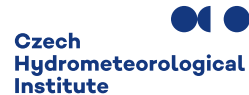


*Regional Cooperation for
Limited Area Modeling in Central Europe*



B matrix computation

A. Bučánek, 23.9.2021



ARSO METEO
Slovenia

- ▶ B matrix properties
- ▶ Computation methods used in ALADIN
 - ▶ NMC
 - ▶ Ensemble based B
 - ▶ Downscaled EDA
 - ▶ EDA (Ensemble Data Assimilation)
- ▶ Technical steps explained
 - ▶ Choosing period
 - ▶ Creation LBC from global EDA
 - ▶ Obs perturbations
 - ▶ Computation of differences
 - ▶ Computation of B (festat)
 - ▶ Tuning of B

B matrix properties in ALADIN (1)

- ▶ The role of B matrix is to scale, spatially filter and propagate the observed information away from observation point.
- ▶ Due to balance properties the corrections are propagated from one variable to the other
- ▶ The bg error statistics depend on:
 - ▶ model and its resolution (global, lam)
 - ▶ geographical area (midlatitude, tropical)
 - ▶ weather regime, time of day, season
 - ▶ density of the observation network. (dense x sparse obs)

- ▶ In ALADIN we use incremental formulation of 3D-VAR with transform to control variable χ (Courtier et al., 1998)

$$J(\chi) = \frac{1}{2}\chi^T\chi + \frac{1}{2}(\mathbf{H}\mathbf{B}^{\frac{1}{2}}\chi - \mathbf{d})^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{B}^{\frac{1}{2}}\chi - \mathbf{d}),$$
$$\nabla J(\chi) = (1 + \mathbf{B}^{\frac{T}{2}}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{\frac{1}{2}})\chi - \mathbf{B}^{\frac{T}{2}}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{d}.$$
$$\delta\mathbf{x}_a = \mathbf{B}^{\frac{1}{2}}\chi_a.$$

- ▶ By series of operators we remove the correlation of errors from state vector and we get new variable χ (control variable) which has identity covariance matrix

$$\mathbf{B}^{1/2} = \mathbf{L}\mathbf{D}\mathbf{V}\mathbf{W}.$$

- ▶ No need for inversion of B matrix.

B matrix properties in ALADIN(3)

- ▶ Contains cross covariances between variables - balances (vorticity, divergence, surf pressure and spec. humidity)
- ▶ Contains horizontal autocovariance and vertical autocovariance (see [3D-var report, cnrm, Fischer 2007](#))
- ▶ As a simplification we expect that horizontal autocovariances to be homogeneous and isotropic
 - ▶ In spectral representation this means that waves with different wave numbers are decorrelated.
 - ▶ In other words, each three-dimensional auto-covariance matrix is block-diagonal with one block of vertical correlations for each wave number

- ▶ B matrix properties
- ▶ **Computation methods used in ALADIN**
 - ▶ **NMC**
 - ▶ **Ensemble based B**
 - ▶ **Downscaled EDA**
 - ▶ **EDA (Ensemble Data Assimilation)**
- ▶ Technical steps explained
 - ▶ Choosing period
 - ▶ Creation LBC from global EDA
 - ▶ Obs perturbations
 - ▶ Computation of differences
 - ▶ Computation of B (festat)
 - ▶ Tuning of B

- ▶ B matrix is computed in two step:
 1. To simulate background (first guess) errors we use differences of model or ensemble forecasts
 2. To compute the B matrix from differences is used program festat.

- ▶ + cheap on computation cost (all you need you have in house)
- ▶ + no need to prepare special LBC
- ▶ - statistics are mixture of analysis and background error
- ▶ - in theory it should overestimate the errors (Berre 2006)
- ▶ Steps:
 1. Compute differences (fc+6h) – (fc+30h) valid at the same time
 2. Compute B matrix (festat)

- ▶ This is a must before computation of full EDA B matrix
- ▶ - need to get LBC from global EDA
- ▶ - need to run forecasts
- ▶ - bg statistics are influenced by global model analysis error
- ▶ Steps:
 1. Create LBC from global EDA (on provider side)
 2. Run oper. dyn. adapt. of each downloaded member (e.g. +6h)
 3. Compute differences between members
 4. Compute B matrix (festat)

- ▶ + diagnoses the statistics of actual analysis system
- ▶ - the most expensive,
- ▶ - assumes statistics of observation error (and SST, etc.) are well known.
- ▶ - danger of feedback. (Noisy analysis system => noisy stats => noisier system.)

▶ Steps:

1. Create LBC from global EDA
2. Create LAM EDA
 - ▶ Need of working, well tuned 3d-var system before start with perturbations of observations (pert SST, pert FG).
 - ▶ The upper air observations can be perturbed offline (pertcma) or in screening
 - ▶ OI observation can be perturbed only offline (pertcma)
 - ▶ Sea surface temperature could be perturbed ([Khalfaoui, 2019](#))
 - ▶ First guess perturbations could be inflated ([Raynaud 2012](#))
3. Compute differences between members
4. Compute B matrix (festat)
5. Tune the B matrix by REDNMC (theoretical value for EDA is REDNMC=0.7)

- ▶ Initial perturbation of EDA is not important since it is forgotten during cycling ([El Ouaraini, R., and Berre, L., 2011](#)) but some warmup period would be better to remove from computation of B.
 - ▶ So we can start the LAM EDA from the same file (e.g. deterministic guess)
- ▶ LAM EDA forecasts should be long enough to get rid of spinup caused by large jumps in resolutions (AEARP ~40km, EDA IFS ~30km)
 - ▶ So be sure that you do not have spinup when short cycling frequency (1, 2, 3h) is used in sampling of B

- ▶ B matrix properties
- ▶ Computation methods used in ALADIN
 - ▶ NMC
 - ▶ Ensemble based B
 - ▶ Downscaled EDA
 - ▶ EDA (Ensemble Data Assimilation)
- ▶ **Technical steps explained**
 - ▶ **Choosing period**
 - ▶ **Creation LBC from global EDA**
 - ▶ **Obs perturbations**
 - ▶ **Computation of differences**
 - ▶ **Computation of B (festat)**
 - ▶ **Tuning of B**

- ▶ The minimal length of the period(s) should be that you get at least the same number of differences as you have vertical levels
 - ▶ This produce the positive definite B matrix
- ▶ We know that errors depend on period of year and change even during a day
- ▶ So we would like to have some averaged B for the different weather regimes which can appear.

- ▶ AEARP – conf. e927
 - ▶ Unmaintained example of e927
belenos: ~bucanek/exp/lancelot/aearp_beaufix_479
- ▶ IFS EDA – conf. e903
 - ▶ IFS uses completely different surface scheme so one should be careful when using IFS in dynamical adaptation.
This could increase the spinup.
 - ▶ Yelis Cengiz will explain how to create LBCs from IFS EDA

OBS perturbations (1)

- ▶ Pertcma.F90 (odb/tools/Pertcma.F90)
 - ▶ Missing config in gmckpack (user has to make his own linking)
 - ▶ Input ECMA or CCMA databases
 - ▶ Output perturbed DB with random draw from Gaussian distribution $N(0, \text{sigma_o})$
 - ▶ `sigma_o` is read from ODB itself (`final_obs_error`)

```
seed=$1 # seed for perturbation, i.e. member number (1, 2, 3, ....)
DB=$2   # DB = NAME of db to be perturbed (ECMA, CCMA)
```

```
export SWAPP_ODB_IOASSIGN=ioassign
export ODB_SRCPATH_ECMA=/Path/to/your/odb/${DB}
export ODB_DATAPATH_ECMA=/Path/to/your/odb/${DB}
export IOASSIGN=/Path/to/your/odb/${DB}/IOASSIGN
```

```
./pertcma $seed $DB
```


- ▶ Perturbations inside screening (e002)
 - ▶ This is done before the screening
 - ▶ It might happen that the perturbed obs will be rejected by screening
- ▶ Changes in namelist NAMSCC
 - ▶ LPERTURB=.T.
 - ▶ NAEMEMBER=MEMBno (e.g. 1,2,3...)
 - ▶ NAENSEMBLE=1 (Ensemble group number, any integer)

- ▶ Run configuration e001 (standard forecast) with modified namelist:
 - ▶ LFEMARS=.T.
 - ▶ LSPRT=.F. # to use normal temperature instead of virtual
 - ▶ Suppress inline fullpos (NFPOS=0)
 - ▶ Suppress computation of fluxes (NAMXFU – LXFU=.F.)
 - ▶ Run it on **1 CPU only!**
 - ▶ Two more inputs
 - ▶ ICMSH\${CNMEXP}FGIN = ICMSH\${CNMEXP}INIT = MEMB1
 - ▶ ICMSH\${CNMEXP}ANIN = MEMB2
 - ▶ Result is MEMB2-MEMB1 written to gribfile “gribdiff”
 - ▶ Strategy for differences is usually to subtract every second member m2-m1, m4-m3, m6-m5, ... but there are other possibilities ...

- ▶ Program festat is in ALADIN export (utilities/bcov_lam), compile gmckpack
- ▶ Inputs are gribdifs called "ensdiff\${ncase}",
 - ▶ where ncase is 3digit order number of gribdiff produced in previous step. It is necessary to prepare special namelist.
- ▶ On cycle 43 festat allows parallel execution
- ▶ It is memory consuming so don't be surprised by crashes when not enough memory is supplied
- ▶ Program fediacov – computes B matrix diagnostics
- ▶ Package to compute B (scripts, namelists, visualization):
<https://www.rlace.eu/forum/viewtopic.php?f=30&t=62&p=2096#p2143>

- ▶ REDNMC – scaling of “B”
 - ▶ it is applied as REDNMC²
 - ▶ theoretical value for EDA B matrix is 0.7
- ▶ SIGMAO_COEF – scaling of “R”
 - ▶ It is vector of obs error scalings per obstype
 - ▶ Should be set in Bator, screening (e002), minim (e131)
 - ▶ It is not applied to all obs parameters!
- ▶ TuneBR PACKAGE using Desroziers method:
 - ▶ <https://www.rclace.eu/forum/viewtopic.php?f=30&t=248&p=1092&hilit=Tunebr#p1092>
- ▶ Use impact studies for tuning

Questions (1)

- ▶ i) how to optimise the efforts of those countries who never computed a B-matrix so that they may get a B-matrix very efficiently (and this requires a more practical information), preferably from home and with local resources
- ▶ Countries using the same telecom domain can agree on period for computation of EDA LBC to minimize the effort
- ▶ I am not sure if it is feasible to prepare common scripts for EDA
- ▶ To compute the NMC or downscaled EDA B matrix use LACE package.

Questions (2)

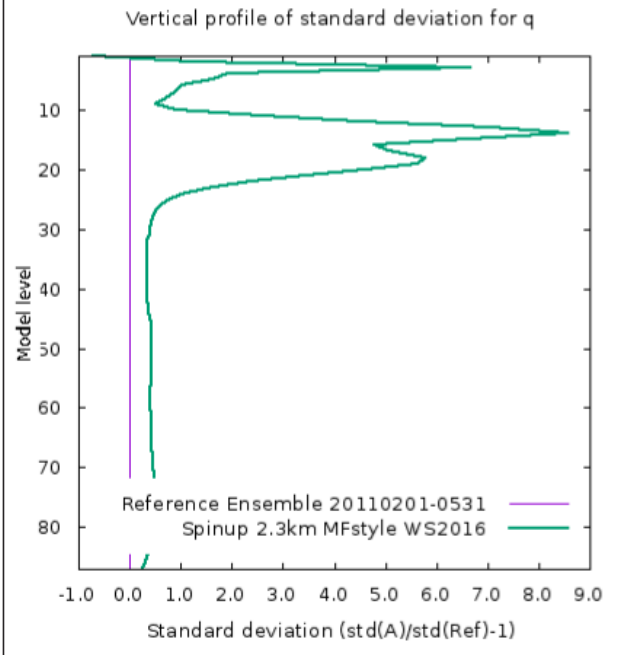
- ▶ ii) how suitable can be this B-matrix to: (a) be implemented in to operations; (b) to improve our model scores, that is to make our DA schemes/approaches optimal (and I would say here people should have a more scientific view of B-matrix role and numerical limits).
- ▶ Well selected period for B calculation and tuned REDNMC should be enough for operational B matrix. But not only B influences the DA performance (e.g. blacklisting, thinning, selection of obs parameters which will be assimilated, setup of VARBC, ...)

Questions (3)

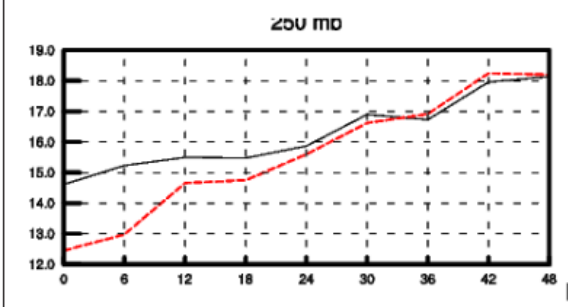
- ▶ Which are the different aspects to take into account, when computing B-matrix by an ensemble technique, if the model is coupled by either IFS/ECMWF or ARPEGE ?
 - ▶ The main difference is that IFS is using different surface scheme. So when you solve the surface issue the rest is the same in computation of B.
- ▶ What is the impact of different B-matrices over different types of observations.
 - ▶ In theory the B matrix contains covariances of first guess errors which should be independent of observations. The B is not applied to obs type but to vor, div, T, ps, q. Covariances vary with height and total wave number so the impact can change for observation which are measured on different heights...
 - ▶ I am not able to comment on different types of B (nmc, ens..) but I can show one particular case where we computed downscaled EDA B for 2.3km resolution and first test showed degradation at 250hPa which is height were one of wv seviri channels is peaking. Investigation showed that we have more than 8 time larger error in new B than in old one

Problem with new B matrix on high resolution

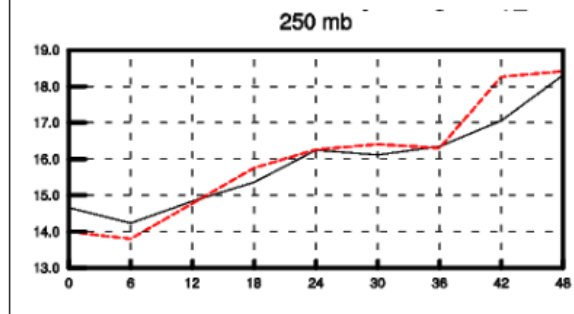
Relative vertical profile of q stde (B matrix diagnostics)



RMSE RH250hPa before tuning of B matrix, ALADIN 4.7km (red), ALADIN 2.3km (black)



RMSE RH250hPa after tuning of B matrix



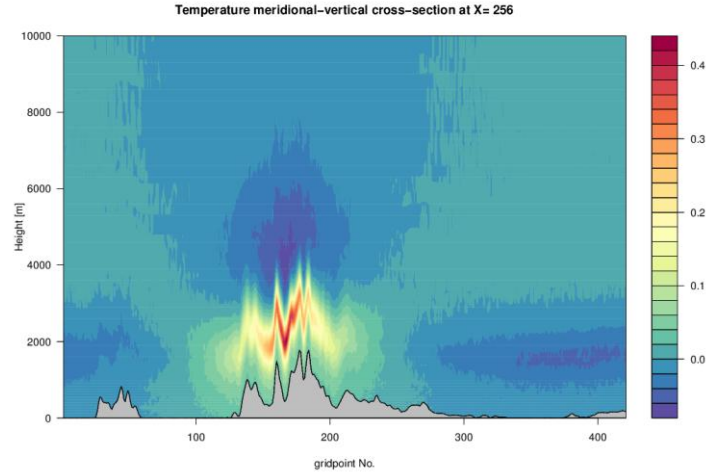
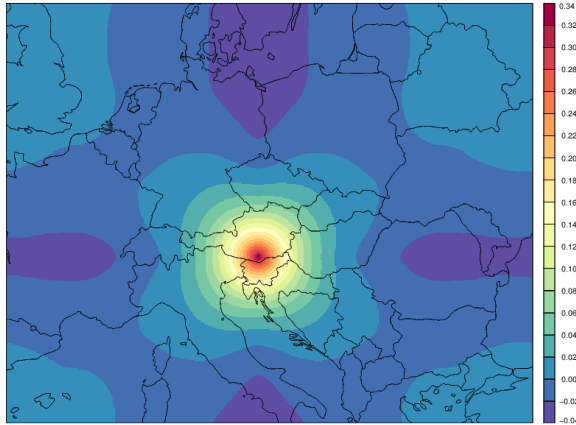
Questions (4)

- ▶ How much is it worth to invest on B-matrix before implementing a local operational DA system
 - ▶ The politically correct answer would be as much as you can effort :)
 - ▶ But in reality since you do not have any other working assimilation system beside dynamical adaptation it would be great to start with something, which is not making your system worse than dynamical adaptation (personally I would be happy if I got neutral results compared with Dyn.Adp.)
 - ▶ As an investment you can also think about computation cost, if you are limited with resources (without possibility to compute elsewhere) and you are keen to start first trials with 3DVAR I would recommend to use NMC method, since you have everything in house.
 - ▶ If you are thinking about EDA or downscaled EDA then you should be prepared to
 - ▶ Create telecom files from the global EDA members for your targeted period (this must be done on provider side and can take some time)
 - ▶ Download those members (could take about 5 days on slow RMDCN lines)
 - ▶ Be ready to run at least 200 forecasts. Do I have enough resources?

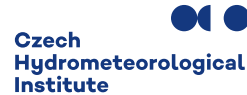
*Regional Cooperation for
Limited Area Modeling in Central Europe*



Analysis increment
S060TEMPERATURE
2019/01/06 z18:00 +6h



Thank you for your attention.



ARSO METEO
Slovenia

Possible strategies for differences

1. Subtract every second member
m2-m1, m4-m3, m6-m5
2. Subtract each not repeating, eg 4memb ens
m2-m1, m3-m1, m4-m1, m3-m2, m4-m2, m4-m3
3. Subtract the mean member,
m1-m, m2-m, m3-m, m4-m