Draft proposal for the Aladin mid-term scientific program in the field of data assimilation
C. Fischer, 2005-05-26

1. preamble

Let us start with an excerpt from the 2004 Aladin DA workplan:
"*In the field of data assimilation, there has been an early recognition that the tools and methodologies chosen for the 100-20 km resolution (global IFS and Arpège) can generally talking be downscaled to 10 km (Aladin-1) and below (Arome). Indeed, all these projects will focus on variational, incremental techniques, and they might (or not, but depending on so far unpredictable scientific performances) give raise to new assimilation methods around improved 4D-VAR (for instance, with microphysics), ensemble-based techniques (EKF, ETKF,...), exotic observations (reflectivities, precipitations, GPS,...). The reasons for the wide range of common scientific and technical issues probably are various, but at least two of them are worth to be mentioned. Firstly, data assimilation based on optimal control as well as ensemble techniques heavily rely on the specification and/or estimation of error statistics (forecast model, observations, physical constraints). The specification of these error statistics is limited by a number of shortcomings, that make their practical estimation generally rather crude, as compared to the desired precision of the deterministic forecast itself (one usually is not happy with a displaced or absent cyclone in the "actual" forecast). Some of the shortcomings are: lack of dense and good-quality measurements to compare with, too big phase space (10^7 for observations, 10^6 for model), sampling errors, common use of Gaussian error models and linear error propagation. Misspecification of error statistics is a general plague, at whatever scale the associated assimilation and NWP system is supposed to be operated. Secondly, data assimilation requires a significant manpower effort on upstream developments, such as transmissions, database, observation monitoring and quality control, along with a wide spread of scientific skills (NWP in its forecast component, linear algebra, knowledge of observation properties, management of complex operational suites). Thus, the definition of several, very different, development projects in this field, inside a very same NWP Center, probably is outside of range (the building up of the Aladin 3D-VAR using the IFS/Arpège software as backbone is a striking example of an economic "re-routing").*
The main move when comparing the present and future Alaro/Arome interaction with the Aladin-1 scientific plan concerns the departure from the time-continuous TL/AD models (and associated "quick" step towards 4D-VAR in the LAM), towards a more observation-oriented and space-continuous approach (3D-FGAT at full LAM forecast resolution)."
Bouncing back from this stratospheric statement, and focussing into detailed work, a few specific constraints however do appear, which make future developments potentially scale-dependent:

1. cloud and rain affected observations should be treated consistently with the physical parametrizations relevant for the scale under consideration: Méso-NH microphysics at 2km, Alaro (or Hirlam?) physics at 10 km, any other in the grey zone (5km). This issue holds in principle for cloud or precipitation-affected radiances (1D-var retrievals) and radar reflectivities.

2. If more model physical information has to enter the B matrix formulation, either as microphysical information for the background state or physics tendencies, then these fields also will depend on the physics parametrization.
3. Additional hypothetical background initialisation constraints ("Jc") may be scale dependent as they might depend on the spurious inertia-gravity waves present in the trajectory.

It is the author's conviction that this mid-term period will not unveil all of the above questions. However, the 2005-8 period will be the first one where the Aladin DA scientists will need to somehow, if not adjust, at least position their own desires and (operational) needs with respect to the overall scientific and technological environment: the speeding-up of Arome (operations planned … in 2008), the convergence with the Hirlam project.

Finally, to close this preamble, it is wise to recall the major lessons from the last, three-year workplan (2002-2004):

1. 4D-var issues were mostly left untouched. The quite pioneering work in the frame of Alatnet rather has raised extra questions. More generally, the Aladin variational community has clearly done an implicit by-passing on 4D-var in the last years. Even at the 10km scale, an operationally valuable 4D-var looks out of range in the frame of Aladin. At the present state, the question whether the closer collaboration with Hirlam can improve the situation is left open.
2. in the previous plan, the work around surface analysis also was quite over-optimistic. Furthermore, we recall the progressive move from CANARI to varpack type of applications (and the desire to use as much as possible the 3D-var algorithm for nowcasting).
3. major areas where either scientific internationally recognized work or significant take-off has been noticed are: B-matrix issues, cycling strategies (blending of any type) and observations. The present mid-term plan certainly should reflect this, and make sure that momentum in these three fields of activity is maintained.

2. general priorities for the 2005-2008 period

The main goal is to achieve a stable, cost-reasonable, application, using as many as possible local observations and observation types. For short:

- 3D-Var FGAT
- digital filters (incremental ?)
- surface observations, radiosonde data, aircraft, ATOVS data, geostationary satellite data, GPS and radar
- an efficient dynamical core for the TL/AD HPE models
- further experience with ensemble sampling and system simulation
- reliable and efficient wavelet transforms, associated research work on non-homogeneous, anisotropic structure functions

3. 3D-var and assimilation cycles as the core part of our operating DA systems

B matrix:

Main topic	Sub-topic	Priority	Persons and manpower	Comments

Sampling techniques				
NMC standard or lagged (Bnmc)	maintenance	P1	LB, GB, KH, Morocco, BC for "Jk version"	No innovation planned, but existing tools will remain useful to test new ideas in the area of balance conditions or tuning or time/flow dependency
Ensemble of analyses (Bens)	Native Aladin ensembles using various combinations (surface obs, model parameter perturbations, …)	P0	LB, CF, 1 student from ENM ?	Presently, Bens is obtained from dynamical adaptation of an Arpège ensemble of analyses
	Re-assess the evaluation of the "V-metric" term representing the Arpège analysis error projected on the LAM domain: truncation, tuning	P1	BC	This term will probably remain monovariate
Hybrid B (Bhyb)	Mix climatological and situation-dependent statistics	P2	GB ?	
Structure functions				
Off-diagonal terms in the bi-Fourier representation	Development in festat, then evaluation inside 3D-var	P3		To help cut off the Extension zone and introduce a flavour of heterogeneous correlations
Wavelets	Chose adequate formulation for a LAM	P1	AD, LB	
	Code the transforms	P1	AD	Help for the distributed memory / message passing version ?
	Evaluate monovariate version in 3D-var	P1	AD, LB	
	Multivariate case and/or link with horizontal balances (β-plane, ω-equation etc…)	P2	AD, LB	
Dependency on background	Any ideas ?			

Dynamical balance				
Non-linear balance and ω-equation	Adapt to Aladin	P0	LB + ?	Inspired from IFS
Other balances taking into account surface friction ?		P2		
σb tuning and cycling (and mapping)				
Application of Lönnberg-Hollingsworth method				Ready at HMS, could be used for an independent validation of σb's
DFS method	Further evaluation in Aladin	P0	BC, CF, TM	Sb term, link with Jk, So term for Seviri
	Link with the ensemble estimation of B	P2	CF	
	Other a posteriori diagnostics (residual, V-term …)	P1	BC, CF	Cross-references with work in Arpège by G. Desroziers
Maps of σb's from the ensemble analysis method	Compute, store and read in 3D-var; test its impact	P2	LB, CF, 1 student from ENM ?	
	Assess daily variability and flow dependency	P2		
Formulation of control variable				
New humidity control variable		P0	LB + ?	Work of Elias Holm at ECMWF
New microphysical fields ? => could be simply total liquid water content in the control		P1	?	Only needed if one really expects that the associated error statistics can be efficiently modelled, and observations do give a valuable information on them
Coupling inside variational formulation				
Jk		P0	BC, CF	Coordinated with the end of V. Guidard's PhD work

Assimilation cycle (as a whole):

The existing variational assimilation system contains the following parts:

1. 6 hourly 3D-var cycled analyses (run quasi-operationally in Budapest, Casablanca, Toulouse)
2. NMC or ensemble statistics
3. ODB installations, but with difficulties to cope with the Arpège/IFS cycle updates
4. 3 blending techniques: DF-blending (run operationally in Prague), explicit blending (tested in Budapest), variational blending (so-called "Jk", evaluated in Toulouse)

In the coming mid-term period, the choices as concerns the cycling strategy will probably continue to depend on the local installation. The reasons here are both human (affinities) and technical (density and volume of observations, readiness of the coupling data). However, more work will be devoted to the following issues:

- 3 hourly analyses and cycling
- link of 3D-var and its performances with the blending technique

The DA community has also to provide manpower for the ***maintenance of a Reference Assimilation System***, which would most likely look like:

- 4D screening and 3D-Var FGAT (to process high frequency observations)
- LAMFLAG tool (small but sturdy !)
- All possible types of observations operationally used in *ANY* 3D-var operating Center (Arpège observations, SEVIRI, etc …)

Topic		Priority	Persons and manpower	Comments
Reference assimilation system		P0	All var. specialists & phasers	
Link 3D-var / blending technique		P0	Local specialists: BC+CF, GB, FH	
3 hourly analyses and cycling		P1	CF, BC	Can be linked with initialization; needs choices for coupling with Arpège

Initialization and "Jc":

Main topic	Sub-topic	Priority	Persons and manpower	Comments
Which imbalances do we actually have ?				
Understand what sort of imbalances do exist in the analysed state and in the model spin-up phase, with respect to:	Orography; coupling; physics; adiabatic (analysis increments)	P0	CF + ?	Should remain a domain for continuous watch and diagnoses, for the least
Incremental digital filters		P1	CF	Already used in the DF-blending suite (Prague)

Link with alternative dynamical balance formulations in B		P1	LB, CF + ?	
Jc-NNMI		P4		Unclear utility at very high resolution; never coded in Aladin
Jc-DFI		P3		Only for 4D-var

4. observations

Outlook on the assimilation of observations within the 2005-2008 period

Florence Rabier and Élisabeth Gérard, 10 March 2005
C. Fischer, 26 May 2005

1. Observations in Arpège useful for Aladin

a. List of sounders

- ATOVS – ongoing developments on sampling, blacklisting…
- SSM/I in clear sky conditions (humidity, surface win speed over the oceans)
- Advanced sounders – AIRS and IASI – assimilation of subsets of channels
- MSG Clear Sky Radiances (CSR) (to be extended later on to GOES data)

b. Wind data and GPS

- Assimilation of MSG winds
- Assimilation of MODIS winds
- Preparatory developments for ADM/AEOLUS mission
- Improvement of the assimilation of QuikSCAT data (resolution down to 25 km instead of 50 km)
- Assimilation of ASCAT data on METOP
- Assimilation of ground-based GPS: specific work for Arpège/Aladin using the ECMWF obs. operator (Fatima-Zohra El Guelai, P. Moll, P. Poli) and at GMME team for Aladin/Arome using a more accurate operator (H. Brenot, V. Ducrocq)
- Radio-occultation GPS data, limb sounders

c. Methodologies

- Improved emissivity over land surfaces to assimilate more microwave and infrared channels – ATOVS, AIRS, IASI and CrIS on NPP/NPOESS
- Assimilation of infrared cloudy radiances (work from PhD, M. Dahoui, to be continued)

- Assimilation of SSMI precipitating radiances
- Radiosonde data bias correction
- Advanced sounders data bias correction with neural network

2. Observations for Aladin

- Hungary: work on ATOVS data resolution (sampling), bias correction; impact of surface observations (T2m, RH2m) – R. Randriamampianina, H. Toth
- Morocco: assimilation of locally-received ATOVS data – Zahra Sahloui, Fatima Hdidou, Rashyd Zaaboul
- France: surface observations, link with Varpack (L. Auger)

3. Observations for Arome

- Radar: observation operator and evaluation at 2 km resolution with Méso-NH (O. Caumont, V. Ducrocq); pre-treatment, quality control, dataflow, 1D-Var and adjoint obs. Operator (E. Wattrelot, M. Jurasek)
- wind doppler: opening of a FlySafe position (sept. 2005)
- Meteosat SEVIRI: continuation of the work of Thibaut Montmerle for Aladin/France 3D-Var and specific work for AMMA (M. Nuret)
- Ground-based GPS: see GPS

5. 4D-var and related topics

The proposal is to try to promote a central work with a precise algorithmic goal, which includes two high-priority topics:

- a more efficient, up-to-date dynamical kernel: Semi-Lagrange TL and AD models for the hydrostatic primitive equation model
- a "4D-Var in a nutshell", that is a raw version containing one inner-loop minimisation using N>1 timeslots, the SL TL/AD propagation and (normally) all observation types valid for 3D-Var

more ambitious work is not discouraged (but with lower, variable priority):

- high frequency Doppler radar wind assimilation
- simple simplified physics
- sensitivity studies
- development of the TL/AD of the Non-hydrostatic model

however, all the latter items would be done at the researcher's own risk, in the sense that the Aladin project probably cannot guarantee full supervision and system logistics for all these technically demanding topics.

6. surface analysis

proposal by F. Bouyssel; please refer also to the common workplan for surface modelisation

The points proposed below do not explicitly mention the introduction of the externalised surface scheme, inherited from Méso-NH, into Arpège and Aladin (while it is supposed to be the default for Arome anyway). However, depending on pending decisions about the exact implementation of the externalised surface in these two models, some technical or scientific developments will probably be needed (physiographic database, adapt an incore surface assimilation module or develop a sound interface between the prognostic externalised surface and the "old" surface analysis scheme, …).

- Sea surface temperature: until now, SST analysis is performed every six hours in Arpège only, assimilating surface observations (buoys, ships) and using the NESDIS SST analysis as a relaxation. Geostationnary satellites provide high temporal and spatial SST observations over cloud free areas and are therefore particularly interesting to initialise SST in Aladin. The ocean and sea ice SAF provides such SST data at 0.1° resolution with quality flags over European seas and the Atlantic ocean, derived from Meteosat-8 and GOES-East. A strategy should be developed to use these data.
- Snow analysis: no snow analysis is currently available to initialise Arpège and Aladin. Snow water content is evolving with the model. There is in CANARI a relaxation towards a climatology and an additional melting term which is active when the analysed surface temperature becomes significantly larger than 0 °C. On the opposite, it is not possible to inject an up-to-date snow cover on a daily basis for instance. Thus, an error on snow water equivalent may persist during several days with a negative impact on the forecast of 2m temperature. The NESDIS snow cover analysis is planned to be used in Arpège, and the snow cover provided by the land SAF is of interest for Aladin.
- Soil moisture and temperature:
    - Use of CANARI surface analysis in Aladin: this tool is still used in several centers (diagpack in Toulouse, coupling with 3D-var in Casablanca and Budapest), however, with the increasing input of screen-level data into 3D-var itself, an alternative method may be envisaged with an off-line program to perform the soil variable optimal interpolation, forced by an updated low-level 3D-var analysis.
    - Replace the current statistical optimal interpolation technique with the "dynamical optimal interpolation" technique (formerly called "2D-var")
    - Evaluation of the potential to initialise soil variables from an off-line system (like Safran-Isba-Modcou)
    - Research study to prepare the assimilation of satellite observations (infrared surface temperature from Meteosat-8, microwave surface temperature from SMOS and scatterometer from METOP)

7. varpack

Work on the improvement of a nowcasting-oriented 3D-Var version of Aladin will continue. The goal remains the replacement of CANARI for the 3D altitude analysis. Further work must be devoted to the link between analyzed fields and diagnostic "diagpack" outputs like CAPE, MOCON etc… The optimal use of planetary boundary layer data (T2, RH2, U10) will be addressed, and possible adaptations to the varpack B-matrix will be tackled. The work on varpack will be performed in close interaction with

the one for the cycled data assimilation. An interaction with nowcasting work on imagery and pattern recognition techniques is also possible.

8. further recommendations

We list here items that have been discussed, or have been listed elsewhere in the workplan, and which do significantly interact with data assimilation. They generally concern transversality:

- Keep in mind the link between the EPS system and the ensemble generation of a B matrix. In the long term, a convergence should be sought, as both approaches aim at representing the analysis and background errors (after 2008 ?). If such a convergence proves impossible, then the question of why is it impossible or inefficient to do so would probably become a matter of interest for research.
- The code of the TL/AD dynamics strongly depends on the evolution of the non-linear dynamical core. We would propose the following recommendations:
    - The SL TL/AD code could be much more easily developed with the help of SL specialists. It is not more complicated to learn how to code TL and AD, than to learn about the time-stepping and buffering in the SL scheme.
    - Would the NH code be stabilized enough to code the TL/AD ? if more changes are to come in future, of medium or large size, then their reflection into the TL/AD would demand a deep investigation. So far, changes in the NL core had to be re-phased into the TL/AD by the variational community.
- Similar remarks are valid for the simplified physics: it would be unsafe to put the burden of regular updates or developments in the simplified physics on the shoulders of the DA specialists. What we see in leading 4D-var centersis that physics specialists are the triggering force both for the full and simplified parametrizations.

9. Appendix 1: topics for further scientific collaboration with Hirlam

Topic	Contact person in HIRLAM	Contact person in Aladin	Remarks
Wavelet Jb	Thomas Landevius	Alex Deckmyn (Loïk Berre)	Learning stage, followed by work either on diagnostic studies or application to PBL fields
Radar reflectivities	Guenther Haase	Claude Fischer (Eric Wattrelot)	Exact work to be defined in 2006
Doppler radar winds	Kirsti Salonen ?		
Balance conditions in analysis increments			
New humidity control variable (Holm)			
Total liquid water content in control			
Cloud-affected radiances and 1D-Var			

Nota Bene: ***workshop about code convergence and DA coordination to be held in Budapest in the second week of November 2005***.

10. Appendix 2: summary of the Prague discussions, held in the premices of the 4th WMO workshop on D.A.

```
Summary from the informal data assimilation ALADIN-HIRLAM meeting
held along 4th WMO International Symposium on Assimilation of
Observations
in Meteorology and Oceanography, Prague, Czech Republic.

20/04/2005

17 participants

Informal meeting discussion was chaired by C. Fischer. Discussion was
based on
the three papers prepared by C. Fischer, submitted via aladin web-
page:
- status report on completion of 2002-4 DA WP
- draft of 2005-8 DA WP
- 4dvar dreamplan
Concerning HIRLAM, they have scientific plan for 2005, and for future
only "very
open-minded" strategic document.

Besides an information was given on the ALADIN-HIRLAM
coordination-"convergence"
VAR meeting to be held during the 2nd week of November in Budapest.
```

Firstly the 2002-2004 achievements were summarized: 4DVAR was almost completely forgotten (except work of C. Soci); and surface aspects were also missing. On the contrary, in background error modeling, Aladin is quite active. NG noted that the first convergence was already obtained, because similar topics were tackled in HIRLAM.

The draft proposal of the 2005-8 working plan was then commented. For Jb, spectral representation of B matrix is kept. The work on wavelets has started. HIRLAM is also interested in wavelets (but more interest in surface fields and/or the extension zone problem). The "compact support" approach was abandoned. The sampling techniques usage is to be continued. The question was raised how the others can obtain ensemble statistics: either to download Arpege ensemble and run on them dynamical adaptation; or to perturb observations and run Aladin 3DVAR (in a second stage). At ECMWF they plan to run ensemble data assimilation. HIRLAM may use it for ensemble boundaries.

Common interest was found in the area of imbalances.

Work on a posteriori validation and Jk issues started and is to be continued. Mapping (the normalization) of Sigma_b in gridpoints is not used in ALADIN nor in HIRLAM. In Arpege, the mapping from ensemble is considered. Sigma_b randomization planned in Arpege, ready but not used in HIRLAM.

The work of E. Holm (Log of relative humidity in control variable) is to be ported to Arpege and HIRLAM, it shall be easy to do in ALADIN. Total/liquid water in control variable seems to be reasonable to do in HIRLAM (and probably in ALADIN as well). An interest is also in 1dvar, work on cloudy radiances and link with radar (eumetsat and FP6 support in HIRLAM).

The question from HIRLAM was adressed to ALADIN: how the developments from ECMWF model enter ARPEGE/ALADIN ? The system of common cycles, phasing and merging was explained.

Another question was on how the problems with observations wrapping around the model domain was solved. It was answered that it is sleeping (got low priority), but perhaps Jk term or off-diagonal approach could help. Mostly, to remove

observational data in a rim zone and to use Jk will be tried. The off-diagonal term
approach will receive low priority in the next workplan, except if all other techniques
remain useless and the wrap-around is proven to spoil actual forecast performances.

The miniround-table on the status of 3dvar in aladin countries was done:
FR: currently in e-suite with still some open pb, oper planned for June
Mor: waits for radiances
HU: waiting for new machine to become oper (question of weeks)
ROM: plans for implementing 3dvar, possibility to buy HPC
CZ: only plans, ODB big obstacle, too much manpower demanding
SK: plans to install odb and test blending in parallel

Issue of the Observation pre-processing was raised: how many countries do have
a reliable observational database system and management, that should come upstream
of any data assimilation (or verification) system ? (already pointed out by F. Bouttier, see electronic forum). The point was that an application
like 3DVAR still looks unaffordable in many Aladin sites, and whether a marginal
collaboration on data pre-processing should be considered. Following points
came up from the discussions:
- possibility to obtain data from eumetcast (and use eumetcast input to standardize
some data format for any local meteorological database)
- to archive OBSOUL is not so good idea (loss of information but huge memory consumption
because it is ascii)
- OULAN changed too often
- lack of manpower => advice to convince directors that 3dvar is useful, then
manpower can be defended and asked for => for the time being, still not enough positive
signals from the few leading centers in terms of scores and operational applications,
so that a "director-sensitive" publicity cannot be made.
- ODB: installation and understanding is a heavy task and needs a few good specialists
In HIRLAM every country has its own preprocessing system somehow based on
ECMWF software. No use of ODB, but CMA format. In future, HIRLAM plan to interface the ODB
with HIRVAR (which has a flexible observation format interface) as one of the first
tasks in the code convergence exercise.

4DVAR issues were intentionally put appart because of several reasons:
* need to strongly consolidate 3DVAR/ALADIN to prepare MF AROME 2km priorities.
* still not enough adjoint code specialists ?
* the installation of local 3DVAR in the leading centers has certainly taken away

from pure development many precious manpower. The Var community works more in a
tense streamline since 2-3 years.

Also not so much work was done in last period (work of C. Soci with the paper
to appear). The question of possible work deportation was discussed, mainly
the TL/AD version of SL code (remark: in HIRLAM dynamics is GP, but var is
spectral). But generally the work with HIRLAM can be shared (but no precize discussion yet).
The question of how to re-introduce 4DVAR into a reasonable workplan was
tackled (but probably not completely discussed because of time constraints).
*** *** *** *** *** *** *** *** *** *** *** *** *** ***
The "after-meeting proposal" would be basically to copy the previous items,
from the 2002-4 WP:
* SL TL/AD
* 4DVAR in a nutshell
and to let as open issues any research topics that certainly could be tackled,
provided that the project can provide the above-mentioned minimum application:
high frequency radar wind assimilation, simplified physics,
sensitivity studies etc...
However, these research activities will have to be done at the researcher's risks,
in the sense that the ALADIN project cannot garantee full supervision and
system logistics for all these technically demanding topics.
*** *** *** *** *** *** *** *** *** *** *** *** *** ***

Conclusion: Due to lack of time, only B matrix issues and 4DVAR were discussed in more detail,
as concerns the next DA WP.
Anyhow, the contacts in the frame of ALADIN and HIRLAM activities were established.
People are encouraged to consult and contribute to the WP web page.
CF will try to put the priorities to the WP topics and sub-topics. Final WP is
to be delivered on the 15th ALADIN workshop in Bratislava, June 2005.

Editorial team of the meeting,
Maria Derkova
Claude Fischer