DOCUMENTATION DU SCRIPT 4D-VAR

CNRM/GMAP/ALGO

Gérald Desroziers et Pascale Riber

Version du 13 juin 2002

Table des matières

Première partie

Guide d'utilisation

Chapitre 1

Installation du script

Dans ce chapitre, on passe en revue les configurations nécessaires sur les différentes machines. L'installation du script est ensuite décrite.

1.1 Configuration préliminaire

Pour exécuter le script 4D-Var, il faut que vous soyez capable de soumettre un job sur le VPP (kami), que vous ayez un accès sur l'Archive (`delage`) à travers ftuas (ftp sécurisé depuis le VPP) et que vous soyez capable d'effectuer des remote-copy de fichiers entre le VPP et la machine locale (la machine depuis laquelle vous soumettez le job). Vous pouvez aussi considérer le VPP comme votre machine locale. Tout cela nécessite qu'un certain nombre de fichiers de configuration soient placés sur la machine locale et distante (VPP). La liste et le contenu de ces fichiers sont maintenant décrits.

- Fichier **.rhosts** sur la machine locale (si elle est différente du VPP).

Ce fichier contient par exemple les lignes suivantes (noms de kami et de chacun des ses processeurs, associés au user VPP) :

```
kami.meteo.fr mrpm611
kaze.meteo.fr mrpm611
cbar1.meteo.fr mrpm611
cbar2.meteo.fr mrpm611
cbar3.meteo.fr mrpm611
cbar4.meteo.fr mrpm611
cbar5.meteo.fr mrpm611
cbar6.meteo.fr mrpm611
cbar7.meteo.fr mrpm611
cbar8.meteo.fr mrpm611
cbar9.meteo.fr mrpm611
cbar10.meteo.fr mrpm611
cbar11.meteo.fr mrpm611
cbar12.meteo.fr mrpm611
cbar13.meteo.fr mrpm611
cbar14.meteo.fr mrpm611
cbar15.meteo.fr mrpm611
cbar16.meteo.fr mrpm611
cbar17.meteo.fr mrpm611
cbar18.meteo.fr mrpm611
cbar19.meteo.fr mrpm611
cbar20.meteo.fr mrpm611
cbar21.meteo.fr mrpm611
cbar22.meteo.fr mrpm611
cbar23.meteo.fr mrpm611
cbar24.meteo.fr mrpm611
cbar25.meteo.fr mrpm611
```

```
cbar26.meteo.fr mrpm611
cbar27.meteo.fr mrpm611
cbar28.meteo.fr mrpm611
cbar29.meteo.fr mrpm611
cbar30.meteo.fr mrpm611
cbar31.meteo.fr mrpm611
```

- Fichier **.rhosts** sur le VPP (kami).

Il contient le nom de la machine locale et votre user sur cette machine. Par exemple :

```
aad2.cnrm.meteo.fr desroz
```

- Fichier **.ftuas** sur le VPP.

Il est utilisé durant les transferts ftp sécurisés entre le VPP et la machine d'archivage (`delage`). Pour le définir, vous devez exécuter la commande `ftuas -p` sur le VPP, qui vous demandera un mot de passe. Donnez votre mot de passe sur delage et celui-ci sera encrypté dans le fichier `.ftuas`.

1.2 Installation du script sur la machine locale

Choisissez un répertoire sur votre machine locale. Appelons le `LDIR`. Placez vous sous ce répertoire et dé-assemblez (`tar xf`) le fichier `4dvar_*.tar.gz` fourni. Supprimez le fichier `4dvar_*.tar`. Le contenu de `LDIR/4dvar_*` devrait ressembler à quelque chose comme :

```
-rwxr-xr-x   1 desroz     gmap          6810 21 mai 11:26 configure
drwxr-xr-x   2 desroz     gmap          1024 21 mai 11:26 param_cycle
drwxr-xr-x   4 desroz     gmap          1024 21 mai 11:26 script
```

Sous le répertoire `script`, on trouve un certain nombre de fichiers :

```
./script:

drwxr-xr-x   2 desroz     gmap          1024 21 mai 11:26 envir_delta
-rw-r--r--   1 desroz     gmap          2399 21 mai 11:26 file_date
-rw-r--r--   1 desroz     gmap          5949 21 mai 11:26 file_envir
-rwxr-xr-x   1 desroz     gmap         10723 21 mai 11:26 ifs31to41
-rwxr-xr-x   1 desroz     gmap          3214 21 mai 11:26 ifsan
-rwxr-xr-x   1 desroz     gmap          4462 21 mai 11:26 ifsan_assim_job
-rwxr-xr-x   1 desroz     gmap          3750 21 mai 11:26 ifsan_prod_job
-rwxr-xr-x   1 desroz     gmap          3753 21 mai 11:26 ifsbiaisdfi
-rwxr-xr-x   1 desroz     gmap          8047 21 mai 11:26 ifsbincst
-rwxr-xr-x   1 desroz     gmap          4611 21 mai 11:26 ifscans
-rwxr-xr-x   1 desroz     gmap          1079 21 mai 11:26 ifsdateg
-rwxr-xr-x   1 desroz     gmap          1113 21 mai 11:26 ifsdatev
-rwxr-xr-x   1 desroz     gmap          4061 21 mai 11:26 ifsenv
-rwxr-xr-x   1 desroz     gmap          3037 21 mai 11:26 ifsfc
-rwxr-xr-x   1 desroz     gmap          3830 21 mai 11:26 ifsfc96
-rwxr-xr-x   1 desroz     gmap          2674 21 mai 11:26 ifsfc96_job
-rwxr-xr-x   1 desroz     gmap          3337 21 mai 11:26 ifsh2l
-rwxr-xr-x   1 desroz     gmap          3781 21 mai 11:26 ifsl2h
-rwxr-xr-x   1 desroz     gmap          5462 21 mai 11:26 ifsmin131
-rwxr-xr-x   1 desroz     gmap         14413 21 mai 11:26 ifsnamel
-rwxr-xr-x   1 desroz     gmap          5317 21 mai 11:26 ifsobs
-rwxr-xr-x   1 desroz     gmap          2809 21 mai 11:26 ifsobs_assim_job
-rwxr-xr-x   1 desroz     gmap          2895 21 mai 11:26 ifsobs_prod_job
```

```
-rwxr-xr-x   1 desroz     gmap          2943 21 mai 11:26 ifsrad
-rwxr-xr-x   1 desroz     gmap          5679 21 mai 11:26 ifsscreen
-rwxr-xr-x   1 desroz     gmap          3890 21 mai 11:26 ifsstart
-rwxr-xr-x   1 desroz     gmap          3121 21 mai 11:26 ifsstart_job
-rwxr-xr-x   1 desroz     gmap          4244 21 mai 11:26 ifstraj
-rwxr-xr-x   1 desroz     gmap          3771 21 mai 11:26 ifsvar
drwxr-xr-x   2 desroz     gmap          1024 21 mai 11:26 tools
```

Les fichiers suffixés par `job` correspondent aux différents jobs qui seront lancés au cours de l'assimilation. Les fichiers pré-fixés par `ifs` et non suffixés par `job` correspondent à des sous-scripts appelés par ces divers jobs.

Sous le répertoire `tools`, on trouve enfin des fonctions supplémentaires appelées dans les scripts comme par exemple la modification d'une date (`timdat`) :

```
./script/tools:

-rwxr-xr-x   1 desroz     gmap          1154 21 mai 11:26 adate
-rwxr-xr-x   1 desroz     gmap           705 21 mai 11:26 aget
-rwxr-xr-x   1 desroz     gmap           546 21 mai 11:26 aput
-rwxr-xr-x   1 desroz     gmap          1180 21 mai 11:26 asub
-rwxr-xr-x   1 desroz     gmap          1757 21 mai 11:26 copy
-rwxr-xr-x   1 desroz     gmap           346 21 mai 11:26 timdat
-rwxr-xr-x   1 desroz     gmap       3182660 21 mai 11:26 timdat.ex
```

Les permissions d'exécution de certains fichiers sont essentielles. Il est fortement recommandé de les vérifier et de les restaurer si elles ont été perdues au cours du tar ou d'un `ftp`.

1.3 Définition d'une expérience

Vous devez éditer le fichier `configure` et définir les variables d'environnement apparaissant dans la première partie :

```
########################### Begin of user configuration #######################

#
#        ****************************************************************
#        *                         Definitions                          *
#        ****************************************************************
#

LOCAL=aad2.cnrm.meteo.fr                          # name of the workstation
USER=desroz                                       # user login on the workstation
LDIR=4dvar_scripts/4dvar_cy24t1_op6_02_020521.02 # local root directory for 4D-Var
RDIR=4dvar_scripts/4dvar_cy24t1_op6_02_020521.02 # name of the root directory

#
#        ****************************************************************
#        *                  Parameters for QSUB cards                   *
#        ****************************************************************
#

TIMEAN=10800                # max CPU time for analysis (seconds)
VPP_MBX_SIZE=128000000      # memory size used for mpi mailbox
size=4000                   # basis number (~size for one process) used ...
dsize=4872                  # ... for determining the analysis max memory
```

```
size=$(($size / $2))         # roughly the max memory per process in mb
size=$(($size + $(($VPP_MBX_SIZE / 1000000))))
MEMOAN=$(($size + $dsize))   # total max memory for analysis (mb)
MEMODIAG=1792                # total max memory for diagnostics (mb)
TIMEFC=1200                  # max CPU time for 6h forecast (seconds)
TIMEFC72=11000               # max CPU time for 72h forecast (seconds)
TIMEFC96=11000               # max CPU time for 96h forecast (seconds)
TIMEFC06=1000                # max CPU time for starting 6h forecast (seconds)
TIMEDIAG=300                 # max CPU time for diagnostics
TIMECP=3000                  # max CPU time for compilation (seconds)
MEMOCP=420                   # total max memory for compilation (mb)
size=800                     # basis number (~size for one process) used ...
dsize=1772                   # ... for determining the forecast max memory
size=$(( size / $3))         # roughly the max memory per process in mb
size=$(($size + $(($VPP_MBX_SIZE / 1000000))))
MEMOFC=$(($size + $dsize))   # total max memory for forecast (mb) per process


###########################  End of user configuration  ########################
```

Il faut surtout renseigner les paramètres correspondant à la partie "`Definition`". Les variables de la partie "`Parameters for QSUB cards`" n'ont pas à être modifiées sauf nécessité. Veillez à ne rien modifier après la ligne "End of user configuration".

Exécuter le fichier `configure` sous la forme

`configure XID NPROCA NPROCP`,

où `XID` désigne le nom de l'expérience à créer, en 3 lettres, et `NPROCA`, `NPROCP` sont respectivement les nombres de processeurs pour la partie analyse et prévision.

(ex : `configure AAA 3 4`).

Le lancement de cette commande va ouvrir successivement sous `vi` deux nouveaux fichiers de configuration, `file_date` et `file_envir`, qui sont décrits dans le paragraphe suivant. Une fois que vous avez édité ces deux fichiers, un répertoire `XID` est créé sous `LDIR`, qui contient tous les éléments nécessaires au déroulement du 4D-Var.

Ce répertoire devrait contenir quelque chose comme :

```
-rw-r--r--   1 desroz     gmap          2399 21 mai 11:34 file_date
-rw-r--r--   1 desroz     gmap          5949 21 mai 11:34 file_envir
-rw-r--r--   1 desroz     gmap          4501 21 mai 11:34 ifsan_assim_job
-rw-r--r--   1 desroz     gmap          3789 21 mai 11:34 ifsan_prod_job
-rw-r--r--   1 desroz     gmap          2711 21 mai 11:34 ifsfc96_job
-rw-r--r--   1 desroz     gmap          2850 21 mai 11:34 ifsobs_assim_job
-rw-r--r--   1 desroz     gmap          2936 21 mai 11:34 ifsobs_prod_job
-rw-r--r--   1 desroz     gmap          3157 21 mai 11:34 ifsstart_job
```

Les fichiers suffixés par `job` correspondent aux jobs qui seront soumis au cours du cycle d'assimilation et pour lesquels les paramètres `QSUB` et le nombre de processeurs ont été correctement définis en fonction des valeurs passées dans le fichier `configure`.

On retrouve également les deux fichiers de configuration `file_date` et `file_envir` que l'on décrit au paragraphe suivant : vous devez à partir de maintenant travailler depuis le répertoire `XID/LDIR`.

Chapitre 2

Définition et lancement d'une expérience

On appelle expérience une assimilation sur une période et un jeu d'observations donnés et avec une configuration particulière. De manière à adapter le 4D-Var à cette expérience, vous avez à mettre à jour quelques variables d'environnement dans les fichiers `file_date`, `file_envir` et éventuellement `file_envir_prod`.

2.1 Configuration de l'expérience

2.1.1 Configuration de `file_date`

Dans ce fichier, on spécifie la date synoptique de la première analyse ainsi que le nombre d'analyses successives à effectuer.

La partie configurable de `file_date` ressemble à (version du 13 juin 2002) :

```
########################### Beginning of user configuration ####################

DATRAW=2002011700          # AAAAMMJJHH  (2002011700)
NJOB=0                     # Current analysis job number (0)
NMXJOB=1                   # Max number of assimilation jobs (1)
DATEFC=                    # If set, signals the first prediction date
                           # If not set, THE PREDICTION WILL NOT BE DONE

####################### file_date : end of user configuration ##################
```

Les valeurs par défaut des variables sont indiquées entre parenthèse. La signification de ces variables est la suivante.

`DATRAW`	:	date du réseau courant.
`NJOB`	:	Nombre de jobs déjà effectués (les différentes analyses + éventuellement un premier job pour créer la toute première ébauche)
`NMXJOB`	:	nombre total de jobs à effectuer.

Enfin, en spécifiant une date dans `DATEFC`, alors une prévision à 96 h depuis l'analyse "cut-off" court sera effectuée si `DATEFC` est antérieure ou égale à la date d'analyse courante `DATRAW`.

Le fichier `file_date` est mis à jour par le script.

2.1.2 Configuration de `file_envir`

Ce fichier définit la configuration du schéma d'assimilation. Les valeurs par défaut sont indiquées entre parenthèse.

```
######################### Beginning of user configuration ####################

#        ***********************************************************
#        *                   General definitions                    *
#        ***********************************************************

DATECONF=                             # Date (AAAAMMJJHH) for which you want to
                                      # get the corresponding environment
CONF=cy24t1_op6.02                    # Arpege cycle
CHAINE=oper                           # Stream (oper/dbl) for Arpege
                                      # configuration (oper)

ARPEGE=                               # If set, overrides the GCO arpege executable
NAM_PERSO=                            # If set, personal set of namelist in NAM_perso directory

CHAINE_FK=oper                        # Obs from the (oper|dbl) stream ? (oper)

PRODUCTION=no                         # yes=perform production cycle (no)
P00ONLY=yes                           # yes=prod. analysis at 00H only (yes)
FGPROD=no                             # FGPROD != no or FGPROD != "" : (no)
                                      # assim fg given by prod analysis
                                      # FGPROD = number of the prod minim

#        ***********************************************************
#        *                  Outer model definition                  *
#        ***********************************************************

RESOLTRAJ=                            # Outer resolution for 3D/4D-Var
RSTRET=                               # Outer stretching coefficient

LEVEL=                                # Number of vertical levels
TSCHEME=                              # Time integration scheme :
                                      #  - sli, for semi-lagrangian integration (default)
                                      #  - eul, for eulerian time integration
DTIME=                                # Time step (s)

#        ***********************************************************
#        *          Assimilation and inner models definition        *
#        ***********************************************************

PERIOD_4D=                            # Length of the assimilation period (h) (6 h)
MXUP_TRAJ=                            # Max number of updates

TLTSCHEME=                            # Inner model timestepping: sli or eul

DTIMEINNER1=                          # Inner model time steps (s)
RESOLINNER1=                          # Inner resolutions
L3DFGAT1=                             # yes = identity model for TL and AD
ITERATIONS1=                          # Max. iterations in the minimizations
SIMULATIONS1=                         # Max. simulations in the minimizations
ALGOMIN1=                             # Minim. Algo. (1 for N1CG1, 3 for M1QN3)
LSPHYSINNER1=                         # yes = use Marta's simplified physics
LRAYSPINNER1=                         # yes = include radiation

LWRCKY=                               # yes=use Karim Yessad's radiation coefficients

#        ***********************************************************
#        *      External Digital Filter Initialization parameters    *
```

```
#       **************************************************************

LDFI=                               # yes=DFI initialization
NSTDFI=                             # Number of DFI time steps ( (PERIOD_4D/DTIME)/2 )
TAUS=                               # cut-off frequency (s)

#       **************************************************************
#       *      Internal Digital Filter Initialization parameters     *
#       **************************************************************

LDFIINNER=                          # yes=DFI initialization as a weak constraint
TAUSINC=                            # cut-off frequency for Jc-DFI

#       **************************************************************
#       *                   Diagnostics to be run                    *
#       **************************************************************

LDDHLF=                             # yes=compute and save DDH (defaut=yes)

################## file_envir : end of user configuration ######################
```

Les variables d'environnement de votre expérience sont regroupées en 6 classes, que l'on passe maintenant en revue.

"General definitions"

On spécifie ici l'environnement général de l'assimiliation.

Bien que le script 4D-Var soit un outil de recherche, il est apparu nécessaire que ce script puisse reproduire une configuration opérationnelle en utilisant en particulier la bonne version de l'exécutable Arpège, les namelists correspondant à chacune des configurations lancées dans le script mais aussi les bons fichiers de constantes.

Ceci est possible en demandant au script de récupérer l'environnement opérationnel (ou de la chaine en double) correspondant à une date donnée `DATECONF`. Il est également possible de donner explicitement le nom d'un cycle dans `CONFIG` (par exemple `cy24t1_op6.02`) ; ne pas spécifier de date dans `DATECONF` dans ce cas.

La liste des cycles peut être trouvée sur la page Web de GCO
`http ://mout.meteo.fr/ marp001/cycles/op/index.html`.

Le choix chaine opérationnel ou en double se fait en donnant la valeur désirée (`oper/dbl`) dans `CHAINE`.

Le nom de l'exécutable Arpège que l'on veut utiliser peut être donné dans `ARPEGE` (nom de fichier binaire avec son chemin complet sur le VPP).

De la même manière, on peut décider de travailler à partir de namelists personnelles en spécifiant `NAM_PERSO=yes`. Dans ce cas il faut créer un répertoire `NAM_perso` sous le répertoire de l'expérience (`LDIR/XID`) et placer **toutes** les namelists nécessaires au script sous ce répertoire.

La variable `CHAINE_FK` permet de spécifier où prendre les observations : dans la chaine opérationnelle (oper) ou en double (dbl).

La variable `PRODUCTION` indique si l'on désire effectuer un cycle "production" ("cut-off" court) en plus du cycle assimilation ("cut-off" long).

Si `PRODUCTION=yes`, alors il est également possible de faire démarrer la minimisation "cut-off" long de la solution fournie par la minimisation "cut-off" court. Dans ce cas, la variable `FGPROD` doit être différente de la chaine vide ou de `no` et `FGPROD` est un entier contenant l'indice de la solution produite par le cut-off court (si par exemple le nombre d'itérations externes de la minimisation "cut-off" court est 2, alors on peut choisir de démarrer la minimisation "cut-off" long de la solution issue de la première ou de la deuxième minimisation "cut-off" court).

!!! Pour l'instant cette possibilité demande une modification du code et donc la génération d'un nouvel exécutable.

"Outer model definition"

Paramètres liés à la trajectoire à haute résolution du modèle.

"Assimilation and inner models definition"

Choix de la durée de la période d'assimilation, du nombre de mise à jour de la trajectoire (égal au nombre de minimisations) et du schéma temporel du modèle assimilateur. Chacun des paramètres doit être suffixé par l'indice correspondant de la minimisation.

On trouve ensuite autant de blocs de paramètres additionnels qu'il y a de minimisations. A noter la possibilité d'effectuer une minimisation en mode 3D-Var FGAT.

"External Digital Filter Initialization parameters"

Paramètres liés à l'initialisation externe par Filtres Digitaux.

"Internal Digital Filter Initialization parameters"

Paramètres liés à l'initialisation interne par Filtres Digitaux (Jc-DFI).

"Diagnostics to be run"

Permet d'effectuer la prévision 96 h avec un calcul des DDH (et de les archiver). Le défaut est `yes`.

2.1.3 Configuration par modification de namelist

A noter qu'il est de plus possible d'apporter des modifications limitées aux namelists opérationnelles ou fournies dans le répertoire `NAM_perso` en créant un nouveau répertoire `NAMELIST_delta` toujours sous `LDIR/XID` sous lequel on placera la ou les namelists que l'on veut modifier en les suffixant par `_delta`.

Par exemple pour effectuer le test du gradient de la fonction coût dans la première minimisation, on pourra placer sous `NAMELIST_delta` le fichier `namelistmin1311_delta` contenant

```
&NAMVAR
  LTEST=.TRUE.
/
```

2.1.4 En résumé

Le job d'assimilation est configurable de la manière suivante.

⇒ Si `DATECONF` est spécifiée alors l'environnement opérationnel ou de la chaine en double correspondant à cette date est utilisé.

⇒ Sinon, c'est le nom du cycle Arpège passé dans `CONFIG` qui fixe l'environnement.

⇒ Tout en s'appuyant sur l'environnement ainsi défini, il est possible de spécifier un exécutable particulier dans `ARPEGE`.

⇒ De même, on peut choisir d'utiliser un jeu de namelists particulier en spécifiant `NAM_PERSO=yes` et en plaçant l'ensemble des namelists nécessaires au job sous le répertoire `NAM_perso` (sous le répertoire de l'expérience sur la machine locale).

⇒ Le déroulement du job et le contenu des différentes namelists pourront ensuite être modifés en fonction des paramètres (autres que les définitions générales précédentes) passés dans `file_envir`. C'est en particulier le cas pour l'option `L3DFGAT=yes` (voir la description de cette formulation plus loin dans la documentation).

⇒ Enfin, les namelists produites jusqu'à ce stade peuvent encore être modifiées en ajoutant des modifications dans le répertoire `NAMELIST_delta` comme on vient de l'indiquer un peu plus haut.

2.1.5 Configuration éventuelle de `file_envir_prod`

Si `PRODUCTION=yes` dans `file_envir`, alors un fichier de configuration de l'assimilation "cut-off" court (cycle production) est également créé. Ce fichier de nom `file_envir_prod` permet de spécifier une configuration éventuellement différente pour l'analyse "cut-off" court.

2.1.6 Configuration par défaut

A noter qu'il est suffisant de spécifier une des deux variables `DATECONF` ou `CONF`. Toutes les autres variables seront ensuite initialisées par défaut avec les valeurs correspondant à la configuration Arpège choisie.

2.2 Fabriquer un exécutable

Pour fabriquer un exécutable, il faut désormais utiliser l'utilitaire **gmkpack** développé par Ryad El Khatib. Une documentation de **gmkpack** se trouve sous

http ://intra.cnrm.meteo.fr/gmap/MODELES/SUPPORT/mkpack/gmkpack.html .

2.3 Lancer l'expérience

2.3.1 Cas général

Si vous n'avez pas d'ébauche particulière à utiliser pour votre première (et éventuellement unique) analyse, alors la manière la plus simple d'initialiser le cycle d'assimilation est de spécifier dans `file_date` la date du réseau d'analyse précédent le réseau que vous voulez effectivement analyser (décalé de 6 heures jusqu'à présent) puis de lancer le job `ifsstart_job`, depuis le répertoire `LDIR/XID`, qui effectuera la prévision nécessaire à la fabrication de cette ébauche. Dans ce cas il faudra aussi spécifier `NJOB = 0` dans le fichier `file_date` et `NMXJOB > 1` si vous voulez effectuer au moins une analyse.

2.3.2 Ebauche déjà disponible

Si vous disposez déjà d'une ébauche, alors il est possible d'utiliser cette ébauche sans avoir à effectuer la prévision 6 heures depuis le réseau précédent. Dans ce cas il ne faut pas lancer le job `ifsstart_job`, mais le job `ifsobs_assim_job`.

L'ébauche doit être fournie pour deux échéances : 3 et 6 heures, avec pour nom `icmsharpe+0003` et `icmsharpe+0006`. Ces fichiers doivent être placés sur la machine d'archivage sous le répertoire

`RDIR/XID/assim/AAAA/MM/JJ/rRR`, où `AAAA/MM/JJ/rRR` correspond à la date du réseau précédent. On anticipe ici sur la description du rangement des fichiers au cours du job (voir le paragraphe 4.5).

Deuxième partie

Comment ça marche ?

Chapitre 3

Formalisme général

3.1 Assimilation 4D-Var

3.1.1 Le formalisme 4D-Var standard

Le problème 4D-Var consiste à chercher la trajectoire du modèle qui ajuste au mieux une ébauche et des observations disponibles durant la période d'assimilation choisie.

Ceci revient à trouver l'état $\boldsymbol{x}_0$ du modèle à l'instant initial t_0 (variable de contrôle du problème) qui minimise la fonction coût

$$\begin{aligned}
J(\boldsymbol{x}_0) &= J^b(\boldsymbol{x}_0) + J^0(\boldsymbol{x}_0) \\
&= \frac{1}{2}[\boldsymbol{x}_0 - \boldsymbol{x}^b]^T \boldsymbol{B}^{-1}[\boldsymbol{x}_0 - \boldsymbol{x}^b] \\
&\quad + \frac{1}{2}\sum_{i=0}^{n}[H_i(M_i(\boldsymbol{x}_0)) - \boldsymbol{y}_i^o]^T \boldsymbol{R}_i^{-1}[H_i(M_i(\boldsymbol{x}_0)) - \boldsymbol{y}_i^o],
\end{aligned}$$

où
$\boldsymbol{x}^b$ est l'ébauche spécifiée à l'instant initial t_0,
$\boldsymbol{B}$ est la matrice de covariance d'erreur de prévision,
M_i désigne l'intégration du modèle entre t_0 et t_i,
$\boldsymbol{y}_i^o$ est le vecteur des observations disponibles à l'instant t_i,
H_i est l'opérateur d'observation associé,
$\boldsymbol{R}_i$ est la matrice de covariance des erreurs d'observation.

3.1.2 La formulation incrémentale

La formulation incrémentale revient à écrire la solution sous la forme $\boldsymbol{x}^a = \boldsymbol{x}^b + \delta\boldsymbol{x}$, où $\delta\boldsymbol{x}$ est l'incrément d'analyse que l'on considère comme nouvelle variable de contrôle.

Dans ce cas, le problème peut être linéarisé autour de l'ébauche et on peut écrire une nouvelle fonction coût de $\delta\boldsymbol{x}$:

$$J(\delta\boldsymbol{x}) = \frac{1}{2}(\delta\boldsymbol{x})^T \boldsymbol{B}^{-1}\delta\boldsymbol{x} + \frac{1}{2}\sum_{i=0}^{n}[\boldsymbol{H}_i\boldsymbol{M}_i\delta\boldsymbol{x} - \boldsymbol{d}_i]^T \boldsymbol{R}_i^{-1}[\boldsymbol{H}_i\boldsymbol{M}_i\delta\boldsymbol{x} - \boldsymbol{d}_i],$$

où $\boldsymbol{H}_i$ et $\boldsymbol{M}_i$ sont respectivement les opérateurs linéaires-tangents de H_i et M_i et $\boldsymbol{d}_i$ est le vecteur des innovations à l'instant t_i défini par $\boldsymbol{d}_i = \boldsymbol{y}_i^o - H_i(M_i(\boldsymbol{x}^b))$ (car $H_i(M_i(\boldsymbol{x}_0)) - \boldsymbol{y}_i^o = H_i(M_i(\boldsymbol{x}^b + \delta\boldsymbol{x})) - \boldsymbol{y}_i^o \simeq H_i(M_i(\boldsymbol{x}^b)) + \boldsymbol{H}_i\boldsymbol{M}_i\delta\boldsymbol{x} - \boldsymbol{y}_i^o$).

De manière à réduire le coût du problème on introduit une simplification de cet incrément en lui appliquant un opérateur S correspondant à une réduction de la résolution. Si $\boldsymbol{S}$ est l'opérateur linéaire tangent de S autour de l'ébauche $\boldsymbol{x}^b$, alors on peut définir l'incrément d'analyse simplifié $\delta\boldsymbol{w} = \boldsymbol{S}\ \delta\boldsymbol{x}$. Dans ce cas, on peut également approximer l'analyse avec

$$\boldsymbol{x}^a = x^b + \boldsymbol{S}^{-I}\delta\boldsymbol{w},$$

où $\boldsymbol{S}^{-I}$ désigne l'inverse généralisé de $\boldsymbol{S}$ ($\boldsymbol{S}$ n'a naturellement pas de vrai inverse et $\boldsymbol{S}^{-I}$ correspond simplement à un retour dans l'espace à pleine résolution). On ré-écrit alors la fonction coût sous la forme

$$J(\delta\boldsymbol{w}) = \frac{1}{2}(\delta\boldsymbol{w})^T(\boldsymbol{B}^w)^{-1}\delta\boldsymbol{w} + \frac{1}{2}\sum_{i=0}^{n}[\boldsymbol{H}_i\boldsymbol{M}_i\boldsymbol{S}^{-I}\delta\boldsymbol{w} - \boldsymbol{d}_i]^T(\boldsymbol{R}^w)_i^{-1}[\boldsymbol{H}_i\boldsymbol{M}_i\boldsymbol{S}^{-I}\delta\boldsymbol{w} - \boldsymbol{d}_i],$$

où $\boldsymbol{B}^w$ représente la matrice de covariances des erreurs de prévision dans l'espace simplifié et $(\boldsymbol{R}^w)_i$ les matrices d'erreur d'observation dans ce nouvel espace qui devraient être différentes des $(\boldsymbol{R})_i$ dans l'espace non simplifié (en pratique, on n'introduit pas de différence entre ces représentations dans le schéma d'analyse actuel).

Une nouvelle simplification est introduite en supposant que $\boldsymbol{G}_i\boldsymbol{L}_i$ est une approximation correcte de $\boldsymbol{H}_i\boldsymbol{M}_i\boldsymbol{S}^{-\boldsymbol{I}}$, où $\boldsymbol{G}_i$ et $\boldsymbol{L}_i$ sont respectivement les opérateurs linéaires-tangents des opérateurs d'observation et du modèle dans l'espace simplifié.

Il faut noter que cette approximation dépend de la trajectoire autour de laquelle on écrit ces opérateurs linéaires-tangents : dans le cas du script 4D-Var, cette trajectoire est donnéee par l'intégration du modèle non-linéaire complet (avec en particulier la physique complète) mais à basse résolution depuis $S(\boldsymbol{x}^b)$ (l'ébauche à l'instant initial t_0 projetée dans l'espace à basse résolution).

Avec ces approximations la fonction coût peut être de nouveau reformulée :

$$J(\delta\boldsymbol{w}) \quad = \quad \frac{1}{2}(\delta\boldsymbol{w})^T(\boldsymbol{B}^w)^{-1}\delta\boldsymbol{w} + \frac{1}{2}\sum\nolimits_{i=0}^{n}[\boldsymbol{G}_i\boldsymbol{L}_i\delta\boldsymbol{w} - \boldsymbol{d}_i]^T(\boldsymbol{R}^w)_i^{-1}[\boldsymbol{G}_i\boldsymbol{L}_i\delta\boldsymbol{w} - \boldsymbol{d}_i].$$

Cette formulation, qui ne fait intervenir que l'intégration d'un modèle à basse résolution permet de rendre la formulation 4D-Var applicable avec la puissance de calcul actuelle.

De manière à se rapprocher du problème 4D-Var initial, un certain nombre de remises à jour de la trajectoire sont effectuées à partir du nouvel état initial. Si l'on note $\boldsymbol{x}_k$ la $k$ième trajectoire à l'instant initial t_0, cette remise à jour s'écrit simplement

$\boldsymbol{x}_k = \boldsymbol{x}_{k-1} + \boldsymbol{S}^{-\boldsymbol{I}}\delta\boldsymbol{w}_k$,

où $\delta\boldsymbol{w}_k$ est l'incrément d'analyse à basse résolution produit par la minimisation k.

Le problème à minimiser à l'itération externe k de ce processus s'écrit alors

$$J(\delta\boldsymbol{w}_k) \quad = \quad \frac{1}{2}(\delta\boldsymbol{w}_k)^T(\boldsymbol{B}^w)^{-1}\delta\boldsymbol{w}_k + \frac{1}{2}\sum\nolimits_{i=0}^{n}[\boldsymbol{G}_i\boldsymbol{L}_i\delta\boldsymbol{w}_k{}^* - \boldsymbol{d}_i]^T(\boldsymbol{R}^w)_i^{-1}[\boldsymbol{G}_i\boldsymbol{L}_i\delta\boldsymbol{w}_k{}^* - \boldsymbol{d}_i],$$

A chaque itération externe, les opérateurs linéaires-tangents $\boldsymbol{G}_i$ et $\boldsymbol{L}_i$ sont construits autour de la trajectoire issue de $\boldsymbol{w}^b + \delta\boldsymbol{w}_{k-1}$, mais les écarts $\boldsymbol{d}_i$ aux observations sont donnés par $\boldsymbol{d}_i = \boldsymbol{y}_i^o - H_i(M_i(\boldsymbol{x}_{k-1}))$, c'est-à-dire par la différence entre les observations et la précédente trajectoire à haute résolution (avec $\boldsymbol{x}_{k=0} = \boldsymbol{x}^b$).

Le vecteur $\delta\boldsymbol{w}_k{}^*$ correspond à l'incrément ajouté par la nouvelle minimisation k, c'est-à-dire $\delta\boldsymbol{w}_k{}^* = (\boldsymbol{w}^b + \delta\boldsymbol{w}_k) - (\boldsymbol{w}^b + \delta\boldsymbol{w}_{k-1})$ (c'est ce qui est fait dans le sous-programme `subfgs`).

La première minimisation est démarrée avec $\delta\boldsymbol{w}_0 = \boldsymbol{0}$ et les suivantes avec l'incrément obtenu à la minimisation précédente $\delta\boldsymbol{w}_{k-1}$.

3.1.3 Pré-conditionnement du problème

En pratique, on effectue un nouveau changement de variable pour accélérer la convergence de chacune des minimisations. Cette nouvelle variable s'écrit

$$\boldsymbol{\chi} = \boldsymbol{C}\delta\boldsymbol{w},$$

ou encore

$$\delta\boldsymbol{w} = \boldsymbol{C}^{-1}\boldsymbol{\chi} = (\boldsymbol{B}^w)^{1/2}\boldsymbol{\chi},$$

où $\boldsymbol{C}$ désigne le changement de variable et $\boldsymbol{C}^{-1}$ son inverse.

Chacune des minimisations s'effectue en $\boldsymbol{\chi}$ et est démarrée du $\boldsymbol{\chi}_{k-1}$ obtenu à l'itération précédente. Par contre un retour à la variable $\delta\boldsymbol{w}$ est nécessaire au début de la minimisation pour construire le point de départ de la trajectoire de linéarisation (avec $\boldsymbol{w}_k = \boldsymbol{w}^b + \boldsymbol{C}^{-1}\boldsymbol{\chi}_{k-1}$) puis à chaque itération de la minimisation pour calculer J^o (ainsi que l'adjoint de $\boldsymbol{C}^{-1}$ pour déterminer le gradient de J^o).

3.1.4 Pré-conditionneurs de la minimisation

Les deux algorithmes utilisés dans le schéma d'assimilation (m1qn3 : quasi-Newton et n1cg1 : gradient conjugué) construisent au cours d'une minimisation un pré-conditionneur du problème qui peut être réutilisé à la minimisation suivante pour accélérer sa convergence. En pratique il s'agit de construire des paires de vecteurs, de même taille que la variable de contrôle $\boldsymbol{\chi}$ que l'on stocke à la fin d'une minimisation et que l'on récupère au début de la minimisation suivante.

A noter que ceci n'est utile que si l'on effectue au moins deux minimisations avec le même algorithme de minimisation, ce qui n'est pas le cas pour la version opérationnelle du script correspondant à la date de rédaction de cette documentation (13 juin 2002).

3.1.5 Mise à jour de la trajectoire

L'opérateur de simplication correspond ici à la fois à un changement de géométrie (passage d'une géométrie étirée à une géométrie non étirée) et une diminution de la troncature de base. Ceci se traduit par une réduction très importante de la résolution dans la zone d'intérêt d'Arpège (T161 / (T298 × C3.5 = T1043) pour la version opérationnelle à la date de rédaction de cette note (13 juin 2002)).

L'opérateur S correspond en fait à une application de FULL-POS (FP) à l'ensemble des variables du modèle et le pseudo-inverse S^{-I} de S à une application de FULL-POS dans l'autre sens (FP^{-1}). D'autre part il n'existe pas jusqu'à présent de version linéaire-tangente de FULL-POS.

Les mises à jour de la trajectoire s'écrivent donc en fait sous la forme

$$\boldsymbol{x}_k = \boldsymbol{x}_{k-1} + FP^{-1}(\boldsymbol{w}_k^a) - FP^{-1}(\boldsymbol{w}_k^b),$$

avec $\boldsymbol{w}_k^b = FP(\boldsymbol{x}^b) + \delta\boldsymbol{w}_{k-1}$ et $\boldsymbol{w}_k^a = \boldsymbol{w}_k^b + \delta\boldsymbol{w}_k$ (à la première mise à jour, on a naturellement $\delta\boldsymbol{w}_{k-1} = \delta\boldsymbol{w}_0 = \mathbf{0}$).

3.1.6 La formulation multi-incrémentale

La formulation multi-incrémentale est une généralisation de la formulation incrémentale, dans laquelle on augmente progressivement la résolution de l'incrément d'une minimisation externe à une autre. On réduit ainsi le coût final de l'assimilation tout en espérant obtenir une solution voisine de l'analyse qui serait produite à la résolution la plus forte de l'incrément pour toutes les minimisations.

En pratique, cela revient à définir un opérateur de simplification S_k différent pour chaque minimisation, c'est-à-dire une version de FULL-POS correspondant à chaque résolution de l'incrément, avec les fichiers CLIM associés.

Pour initialiser une nouvelle minimisation à résolution plus forte, il faut faire passer la variable de contrôle $\boldsymbol{\chi}_{k-1}$ obtenue à la minimisation précédente à cette nouvelle résolution. En supposant que les deux incréments ont cependant la même géométrie (non basculée, non étirée jusqu'à présent), ceci se fait jusqu'à présent par un "zero padding" de la partie de la variable de contrôle manquante (pseudo-inverse de l'opérateur de troncature).

Si un préconditionneur existe pour une nouvelle minimisation effectuée avec le même minimiseur, on applique le même pseudo-inverse de l'opérateur de troncature aux paires de vecteurs générées à la minimisation précédente.

3.1.7 Initialisation par Filtres Digitaux

On cherche à produire un état initialisé au milieu de la période d'assimilation, c'est-à-dire à $t_n/2$. L'application des Filtres Digitaux consiste à construire cet état comme une moyenne des états antérieurs et postérieurs sous la forme

$$\overline{\boldsymbol{x}(t_n/2)} = \sum_{i=-N}^{N} \alpha_i \boldsymbol{x}(t_0/2 + i\Delta t),$$

où Δt est le pas de temps du modèle.

Ceci revient à atténuer les structures dont la durée de vie est inférieure à un certain temps et que l'on assimile aux ondes de gravité. Les poids α_i correspondent à un filtre particulier. On utilise ici un filtre non-récursif de DOLPH-CHEBYSHEV. La réponse de ce filtre est controlée par deux paramètres. Le premier paramètre permet de donner la période $T_s = 2N\Delta t$ sur laquelle se fait la sommation ("span" du filtre) : on donne $2N$ dans `NSTDFI` apparaissant dans la namelist `NAMDFI`. Le deuxième paramètre est la période de coupure du filtre τ_s ("stop-band edge period", correspondant au paramètre `TAUS` dans `NAMDFI`).

3.1.8 Initialisation interne par Filtres Digitaux

Un terme supplémentaire J^c basé sur les Filtres Digitaux a été introduit dans la fonction coût ($J^c - DFI$). Son principe est très simple. Ce terme contraint la perturbation à l'ébauche $\delta\boldsymbol{w}$ à rester voisine de cette même perturbation initialisée au milieu de la période d'assimilation $t_n/2$. Ce terme s'écrit

$$Jc(\delta\boldsymbol{w}(t_0)) = \frac{1}{2} < \delta\boldsymbol{w}(t_n/2) - \overline{\delta\boldsymbol{w}(t_n/2)}, \delta\boldsymbol{w}(t_n/2) - \overline{\delta\boldsymbol{w}(t_n/2)} >$$

avec

$$\overline{\delta \boldsymbol{w}(t_n/2)} = \sum \alpha_i \delta \boldsymbol{w}_i,$$

et $\delta \boldsymbol{w}_i = \boldsymbol{L}_i \delta \boldsymbol{w}(t_0)$. Le calcul de ce terme supplémentaire est très peu coûteux dans le cadre 4D-Var

On utilise là-encore un filtre de DOLPH-CHEBYSHEV avec une emprise T_s du filtre égale à la période d'assimilation (on prend pour `NSTDFI` le nombre de pas de temps utilisé pour intégrer le modèle sur la période d'assimilation). En effet, on a intérêt à utiliser un T_s le plus grand possible pour avoir un filtre le plus sélectif possible à τ_s fixé (ceci n'est pas fait dans l'initialisation externe, que l'on décrit plus loin, pour des raisons de coût). La période de coupure du filtre τ_s est fixée égale à 3 h (version opérationnelle du 13 juin 2002).

Après une mise à jour de la trajectoire, le terme J^c ne porte que sur l'incrément supplémentaire apportée par la nouvelle minimisation k, c'est-à-dire sur $\delta \boldsymbol{w}_k - \delta \boldsymbol{w}_{k-1}$ (en supposant que la trajectoire $\boldsymbol{x}^a_{k-1}$ est filtrée).

A noter, qu'il est toujours possible de revenir, en principe, à l'ancienne formulation basée sur les modes normaux.

3.1.9 Initialisation externe par Filtres Digitaux

Bien que la première version opérationnelle du 4D-Var n'incluait qu'une initialisation interne par Filtres Digitaux, une initialisation externe après analyse a été ajoutée pour pallier des problèmes de précipitations trop fortes. Ceci introduit cependant un surcoût de calcul et a tendance à dégrader l'ajustement aux observations qui est mieux respecté avec la formulation interne.

L'initialisation externe utilisée est de type incrémentale, c'est-à-dire que l'on applique un Filtrage Digital à l'incrément.

Si l'on note $\boldsymbol{x}^i(t_0/2)$ l'analyse initialisée, cela s'écrit

$$\begin{aligned}
\boldsymbol{x}^i(t_0/2) &= \boldsymbol{x}^b(t_0/2) + \overline{\delta \boldsymbol{x}(t_0/2)} \\
&= \boldsymbol{x}^b(t_0/2) + \overline{\boldsymbol{x}^a(t_0/2) - \boldsymbol{x}^b(t_0/2)} \\
&= \overline{\boldsymbol{x}^a(t_0/2)} + (\boldsymbol{x}^b(t_0/2) - \overline{\boldsymbol{x}^b(t_0/2)})
\end{aligned}$$

Le dernier terme de cette expression correspond à un biais de l'ébauche. Ce biais est calculé dans une phase préliminaire, au cours du "screening" correspondant au calcul des écarts observations-ébauche et au contrôle des données. Le biais est archivé dans le fichier `biais_dfi` et ré-utilisé au cours de la dernière mise à jour de la trajectoire.

Contrairement au cas $J^c - DFI$, le paramètre `NSTDFI`, passé dans les namelists du screening et de la dernière mise-à-jour de la trajectoire, correspond dans ce cas à la moitié des pas de temps faits en avant.

3.2 Assimilation 3D-Var FGAT

3.2.1 Formulation du 3D-Var FGAT

Malgré la formulation incrémentale, le coût du 4D-Var est très important à cause des intégrations répétées du modèle linéaire-tangent et de son adjoint nécessaires à la minimisation de la fonction coût. Une simplification du problème 4D-Var revient à n'effectuer l'analyse qu'à un instant donné (le milieu de la période d'assimilation 4D-Var) : ceci conduit à la formulation 3D-Var dans laquelle on omet les intégrations du modèle.

Il existe aussi une formulation hybride entre 4D-Var et 3D-Var dans laquelle on introduit le modèle pour comparer l'ébauche évoluée aux observations distribuées le long de la période d'assimilation. Cette formulation est désignée sous le terme 3D-Var FGAT (3D-Var with First Guess at Appropriate Time).

Dans ce cas, on cherche à trouver l'incrément simplifié $\delta \boldsymbol{w}$ qui minimise la fonction coût suivante :

$$J(\delta \boldsymbol{w}) \;=\; \frac{1}{2}(\delta \boldsymbol{w})^T (\boldsymbol{B}^w)^{-1} \delta \boldsymbol{w} + \frac{1}{2}\sum_{i=0}^{n} [\boldsymbol{y}_i^o - H_i(M_i(\boldsymbol{x}^b)) - \boldsymbol{G}_i \delta \boldsymbol{w}]^T (\boldsymbol{R}^w)_i^{-1} [\boldsymbol{y}_i^o - H_i(M_i(\boldsymbol{x}^b)) - \boldsymbol{G}_i \delta \boldsymbol{w}],$$

ce qui revient à omettre l'évolution de l'incrément par le modèle simplifié linéaire-tangent $\boldsymbol{L}$, ou à remplacer formellement ce modèle par l'identité. C'est ce qui est fait en pratique dans le code Arpège/IFS lorsque l'on spécifie `L3DFGAT=.TRUE.` dans la namelist `NAMVAR`). Ceci permet aussi de conserver une structure des fichiers d'observations identique à celle utilisée en mode 4D-Var (éclatement du fichier sur des tranches horaires de 1 h).

A noter également que dans ce cas, les opérateurs d'observation sont linéarisés autour de l'ébauche intégrée à basse résolution.

3.2.2 Initialisation interne

Dans le cas du 3D-Var FGAT, l'initialisation interne par Filtres Digitaux n'a plus de sens, car le modèle linéaire-tangent n'est pas intégré au cours des itérations de l'algorithme de minimisation. Si l'on veut conserver un terme de pénalisation des ondes rapides au cours du processus de minimisation, on doit revenir à l'ancienne formulation du terme J^c basée sur une décomposition en modes de Rossby et de gravité. Ce terme s'écrit

$$J^c(\delta \boldsymbol{x}_k) = \epsilon \|(\dot{\boldsymbol{x}_k} - \dot{\boldsymbol{x}^b})_G\|^2,$$

avec $\boldsymbol{x}_k = \boldsymbol{x}_{k-1} + \delta \boldsymbol{x}_k$ et ϵ une constante.

$\dot{\boldsymbol{x}_k}$ et $\dot{\boldsymbol{x}^b}$ représentent respectivement les tendances de $\boldsymbol{x}_k$ et $\boldsymbol{x}^b$ et l'indice G se rapporte à la partie modes de gravité de la différence entre ces tendances.

La norme au carré $\| \|^2$ est basée sur une somme pondérée des carrés des coefficients spectraux.

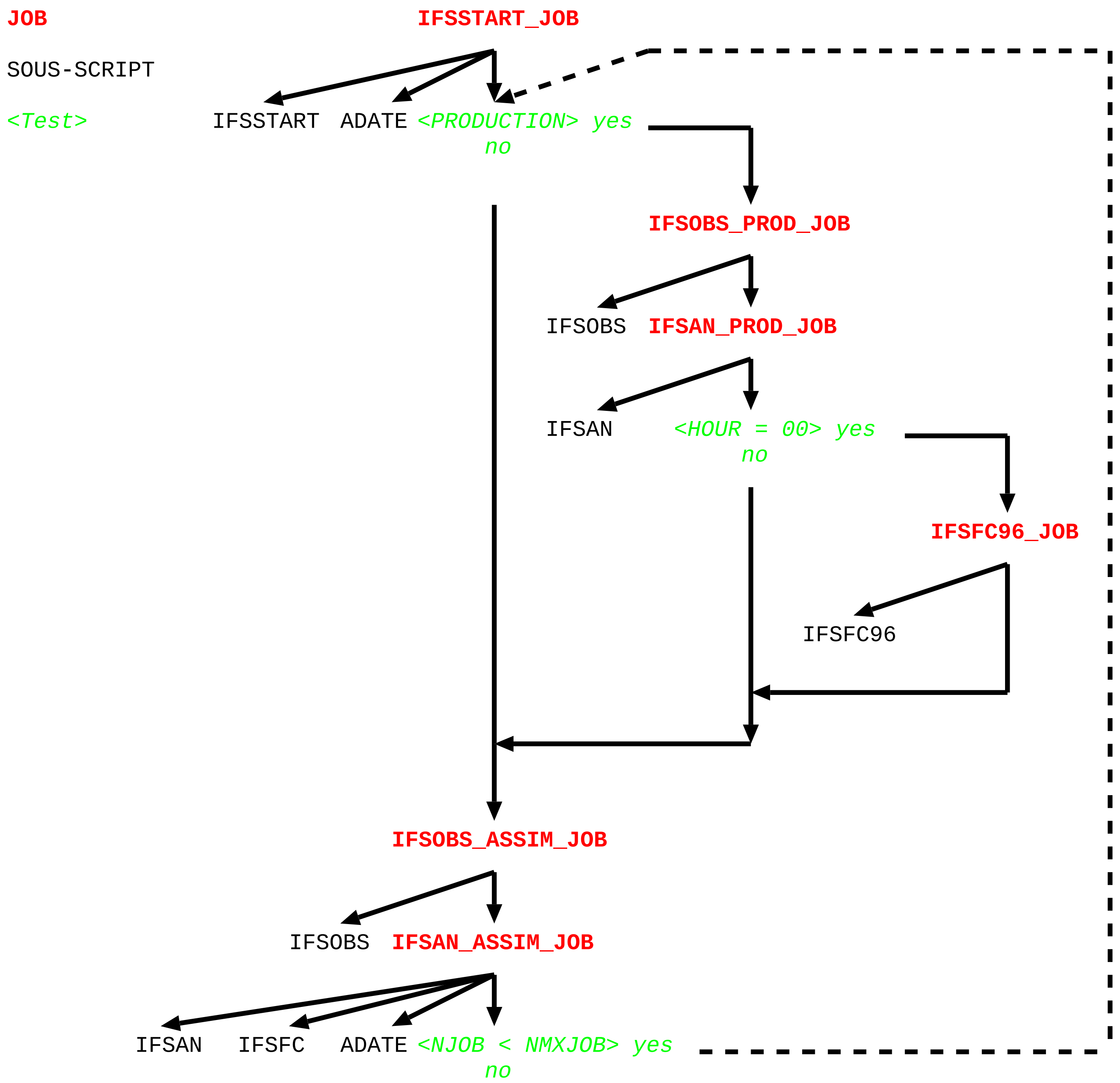
A noter également que ce terme est non quadratique par rapport à $\delta \boldsymbol{x}_k$, contrairement au terme $J^c - DFI$ utilisé dans la formulation 4D-Var, ce qui ne permet pas d'utiliser un algorithme de minimisation de type Gradient-Conjugué (`n1cg1`) pour minimiser la fonction coût globale : on utilisera dans ce cas le minimiseur de type Quasi-Newton (`m1qn3`).

Il faut également signaler que le biais se trouve dans `ICMSHscre+0000` en sortie du screening et que c'est l'analyse initialisée valide à $t_n/2$ que l'on trouve dans `ICMSHtraj+0000` à l'issue de la dernière mise-à-jour de la trajectoire (le `+0000` pouvant porter à confusion dans ce cas).

Chapitre 4

Structure du script

4.1 Arbre des appels



Commentaires

- IFSSTART_JOB : lance un cycle d'assimilation
  - IFSSTART : prévision 6 h,
  - ADATE : incrémente la date de 6 h,
  - lance IFSOBS_PROD_JOB ou IFSOBS_ASSIM_JOB.

- IFSOBS_PROD_JOB : récupération des observations cut-off court (production)
  - IFSOBS : récupération des observations,
  - lance IFSAN_PROD_JOB.

- IFSAN_PROD_JOB : analyse cut-off court (production)
  - IFSAN : analyse,
  - lance IFSFC96_PROD_JOB pour un réseau de OO H,
  - lance IFSAN_ASSIM_JOB.

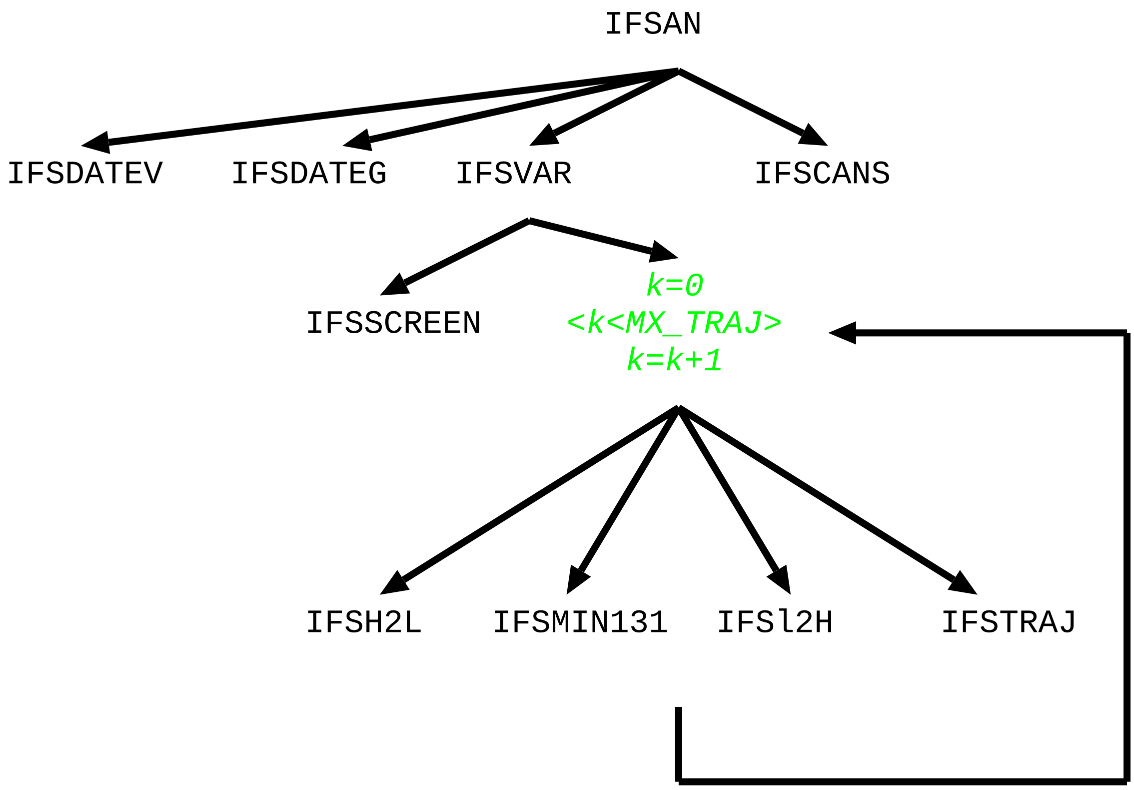
- IFSFC96_JOB : prévision 96 h
  - IFSFC96 : prévision.

- IFSOBS_ASSIM_JOB : récupération des observations cut-off long (assimilation)
  - IFSOBS : récupération des observations,
  - lance IFSAN_ASSIM_JOB.

- IFSAN_ASSIM_JOB : analyse cut-off long (assimilation)
  - IFSAN : analyse,
  - IFSFC : prévision 6h pour constituer l'ébauche suivante,
  - ADATE : incrémente la date de 6 h,
  - lance IFSAN_PROD_JOB ou IFSAN_ASSIM_JOB.

Le choix de la relance de IFSOBS_PROD_JOB de IFSOBS_ASSIM_JOB, effectué dans IFSSTART_JOB et IFSAN_ASSIM_JOB, est en fait un peu plus compliqué que sur le graphique : on teste aussi la variable P00ONLY et lorsque P00ONLY=yes s'il s'agit du réseau de OOH.

Le sous-script IFSAN est lui-même éclaté de la manière suivante.



- IFSDATEV : calcul de la date de début de la période d'assimilation,
- IFSDATEG : calcul de la date de l'ébauche,

- IFSSCREEN : calcul des écarts ébauche/observations et contrôle de qualité (CONF 2),

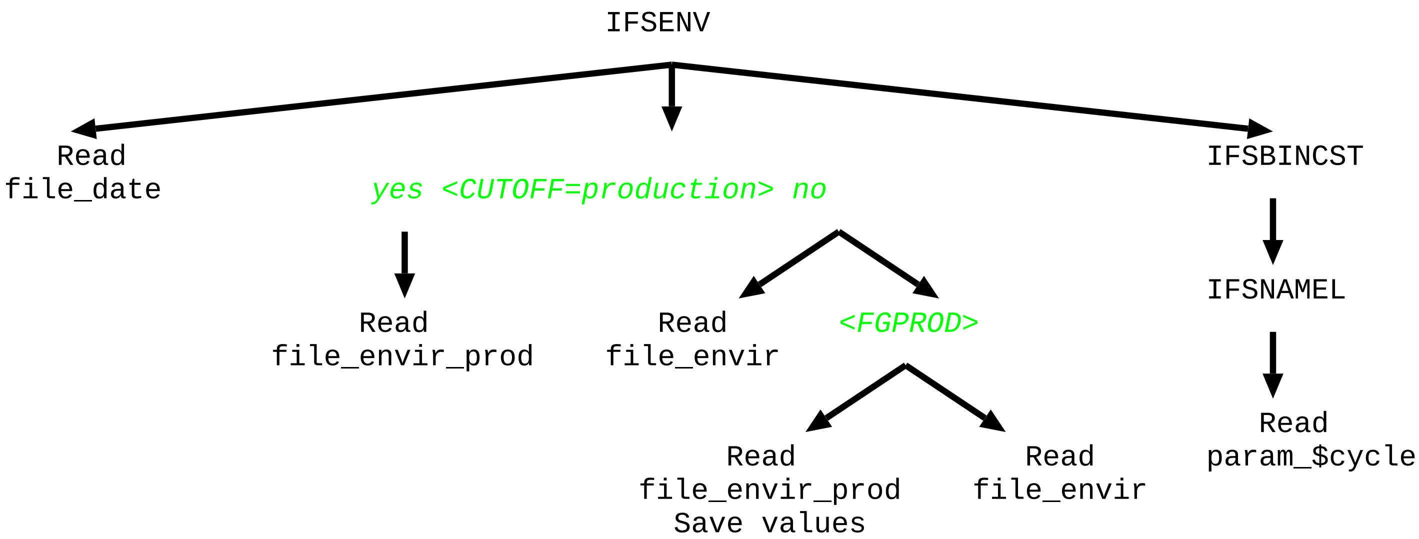
- IFSH2L : passage de l'ébauche à basse résolution (CONF 1 / FULL-POS),
- IFSMIN131 : minimisation (CONF 131),
- IFSL2H : remontée de l'ébauche et de l'analyse à haute résolution (CONF 1 / FULL-POS),
- IFSTRAJ : mise-à-jour de la trajectoire et comparaison aux observations (CONF 2).

- IFSCANS : analyse de surface (CONF 701).

Les jobs

IFSSTART_JOB, IFSOBS_PROD_JOB, IFSAN_PROD_JOB, IFSOBS_ASSIM_JOB, IFSAN_ASSIM_JOB et IFSFC96_JOB

appellent également le sous-script d'initialisation IFSENV. La structure de IFSENV est elle-même la suivante.



Le sous-script IFSBINCST permet de récupérer l'ensemble des exécutables et fichiers de constantes correspondant au cycle choisi. Il appelle IFSNAMEL qui lit un fichier de paramètres correspondant au cycle Arpège choisi. Ce fichier contient par exemple la résolution du modèle ou le nombre de boucles externes de la minimisation. IFSNAMEL récupère aussi les namelists associées au cycle choisi et les modifie éventuellement en fonction des paramètres passés dans le fichier `file_envir` (la subsitution des valeurs de ces namelists n'est initialisée que si le paramètre `RESOLTRAJ` est explicitement donné dans `file_envir`).

Le job IFSFC96_JOB lit aussi le fichier `file_date_prod` créé par le IFSAN_PROD_JOB.

Les différents scripts utilisent également les fonctions ou exécutables `timdat`, `copy`, `adate`, `aget`, `aput` et `asub` permettant de simplifier le script 4D-Var.

`timdat` : incrémentation de la date.
`copy` : simule le comportement de `gget`.
`adate` : incrémentation de la date de 6h.
`aget` : récupération d'un fichier archivé sur delage ou sur la WORKDIR sous l'arborescence de l'expérience.
`aput` : archivage d'un fichier sur delage et sur la WORKDIR sous l'arborescence de l'expérience.
`asub` : soumission d'un job depuis le script.

Ces fonctions sont placées sur la machine locale dans le répertoire `/script/tools` (voir paragraphe 1.2) et sont chargées par IFSENV.

Enfin, IFSBINCST utilise les fonctions `genv` (récupération de l'environnement Arpège) et `gget` (récupération des fichiers Arpège correspondants) développées par GCO et IFSNAMEL utilise `gnam` (modification de namelists) également développé par GCO.

4.2 Algorithmique de la minimisation

```
cas de la minimisation
NCONF=131
LJACDAV=F

Control variational
assimilation job
level 1             --> CVA1 --> SU1YOM        setups
                             --> <LSIMOB = F>  Real observations
                                               Set up observations
                                 <LOBS> SUOBS
                             --> SUCOS         Initialize cost-fonction
                             --> CNT2          Trajectory for linearization
Control variational
assimilation job
level 2                      --> CVA2          Compute the cost
                                               fonction and its gradient
    --> SIM4D(4,NVADIM,VAZX,ZF,VAZG,NSIM4D=0)
          (control variable, cost function, gradient)
          --> initialization:MONVAR
                             SU2YOM
                             SUCOS
                             SUALLT
                             SUALLDFI
          --> from control variable
              to model variables
              CHAVARIN
          --> direct integration :
              CNT3
          --> adjoint integration :
              CNT3AD
          --> Change of variables :
              CHAVARINAD
          --> Sum up cost fonction :
              EVCOST
              SCAAS
          --> Gradient
              ECAIN or CAIN
              STEPO
          --> Write analysis
              ECAIN or CAIN
              STEPO
              SPNORM
              SCAAS
Print initial gradient norm   --> SCAAS(NVADIM,VAZG,VAZG,ZVAZGNO)
                              --> <LTEST> GRTEST
Minimization                  --> <LMULTINC=T> SUHMINC
                                  (ini the hessian estimate for
                                   warm N1QN3 restart)
                                  <LN1CG1=T> N1CG1
                                  SAVHESS
Test de qualité
de la minimisation            --> SIM4D(4,NVADIM,VAZX,ZF,VAZG,
                                  NSIM4D=999)
                              --> SCAAS
Save results of variational job
                              --> SAVEMINI(VAZX)
                              --> WRITEOBA
```

4.3 Namelists

```
namelistcans    : analyse de surface
namelistfc06    : prévision à 6h
namelistfc96    : prévision à 96h
namelisth2l107  : passage de l'ébauche à la troncature T107
namelisth2l161  : passage de l'ébauche à la troncature T161
namelistl2h     : passage de la basse résolution à la haute résolution
namelistmin1311 : première minimalisation
namelistmin1312 : deuxième minimalisation
namelistobsort  : distribution des observations d'un cma avant le passage à odb
namelistscreen  : calcul de la trajectoire issue de l'ébauche; contrôle de qualité
namelisttraj1   : première mise-à-jour de la trajectoire
namelisttraj2   : deuxième mise-à-jour de la trajectoire
```

4.4 Les principales variables du script

4.4.1 Dates et heures

Nom	Description
DATRAW	date de l'assimilation à traiter (YYYYMMDDRR)
DATEV	date de début de la période d'assimilation
HOURV	heure de début de la période d'assimilation
DATEG	date de l'ébauche
HOURG	heure de l'ébauche

4.4.2 Les répertoires sur la machine d'archivage (delage)

R désigne les répertoires sur delage.

Nom	Valeur	Description
XRDIR	\$RDIR/\$XID/\$CUTOFF	répertoire d'archivage pour l'expérience
XRDATE	\$XRDIR/\$YEAR/\$MONTH/\$DAY/\$RR	répertoire d'archivage pour l'assimilation élémentaire
XROUTPUT	\$XRDATE/output	répertoire contenant les différents listings

4.4.3 Les répertoires sur kami

\$RDIR est le répertoire racine. \$XID est l'identificateur de l'expérience. Ils sont déterminés par l'utilisateur dans configure.

W concerne les répertoires sur kami.

Nom	Valeur	Description
XWDIR	\$WORKDIR/\$XRDIR/\$DATRAW	répertoire pour l'expérience
XWDATE	\$XWDIR/current_work	répertoire pour l'assimilation élémentaire

4.4.4 Les variables sur la machine locale

Nom	Description
LOCAL	adresse de la station de travail
LDIR	racine du répertoire local pour la station de travail

4.5 Résultats d'une expérience

Le job produit un ensemble d'analyses et de prévisions. De plus, un grand nombre de fichiers intermédiaires est créé. La structure du répertoire de l'expérience sur la machine d'archivage et `kami` est décrite ici.

4.5.1 Archivage des données

Au cours d'une assimilation élémentaire

La structure de l'arborescence sur la machine d'archivage est la suivante ($RDIR/$XID/$CUTOFF/$YEAR/$MONTH/$DAY/r$RR est le principal répertoire d'archivage).

Nom		*Description*	*Sortie de*
$VATRH\$k$	=>	$Precond.\ issu\ de\ la\ mini.\ \$k$	$(ifsmin131)$
$analyse$	=>	x^a	$(ifscans)$
$analyse\_b\$k$	=>	w_k^a	$(ifsmin131)$
$biais\_dfi$	=>	$biais\ DFI\ ebauche$	$(ifsscreen)$
$guess + 0003$	=>	$x^b(t_0)$	$(ifsfc,\ analyse\ precedente)$
$guess + 0006$	=>	$x^b(t_0 + 3h)$	$(ifsfc,\ analyse\ precedente)$
$guess\_b\$k$	=>	$w_k^b + \delta w_{k-1}$	$(ifsmin131)$
$guess\_b\$RESOLINC(\$k)$	=>	$S_k(x^b)$	$(ifsh2l)$
$icmsharpe + 0003$	=>	$x^f(t_a + 3h)$	$(ifsfc)$
$icmsharpe + 0006$	=>	$x^f(t_a + 6h)$	$(ifsfc)$
$icmshtrajanin\$k$	=>	$S_k^{-I}(w_k^a)$	$(ifsl2h)$
$icmshtrajfgin\$k$	=>	$S_k^{-I}(w_k^b)$	$(ifsl2h)$
$lissef$	=>	$Fichier\ de\ lissage\ analyse\ de\ surface$	$(ifscans)$
$odb\_canari.tar$	=>	y_S^0 $(Observations\ pour\ analyse\ de\ surface)$	$(ifsscreen,\ ifscans)$
$odb\_ccma.tar$	=>	y_C^0 $(Observations\ compressees)$	$(ifsscreen,\ ifsmin131,\ ifstraj)$
$odb\_ecma.tar$	=>	y^0 $(Observations\ etendues)$	$(ifstraj)$
$traj0 + 0000$	=>	$M(x^b, t0)\ (= x^b)$	$(ifstraj)$
$traj0 + 0003$	=>	$M(x^b, t0 + 3h)$	$(ifstraj)$
$traj0 + 0006$	=>	$M(x^b, t0 + 6h)$	$(ifstraj)$
$traj\$k + 0000$	=>	$M(x_k, t0)\ (avec\ x_k = x_{k-1} + S_k^{-I}(w_k^a))$	$(ifstraj)$
$traj\$k + 0003$	=>	$M(x_k, t0 + 3h)$	$(ifstraj)$
$traj\$k + 0006$	=>	$M(x_k, t0 + 6h)$	$(ifstraj)$
$traj\$k + 0000000$	=>	χ_k	$(ifsmin131)$

Ces fichiers sont également présents sur kami sous $WORKDIR/$RDIR/$XID/$CUTOFF/$DATRAW. On trouve également sous $WORKDIR/$RDIR/$XID/$CUTOFF/ les exécutables Arpège ainsi que les différents fichiers de constantes.

Après la fin d'une assimilation élémentaire

Après la fin d'une assimilation élémentaire, la plupart des fichiers intermédiaires est supprimée sur `delage`. On ne trouve plus alors que les fichiers suivants :

$$\begin{array}{rcl}
analyse & => & x^a \\
guess + 0006 & => & x^b(t_0 + 3h) \\
icmsharpe + 0003 & => & x^f(t_a + 3h) \\
icmsharpe + 0006 & => & x^f(t_a + 6h) \\
lissef & => & Fichier\ de\ lissage\ analyse\ de\ surface \\
odb_canari.tar & => & y_S^0\ (Observations\ pour\ analyse\ de\ surface) \\
odb_ccma.tar & => & y_C^0\ (Observations\ compressees) \\
odb_ecma.tar & => & y^0\ (Observations\ etendues)
\end{array}$$

De même le répertoire $WORKDIR/$RDIR/$XID/$CUTOFF/$DATRAW est supprimé et le répertoire $WORKDIR/$RDIR/$XID/$CUTOFF est supprimé à la fin du cycle d'assimilation complet.

4.5.2 Les listings

Les listings des différentes configurations d'Arpège utilisés par le script du 4D-Var se trouvent sur la machine d'archivage sous le répertoire $RDIR/$XID/assim/$YYYY/$MM/DD/rRR/output au cours d'une assimilation élémentaire. A la fin de cette assimilation le répertoire output est "taré" et placé sous \$RDIR/\$XID/assim/\$YYYY/\$MM/\$DD/r\$RR.

Un des listings particulièrement intéressant est celui issu du calcul de la minimisation à basse résolution (listing.min131\$k), car il permet de représenter la fonction coût, ainsi que le gradient en fonction du nombre d'itérations.