability analysis. This section presents an example of these three kind of applications.

4.1 Sensitivity analysis and data assimilation

In the variational context, sensitivity analysis is the computation of gradient of so-called response (or cost or objective or criterium) functions respect to given control vectors. In other words, given the model's state $\boldsymbol{x} = (x_1, \ldots, x_n)^\mathsf{T} \in \mathcal{X} \subset \mathbb{R}^n$ and a set of control

parameters $\alpha = (\alpha_1, \ldots, \alpha_p)^\mathsf{T} \in \mathcal{P} \subset \mathbb{R}^p$ one is interested in computing gradients respect to α of a given response function:

$$J : \mathcal{P} \to \mathbb{R}$$

$$\alpha \mapsto J(\alpha) = \int_0^\mathsf{T} \|\phi(\mathbf{x}, \alpha, t)\|^2 \mathrm{d}t \qquad (12)$$

where ϕ is a (possibly non-linear) function with values in $\mathbb{R}^m$. We considered here the mono-criterium case (J with values in $\mathbb{R}$) but the following can easily be extended to multi-criteria problems.

The scope of local sensitivity analysis is to compute exactly and efficiently the sensitivities of the system's response to variations in the system's parameters, around their nominal values.

This is translated by finding

$$S_\alpha = \nabla_\alpha J(\alpha) = \left(\frac{\partial J}{\partial \alpha_1}(\alpha), \ldots, \frac{\partial J}{\partial \alpha_p}(\alpha) \right)^\mathsf{T} \qquad (13)$$

is the local sensitivity vector of J to variations in α, it is a local sensitivity because it depends on the current estimate of α (and of $\boldsymbol{x}_0$).

Rewriting Eq. (12) as $J(\alpha) = \langle \varphi(\alpha), \varphi(\alpha) \rangle$, where $\langle ., . \rangle$ defines a dot product on $\mathbb{R}^m$. The Gâteaux-derivative of J in any direction $\delta\alpha$ reads

$$\mathrm{d}J(\alpha)[\delta\alpha] = \left\langle \varphi(\alpha), \frac{\partial\varphi}{\partial\alpha}.\delta\alpha \right\rangle = \left\langle \left[\frac{\partial\varphi}{\partial\alpha}\right]^* .\varphi(\alpha), \delta\alpha \right\rangle$$

$$\equiv \langle \nabla J, \delta\alpha \rangle \qquad (14)$$

Hence $\left[\frac{\partial\varphi}{\partial\alpha}\right]^*$ the adjoint of φ allows for the exact computation of the p components of ∇J at once.

One can find an example of application of such methods with NEMOTAM on the Mercator-ocean's GLORYS 1/4° global ocean reanalysis in Vidard et al. (2011). The initial objective was to try to estimate the influences of geographical areas to reduce the forecast error using an adjoint method to compute the sensitivities (while the GLORYS assimilation system is based on optimal interpolation). We conducted a preliminary study by considering the misfit to observations as a proxy of the forecast error and sought to determine the sensitivity of this difference to changes in the initial condition and/or to forcing. That should give an indication about the important phenomena to consider to improve this system.

The most easily interpreted case in this study is to consider a sensitivity criterion coming from the difference in SST maps at the final instant of the assimilation cycle, because of its dense coverage in space. This can be translated into computing the gradient:

$$J(\boldsymbol{x}_0, \boldsymbol{q}) = \frac{1}{2} \sum_{n=1}^{N_{\mathrm{SST}}} \| H_{\mathrm{SST}}(\boldsymbol{x}_n) - \mathrm{SST}^{\mathrm{obs}} \|^2_{\mathbf{R}^{-1}} \tag{15}$$

with a control vector made of $\boldsymbol{x}_0 = (u_0, v_0, T_0, S_0, \eta_0)^{\mathsf{T}}$ the initial state vector (current velocities component, temperature, salinity and sea level) and of $\boldsymbol{q} = (q_{\mathrm{sr}}, q_{\mathrm{ns}}, \mathrm{emp})^{\mathsf{T}}$ (radiative fluxes, total heat fluxes, fresh water fluxes). One can see an example of sensitivity to initial temperature (surface and 100 m) as shown in the two bottom panels of Fig. 1. High sensitivity will give a signal similar to the gap in observations (top left), while low sensitivity will show a white area. In this example it is clear that the SST misfit is highly sensitive to changes in surface temperature where the initial mixed layer depth (top right) is low and insensitive elsewhere. The opposite conclusion can be drawn from the sensitivity to the initial temperature at 100 m. This is obviously not a surprise, and corresponds more to the purpose of verification of the model rather than the original goal of assimilation system improvement. However it highlights the importance of having a good estimate of the vertical mixing and echoes the fact they

Full Screen / Esc

this vertical mixing is often perturbed by data assimilation. Other components of the gradient show the important role of atmospheric forcing (again we could have guessed) and ways to improve the system also appear to point to that direction. With the objective of improving the data assimilation system, this approach is obviously not completely satisfactory because, strictly speaking, the assimilation system should be included in the optimality system. In theory, this assimilation system being linear and made of matrix multiplication, to derive its adjoint should be easy, in practice it's a different story, manipulating an operational system is never easy.

The sensitivities are of interest by themselves, but they can also be used for optimising the system. In particular this way of computing gradient is extensively used in variational data assimilation for the minimisation of similar cost function (4D-Var). For ocean application, historically the preferred choice of data assimilation technique has been (and still is for many cases) that of Optimal Interpolation or 3D-Var types schemes. These algorithms make the assumptions that the system state (or the increment in their FGAT formulation) is stationary over a given time window (typically 1 to 10 days) which can be a crude approximation. 4D-Var does not make this assumption and uses the adjoint model to compute the gradient of a cost function of the form:

$$J(\boldsymbol{x}_0) = \frac{1}{2}\|\boldsymbol{x}_0 - \boldsymbol{x}^{\mathrm{b}}\|_{\mathbf{B}^{-1}}^2 + \frac{1}{2}\sum_{t=1}^{T} \| H_t(\mathcal{M}(\boldsymbol{x}_0, t)) - \boldsymbol{y}_t^{\mathrm{obs}} \|_{\mathbf{R}^{-1}}^2 \tag{16}$$

where $\|z\|_C^2 = \langle z, Cz \rangle$ and $\mathbf{B}$ (resp $\mathbf{R}$) is the background (resp observation) error covariance matrix. $\boldsymbol{x}^{\mathrm{b}}$ its the background state, and $\boldsymbol{y}^{\mathrm{obs}}$ are the observations. The gradient ∇J of this cost function can be computed using relation (14):

$$\nabla J = \mathbf{B}^{-1}(\boldsymbol{x}_0 - \boldsymbol{x}^{\mathrm{b}}) + \sum_{t=1}^{T} \mathbf{L}^*\mathbf{H}_t^*\mathbf{R}^{-1}(H_t(\mathcal{M}(\boldsymbol{x}_0, t)) - \boldsymbol{y}_t^{\mathrm{obs}}) \tag{17}$$

To illustrate the application of 3D-Var and 4D-Var type schemes, one can perform single observation experiments, where only one observation at the end of the

assimilation window is assimilated. In that case, after a bit of algebra and assuming $\mathcal{M}(x_0 + \delta x, T) = \mathcal{M}(x_0, T) + \mathbf{L}.\delta x$ one can write the optimal state x^a that minimises J as:

$$x^a = x^b + \mathbf{BL}^*\mathbf{H}^* (\mathbf{R} + \mathbf{HLBL}^*\mathbf{H}^*)^{-1} (H_T(\mathcal{M}(x_0, T)) - y_T^{obs})$$

For a single observation experiment, it is easy to see that $(\mathbf{R} + \mathbf{HLBL}^*\mathbf{H}^*)^{-1} (H_T(\mathcal{M}(x_0, T)) - y_T^{obs})$ is a scalar, and when multiplied by $\mathbf{H}^*$ it becomes a vector in the state space, with only one non-zero value (assuming the observation is at a grid point). In 3D-Var formulation, $\mathbf{L}^*$ is approximated by the identity operator, so the correction to the initial condition outside the observed grid point is solely driven by the prescribed background error statistics in $\mathbf{B}$, while in 4D-Var the model dynamics is accounted for through the adjoint model $\mathbf{L}^*$.

An example of such differences is given in Fig. 2 where a single synthetic SSH observation, close to the middle of the regional model, at the end of DA time window, is assimilated using both 3-D-FGAT and incremental 4D-Var algorithms from the NEMOVAR system with NEMO's SEABASS configuration (see Appendix A). The observation misfit value is 0.5 m.

The 3D-Var increment (top figure) shows a perfect gaussian shape, centred around the observation location, with an maximum amplitude close to the observation value. This gaussian shape is exactly what is prescribed in the background error covariance matrix $\mathbf{B}$, and the computed increment is independent of the length of the assimilation window. On the other hand, 4D-Var increment is sensitive to the assimilation window length. Two examples are given, first with a 5 day window (bottom left) and second with a 30 days window. In the first case, the 3D-Var approximation is not that bad, so both 3-D- and 4D-Var are similar, even though the latter is slightly deformed and displaced to account for the short term dynamics. For the longer assimilation window (bottom right) however, the effect of the dynamics is more complex, in particular the non linearities are more developed. As a consequence the 3D-Var approximation is no longer valid and the shape of the optimal correction is completely different.

4.2 Singular vectors

Another application of tangent and adjoint models is the stability analysis, that is the study of perturbations on the system. Particular tools for such analysis are the so-called singular vectors.

We classically define the growth rate of a given perturbation δx_0 by

$$\rho(\delta x_0) = \frac{\|\mathcal{M}(x_0 + \delta x_0, T) - \mathcal{M}(x_0, T)\|}{\|\delta x_0\|} \tag{18}$$

where $\|.\|$ is a given norm.

One can then define the optimal perturbation δx_0^1 so that $\rho\left(\delta x_0^1\right) = \max\limits_{\delta x_0} \rho(\delta x_0)$ and then deduce a family of maximum growth vectors

$$\rho\left(\delta x_0^i\right) = \max\limits_{\delta x_0 \perp \mathrm{Span}\left(\delta x_0^1, \ldots, \delta x_0^{i-1}\right)} \rho(\delta x_0), \quad i \geq 2 \tag{19}$$

By restricting the study to the linear part of the perturbation behaviour, the growth rate becomes (denoting $\mathbf{L} = \mathbf{L}(x, T)$ for clarity).

$$\rho^2(\delta x_0) = \frac{\|\mathbf{L}\delta x_0\|^2}{\|\delta x_0\|^2} = \frac{\langle \mathbf{L}\delta x_0, \mathbf{L}\delta x_0 \rangle}{\langle \delta x_0, \delta x_0 \rangle}$$
$$= \frac{\langle \delta x_0, \mathbf{L}^*\mathbf{L}\delta x_0 \rangle}{\langle \delta x_0, \delta x_0 \rangle} \tag{20}$$

$\mathbf{L}^*\mathbf{L}$ being a symmetric positive definite matrix, its eigenvalues are positive real and its eigenvectors are (or can be chosen) orthogonal. The strongest growth vectors are the eigenvectors of $\mathbf{L}^*\mathbf{L}$ corresponding to the greater eigenvalues. They are called forward singular vectors (FSV).

$$\mathbf{L}^*\mathbf{L}f_i^+ = \mu_i f_i^+ \tag{21}$$

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

$\mathbf{L}f_i^+$ is an eigenvector of $\mathbf{LL}^*$.

The backward singular vectors (BSV), noted f_i^-, are noted by:

$$\mathbf{L}f_i^+ = \sqrt{\mu_i}\, f_i^-$$

5 the eigenvalue corresponding to f_i^- is μ_i as well. FSVs represent the directions of perturbation that will grow fastest, while BSVs represent the directions of perturbation that have grown the most.

The computation of the f_i^+ and f_i^- generally requires numerous matrix-vector multi-plications, i.e. direct integrations of the model and backward adjoint integrations. The
10 result of these calculations depends on the norm used, the time window and the initial state if the model is nonlinear. Examples of such vectors are shown in Fig. 3, they were computed using an energy norm to define $\|.\|$ and the parpack (Lehoucq et al., 1997) external library to perform the singular value decomposition.

These vectors, thanks to the information they contain about the system behaviour,
15 have many applications. Among them, one can cite ensemble forecast, sensitivity stud-ies (Rivière et al., 2009, for a recent application), the order reduction in data assimila-tion (Durbiano, 2001; Blayo et al., 2003), improving the monitoring network (Qin and Mu, 2012) or allow to better select targeted observations (Mu et al., 2009). Apart from the generation of ensemble, however, the potential of these vectors has yet been little
20 exploited.

5 Conclusions

The tangent and adjoint models of NEMO (NEMOTAM) is now available (with respect to the 3.4.1 version of NEMO at the time of writing this paper). It is part of the NEMO-ASSIM tools (Bouttier et al., 2012), which aim is to ease the interface between NEMO
25 code and a most Data Assimilation algorithms. In the few preceding pages, these mod-els, the technical choices made for their developments and their validation were dis-cussed. Additionally some applications were presented as illustration of potential use.

When developing a TAM, two main difficulties have to be addressed: the handling of the non-linearities and of the non-differentiable parts. Indeed non linear equations require the storage and/or recompilation of the non linear trajectory to differentiate around. In NEMOTAM this is done through the so-called checkpointing strategy, con-
5 sisting in saving part of the non-linear trajectory at a given frequency (checkpoints), recomputing the missing part and, if needed, interpolating it linearly between check-points. On the other hand the non-differentiability issues are dealt with using three different approaches approaches were used, depending of the discontinuity nature: numerical or physical regularisation, numerical or physical approximation and non-linear
10 branching.

All these choices have to be validated along with the coding itself. To that end a significant effort has been done in NEMOTAM. First, adjoint tests are systematically implemented for each adjoint routines and gives an exact indication of the validity of the adjoint code. For the tangent linear model, there is no exact test which can strictly vali-
15 date the development. However, comparing the propagation of a small perturbation by TL model and and the direct model gives an idea of the validity of the TL hypothesis. Finally, an estimator of the errors due to the approximations on NEMO non-differentiable parts is also provided for the corresponding routines.

The range of applications using NEMOTAM is wide. To illustrate that, three example
20 applications were exposed. First a local sensitivity analysis with a realistic NEMO configuration. Then a very simple data assimilation experiment, using a single observation is also performed, illustrating the impact of the use of an adjoint model. And finally some singular vectors were computed using NEMOTAM.

The scope of the current NEMOTAM implementation leaves the room for different
25 extensions (e.g. taking in account other NEMO modules, as LIM (sea ice) model, AGRIF, etc.). In order to handle properly the MPP aspects and optimize computing cost, the current NEMOTAM is hand-coded. However, in order to provide more flexibility in the choice of variable to differentiate around and to ease the process of updating the NEMOTAM code with new features and following the evolution of the direct code,

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

a subtle trade-off between automatic differentiation and manual intervention could be very beneficial.

Appendix A: SEABASS configuration

The SEABASS reference configuration for NEMO is mentioned several times in this paper. This configuration is an academic basin presenting a double-gyre circulation. The horizontal domain extends from 24 to 44° N and over 30° in longitude. For a 1/4° horizontal resolution, the grid contains only 121 points in longitude and 81 points in latitude. The time step is 900 s. The ocean is sliced into 11 verticals levels, from surface to 4000 m, described with a z coordinate. The domain is closed and has a flat bottom. Lateral boundaries conditions are frictionless and bottom boundary condition exerts a linear friction. The circulation is only forced by a zonal wind. Lateral dissipation is performed on dynamics and tracers with a biharmonic diffusion operator. The salinity is constant over the whole domain and the initial stratification is produced using an analytical temperature profile. Details can be found in Cosme et al. (2010), for example.

Even if SEABASS is an academic configuration, it exhibits a turbulence level statistically meaningful regarding the eddy activity and the non-linearity amplitude of the actual Gulf Stream system. This SEABASS characteristics is interesting in a data assimilation context, as the oceanic turbulence is one of the major current stakes for data assimilation methods in oceanography.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

References

Blayo, E., Durbiano, S., Vidard, A., and Le Dimet, F.-X.: Reduced order strategies for variational data assimilation in oceanic models, in: Data Assimilation for Geophysical Flows, edited by: Sportisse, B. and Le Dimet, F.-X., 18 pp., Springer-Verlag, 2003. 6722

Bouttier, P.-A., Blayo, E., Brankart, J.-M., Brasseur, P., Cosme, E., Verron, J., and Vidard, A.: Toward a data assimilation system for NEMO, Mercator Ocean Quarterly Newsletter, 24–30, 2012. 6722

Daget, N., Weaver, A. T., and Balmaseda, M. A.: Ensemble estimation of background-error variances in a three-dimensional variational data assimilation system for the global ocean, Q. J. Roy. Meteor. Soc., 135, 1071–1094, 2009. 6707

Deltel, C.: Estimation de la circulation dans l'océan Atlantique Sud par assimilation variationnelle de données in situ. Impact du contrôle optimal des forçages et de l'hydrologie aux frontières ouvertes. 185 pp., Ph.D. thesis, Univ. de Bretagne Occidentale, 2002. 6707

Durbiano, S.: Vecteurs caractéristiques de modèles océaniques pour la réduction d'ordre en assimilation de données, Ph.D. thesis, Université Joseph Fourier, 2001. 6722

Errico, R. M.: What is an adjoint model?, B. Am. Meteorol. Soc., 78, 2577–2591, 1997. 6716

Forget, G., Ferron, B., and Mercier, H.: Combining ARGO profiles with a general circulation model in the North Atlantic. Part 1: Estimation of hydrographic and circulation anomalies from synthetic profiles, over a year, Ocean Model., 20, 1–16, doi:10.1016/j.ocemod.2007.06.001, 2008. 6707

Giering, R. and Kaminski, T.: Recipes for adjoint code construction, ACM T. Math. Software, 24, 437–474, 1998. 6709

Hascoët, L. and Pascual, V.: TAPENADE 2.1 user's guide, Technical Report 0300, INRIA, 2004. 6708

Lawless, A., Nichols, N., and Ballard, S.: A comparison of two methods for developping the linearization of a shallow-water model, Q. J. Roy. Meteor. Soc., 129, 1237–1254, 2003. 6715

Lehoucq, R. B., Sorensen, D. C., and Yang, C.: ARPACK Users' Guide, 1997. 6722

Madec, G.: NEMO ocean engine, IPSL, Note No. 27 du Pôle de Modélisation (LOCEAN-IPSL), 2008. 6707

Madec, G., Delecluse, P., Imbard, M., and Lévy, C.: OPA 8.1 ocean general circulation model reference manual, IPSL, Note No. 11 du Pôle de Modélisation (LODYC), 1998. 6707

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◀ | ▶

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Marotzke, J., Giering, R., Zhang, K. Q., Stammer, D., Hill, C., and Lee, T.: Construction of the adjoint MIT ocean general circulation model and application to Atlantic heat transport sensitivity, J. Geophys. Res., 104, 29529–29547, 1999. 6706

Mogensen, K. S., Balmaseda, M. A., Martin, M., and Vidard, A.: NEMOVAR: a variational data assimilation system for the NEMO ocean model, ECMWF Newsletter, 120, 17–21, 2009. 6708

Moore, A., Vialard, J., Weaver, A., Anderson, D., Kleeman, R., and Johnson, J.: The role of air-sea interaction in controlling the optimal perturbations of low-frequency tropical coupled ocean-atmosphere modes, J. Climate, 16, 951–968, 2003. 6707

Moore, A., Arango, H., Di Lorenzo, E., Cornuelle, B., Miller, A., and Neilson, D.: A comprehensive ocean prediction and analysis system based on the tangent linear and adjoint of a regional ocean model, Ocean Model., 7, 227–258, 2004. 6706

Mu, M., Zhou, F., and Wang, H.: A method for identifying the sensitive areas in targeted observations for tropical cyclone prediction: conditional nonlinear optimal perturbation, Mon. Weather Rev., 137, 1623–1639, 2009. 6722

Qin, X. and Mu, M.: Influence of conditional nonlinear optimal perturbations sensitivity on typhoon track forecasts, Q. J. Roy. Meteor. Soc., 138 185–197, 2012. 6722

Rémy, E.: Assimilation variationnelle de données tomographiques simulées dans des modèles de circulation océanique, Ph.D. thesis, Univ. Paris 6, 1999. 6707

Rivière, O., Lapeyre, G., and Talagrand, O.: A novel technique for nonlinear sensitivity analysis: application to moist predictability, Q. J. Roy. Meteor. Soc., 135, 1520–1537, 2009. 6722

Sévellec, F., Huck, T., BenJelloul, M., Grima, N., Vialard, J., and Weaver, A.: Optimal surface salinity perturbations of the meridional overturning and heat transport in a global ocean general circulation model, J. Phys. Oceanogr., 38, 2739–2754, 2008. 6707

Sévellec, F., Huck, T., BenJelloul, M., and Vialard, J.: Non-normal multidecadal response of the thermohaline circulation induced by optimal surface salinity perturbations, J. Phys. Oceanogr., 39, 852–872, 2009. 6707

Tber, M. H., Hascoet, L., Vidard, A., and Dauvergne, B.: Building the Tangent and Adjoint codes of the Ocean General Circulation Model OPA with the Automatic Differentiation tool TAPENADE, Research Report RR-6372, INRIA, 2007. 6708

Thuburn, J.: Adjoints of nonoscillatory advection schemes, J. Comput. Phys., 171, 616–631, 2001. 6711

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◀ | ▶

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Vidard, A.: Vers une prise en compte de l'erreur modèle en assimilation de données 4D-variationnelle, Ph.D. thesis, Université Joseph Fourier, 2001. 6707

Vidard, A., Rémy, E., and Greiner, E.: Sensitivity analysis through adjoint method: application to the GLORYS reanalysis, Contrat no. 08/D43, Mercator Océan, 2011. 6718

5 Vossepoel, F. C., Weaver, A., Vialard, J., and Delecluse, P.: Adjustment of near-equatorial wind stress with 4D-Var data assimilation in a model of the Pacific Ocean, Mon. Weather Rev., 132, 2070–2083, 2003. 6707

Weaver, A. T., Vialard, J., and Anderson, D. L. T.: Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean. Part I: Formulation,
10 internal diagnostics, and consistency checks, Mon. Weather Rev., 131, 1360–1378, 2003. 6707

Weaver, A. T., Deltel, C., Machu, E., and Ricci, S.: A multivariate balance operator for variational ocean data assimilation, Q. J. Roy. Meteor. Soc., 131, 3605–3625, 2005. 6707

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Table 1. Tangent validity tests for the bn2 routine (computation of the Brünt–Väisälä frequency).

Routine (**L**)	γ	ϵ_γ
bn2_tan	1×10^{0}	0.999961862090
bn2_tan	1×10^{-1}	0.999995878199
bn2_tan	1×10^{-2}	0.999999584740
bn2_tan	1×10^{-3}	0.999999958442
bn2_tan	1×10^{-4}	0.999999995846

Full Screen / Esc

Table 2. Approximation error in the tangent linear model for different configurations over 10 days.

Configurations	$\hat{\mathcal{E}}$
SEABASS no simplification	negligible
SEABASS interpolated checkpoint (1 day^{-1})	negligible
SEABASS simplified TVD	3 %
ORCA2 simplified vertical mixing	1 %
ORCA2 idem + TVD + interp. checkpoint (1 day^{-1})	4 %
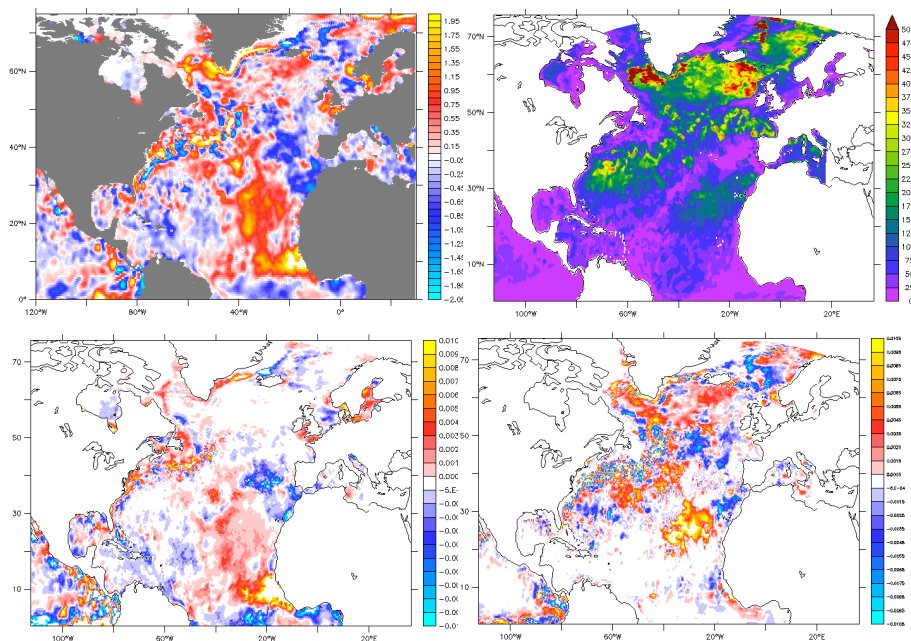
Figure 1. Top: misfit between forecast and observed SST (left) and mixed layer depth (right). Bottom: sensitivity to one week lead time SST error respect to variations in initial surface (left) and 100 m (right) temperature (courtesy E. Rémy, Mercator-Océan).
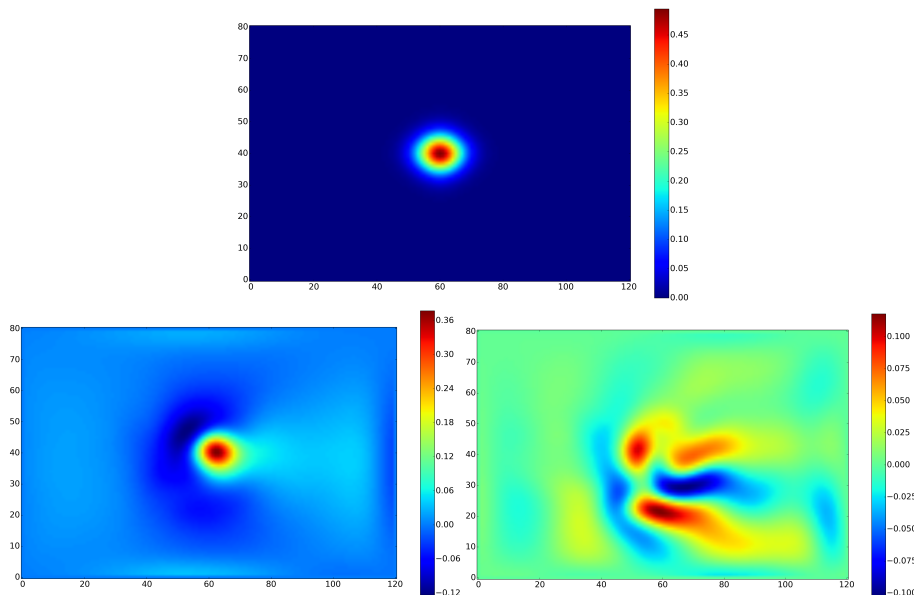
Printer-friendly Version

Interactive Discussion

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Figure 2. Assimilation increments in the SEABASS configuration corresponding to one SSH observation and a misfit of 0.5 m. Coming from NEMOVAR 3-D-FGAT (top) and 4D-Var (bottom) formulation. Two different assimilation widows length are presented for the 4D-Var: 5 days (left) and 30 days (right).
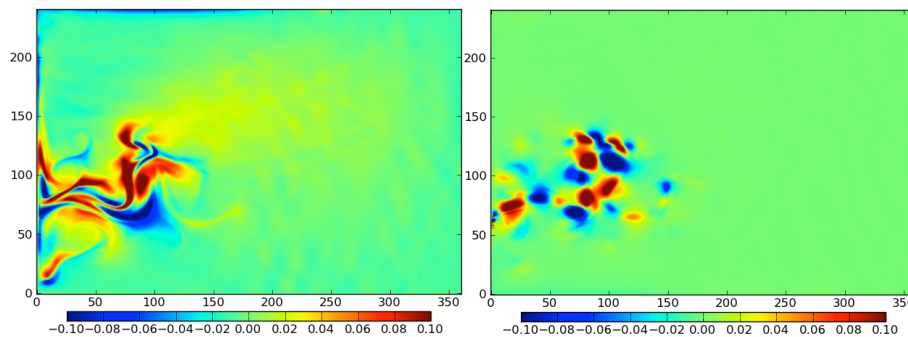
Figure 3. Leading FSV and BVF for a SEABASS configuration at 1/12° and a 10 day window.