BMKG

METEOROLOGICAL TRAINING MODUL

ADVANCE WRF

01.07 DATA ASSIMILATION AND ENSEMBLE FORECAST

Authors:

Dr. Danang Eko Nuryanto

Jaka Anugrah Ivanda Paski S.Tr

EDUCATION AND TRAINING CENTRE

THE AGENCY FOR METEOROLOGY, CLIMATOLOGY AND GEOPHYSICS

2020

FOREWORD

Praised be to God Almighty, teaching materials about of the advance WRF model: data assimilation principles and concepts has been completed. The basic knowledge of data assimilation principles and concepts were provided.

WRF Data Assimilation (WRFDA) is a basic WRF software for combining NWP with product observation data (as the first guess or history forecast) and their respective error statistics to provide an improved atmospheric state estimate (analysis). The WRFDA system is designed to be a modular, state-of-the-art atmospheric data assimilation system that on available parallel computing platforms is portable and powerful.

Observational data is a determining factor in assimilating the data into the WRF model. WRF has been able to process observational data, such as synoptic data, weather radar, and satellite radiation, from many kinds of observations. In this case, the availability and consistency are very important to assimilate data in real time.

Furthermore, input, criticism, and suggestions are needed for future improvements, so that these modules can be kept up-to-date and in tune with developments in science and technology.

Hopefully this teaching material can be useful and I express my gratitude and high appreciation to the drafting team for their contribution and cooperation.

Jakarta, October 2020
HEAD OF EDUCATION AND TRAINING CENTRE
THE AGENCY FOR METEOROLOGY CLIMATOLOGY AND GEOPHYSICS

BADAN METEOROLOGI KLIMATOLOGI DAN GEOFISIKA PUSAT PENDIDIKAN DAN PELATIHAN

Maman Sudarisman

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER I. PREFACE

1.1. Introduction

WRF Data Assimilation (WRFDA) is a specific WRF program to combine observation data with NWP product (the first guess or background forecast) and their respective error statistics to provide an improved estimate (the analysis) of the atmospheric state. Observations are used to make small corrections to a short-range forecast (background), which is assumed to be good, to produce a model analysis.

The WRFDA system is designed to be a flexible, state-of-the-art atmospheric data assimilation system that is portable and efficient on available parallel computing platforms. WRFDA is suitable for use in a broad range of applications, across scales ranging from kilometers for regional and mesoscale modeling to thousands of kilometers for global scale modeling.

Before we can run the WRFDA process we need to prepare observational data as input for WRFDA. Because it takes time to prepare the observational data, it is necessary to pay attention to how long it takes to prepare the data until the WRFDA running process is complete, so that if it will be used for operations it will be very useful. The WRFDA installation and running procedure was provided in this module.

In this module, some advanced WRF application also described for addition knowledge. Instead of using a single prediction, ensemble methods employ a number of predictions. For instance, the multi model ensemble, which combines different models to improve on a single-model forecast, is a common application. Another method utilizes a single model, but disrupts its initial conditions or parameter settings or schemes to achieve various outcomes.

The Ensemble Forecast reflects uncertainty in the initial conditions by generating a series of predictions starting from slightly different states which are similar, but not identical, to our best approximation of the initial atmosphere. Each forecast is based on a model close to our best approximation of the model equations, thus also reflecting the effect of model uncertainties on predictive error.

1.2. Scope of Content

The advance WRF modeling is introduced. We provided an example of data assimilation and ensemble forecast.

1.3. Learning objectives

1.3.1. Goal

After following this lesson, participants are expected to have advance skills regarding WRF modeling techniques, i.e. data assimilation and ensemble forecast.

1.3.2. Objectives

- Participants are able to describe and to explain advance knowledge of WRF about the data assimilation.
- Participants are able to apply the skills in developing WRF modeling mainly in the ensemble forecast.

1.4. Contents

1. Advance WRF : Data Assimilation
   - WRFDA Principles and Concepts
   - WRFDA Techniques (3D-Var and 4D-Var)
   - Bacground of Error (BE) Concept
   - Preparation of Observation Data
   - WRFDA Runnung Procedures
   - Installing WRFPLUS and WRFDA for 4Dvar Run
   - Running Observation Prerecessor (OBSPROC)
   - Running WRFDA
2. Advance WRF : Ensemble Forecast
   - Principles
   - Techniques
   - Types
   - Utilizations
   - Use of Ensemble Forecast in Prediction of Severe Warning and Issue of Warning

CHAPTER II. ADVANCE WRF : DATA ASSIMILATION

Objective : Participants are able to describe and to explain advance knowledge of WRF about the data assimilation.

2.1. WRFDA Principle and Concepts

WRF Data Assimilation (WRFDA) is a specific WRF program to combine observation data with NWP product (the first guess or background forecast) and their respective error statistics to provide an improved estimate (the analysis) of the atmospheric state. Observations are used to make small corrections to a short-range forecast (background), which is assumed to be good, to produce a model analysis.

Various components of the WRFDA system are shown in blue in the sketch below, together with their relationship with the rest of the WRF system.

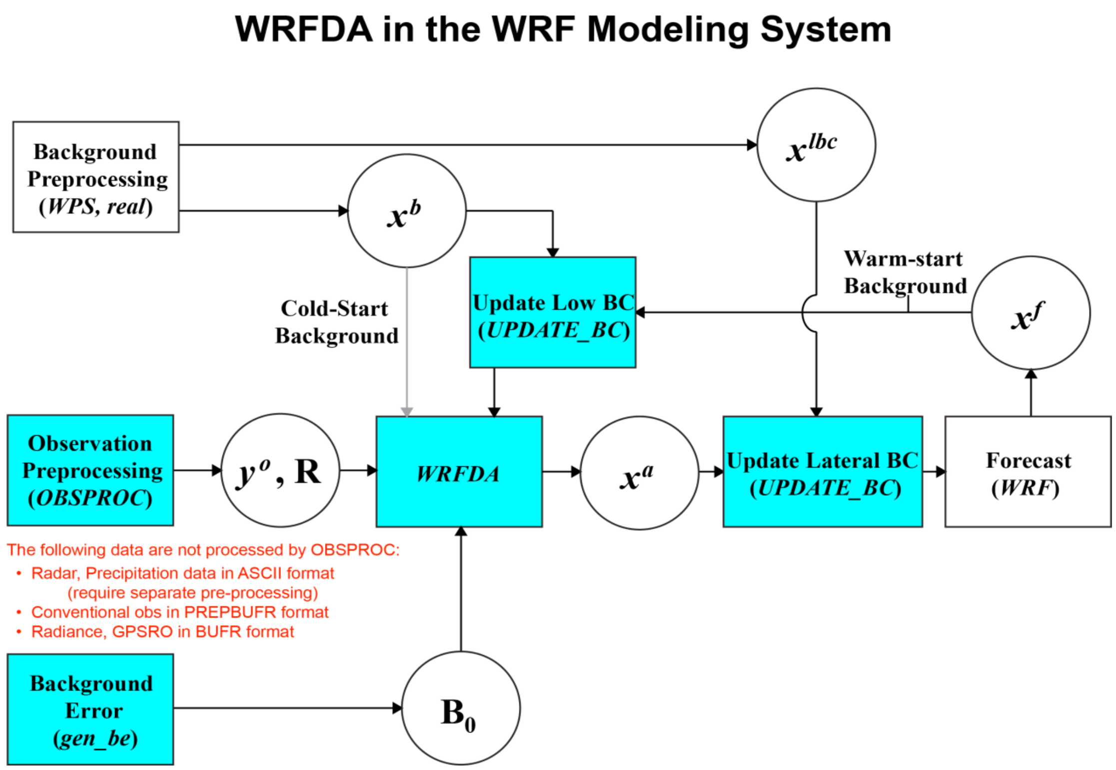


Figure 1. WRFDA process in the system of WRF Model
(sources: http://www2.mmm.ucar.edu/wrf/users/wrfda/index.html)

Table 1. Description of the figure

Label	Description
x^b	First guess, either from a previous WRF forecast or from WPS/real.exe output.
x^{lbc}	Lateral boundary from WPS/real.exe output.
x^a	Analysis from the WRFDA data assimilation system.
x^f	WRF forecast output.
y^o	Observations processed by OBSPROC. (note: PREPBUFR input, radar, radiance, and rainfall data do not go through OBSPROC)
B_0	Background error statistics from generic BE data (CV3) or gen_be.
R	Observational and representative error statistics.

The WRFDA system is designed to be a flexible and sophisticated atmospheric data assimilation system portable and efficient on any available parallel computing platform. WRFDA is suitable for use in various applications, at all scales, from kilometers for regional and mesoscale modeling to thousands of kilometers for global-scale modeling. The MMM Laboratory of NCAR supports a unified model-space data assimilation system (WRFDA) for use by the NCAR staff and collaborators, and is also freely available to the general community, together with further documentation, test results, plans etc., from the WRFDA web-page

(http://www2.mmm.ucar.edu/wrf/users/wrfda/index.html).

2.2. WRFDA Technique (3D-Var and 4D-Var)

In NWP model, there are some assimilation techniques that can be used depending on the method. Thera are statistical methods with 3-Dimensional Variational data assimilation (3D-Var) and 4-Dimensional Variational data assimilation (4D-Var) technique. Another method is advanced methods with Extended Kalman Filter (EKF), Ensemble Kalman Filter (EnKF), and Hybrid VAR/Ens DA technique.

This module chapter will focus on 3D-Var and 4D-Var techniques, which are common and support in WRFDA model. The differences of 3D-Var and 4D-Var techniques has shown on the figure 2.

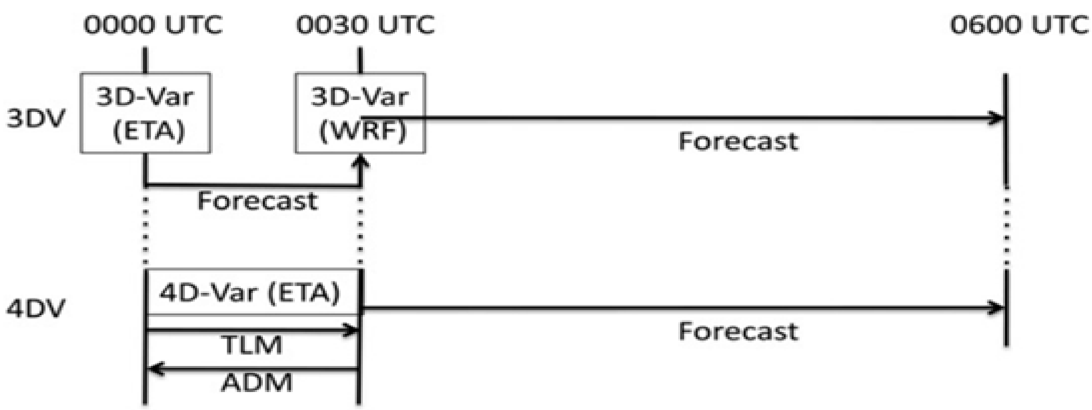


Figure 2. Schematic diagram of 3D-Var and 4D-Var data assimilation experiments. Etamodel and WRF in parentheses indicate the sources of background fields; Tangent Linier Model (TLM) and Adjoint Model Integrations (ADM) (Sun and Wang, 2013)

Three-dimensional variational (3D-Var)

- Observation data is assimilated at a certain time (for example: 00:00 UTC)
- Requires little computational resources

Four-dimensional variational (4D-Var)

- Observation data is assimilated at a certain time range (for example: 00:00 UTC - 00:30 UTC)
- Using TLM and ADM models for propagation analysis in the assimilation time period
- Requires more computational resources than 3D-VAR
- Consider physical and dynamic factors in generating initial conditions (IC) including equilibrium of the convective scale

2.3. Background of Error (BE) Concept

The BE covariance matrix describes the probability distribution function (PDF) of forecast errors, assumed Gaussian. BE is the covariance of (forecast - truth) in analysis control variable space, since the truth is not known (it needs to be estimated). The role of BE is:

3. To spread out the information from the observations.
4. To provide statistically consistent increments at the neighbouring gridpoints and levels of the model.
5. To ensure that observations of one model variable (e.g. temperature) produce dynamically consistent increments in the other model variables (e.g. vorticity and divergence).

Common methods for estimating BE:

- The Hollingsworth and Lönnberg (1986) method - Differences between observations and the background are a combination of background and observation error. - The method tries to partition this error into background errors and observation errors by assuming that the observation errors are spatially uncorrelated.
- The NMC method (Parrish and Derber, 1992) - This method assumes that the spatial correlations of background error are similar to the correlations of differences between 48h and 24h forecasts verifying at the same time.
- The Analysis-Ensemble method (Fisher, 2003) - This method runs the analysis system several times for the same period with randomly-perturbed observations. Differences between background fields for different runs provide a surrogate for a sample of background error.

In the data assimilation method using the WRFDA model both 3D-Var and 4D-Var, the calculation of BE is very important. Background Error is initiation data as model correction data for its prediction. This data is entered as initiation data as well as GFS data for model data and radar data for initiation of observation data. Users have four choices to define the background error covariance options in WRFDA. We call them CV3, CV5, CV6, and CV7. Each of these has different properties, which are outlined in the table below from the WRFDA web-page.

https://www2.mmm.ucar.edu/wrf/users/docs/user_guide_v4/v4.2/users_guide_chap6.html

Table 2. Description of the CV option in WRFDA

CV option	**Data source**	**Control variables**	**cv_options =**
CV3	Provided be.dat file	ψ, χu, Tu, q, Ps,u	3
CV5	GEN_BE	ψ, χu, Tu, RHs, Ps,u	5
CV6	GEN_BE	ψ, χu, Tu, RHs,u, Ps,u	6
CV7	GEN_BE	u, v, T, RHs, Ps	7

With CV3, the control variables are in physical space while with CV5, CV6, and CV7, the control variables are in eigenvector space. The major difference between these two kinds of BE is the vertical covariance; CV3 uses the vertical recursive filter to model the vertical covariance but the others use an empirical orthogonal function (EOF) to represent the vertical covariance. The recursive filters to model the horizontal covariance are also different with these BEs. We have not conducted the systematic comparison of the analyses based on these BEs. However, CV3 (a BE file provided with our WRFDA system) is a global BE and can be used for any regional domain, while CV5, CV6, and CV7 BE's are domain-dependent, and so should be generated based on forecast or ensemble data from the same domain.

2.4. Preparation of Observation Data

In assimilating the data in the WRF model, observation data is a determining factor. WRFDA has been able to process observational data from several types of observations, such as synoptic data, weather radar, and satellite radiation. As shown below.

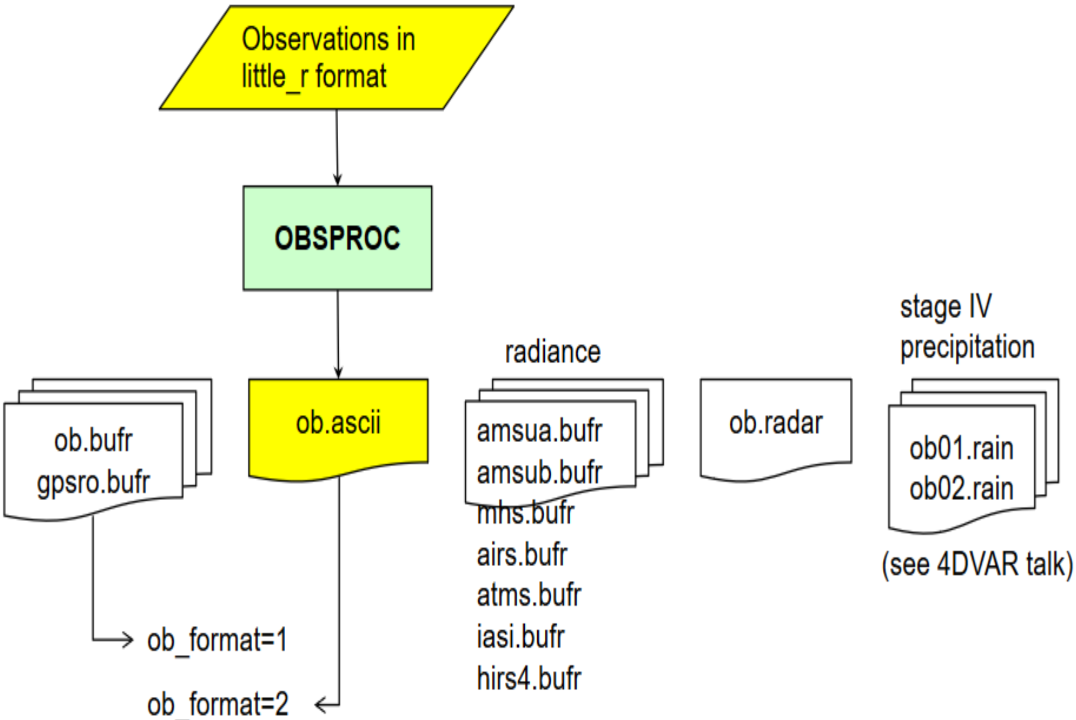


Figure 3. Observation data in WRFDA
(sources: http://www2.mmm.ucar.edu/wrf/users/wrfda/index.html)

Radar Data Assimilation

In recent years, the simulations of Doppler radar data for short-term numerical weather forecasting or nowcasting have become the focal point of research. Assimilate with the 3DVar system to produce a maximum estimate (parameter) of the atmosphere at a time of analysis through an iterative solution of the cost function.

There are some advantages while using radar data as assimilation such as; high spatial and temporal resolutions at the convective-scale, observes wind (radial velocity) as well as microphysics (reflectivity), accurate observations and mostly observation in the lower atmosphere. However, there are also some limitations in radar assimilation, indirect observations, need observation operators, incomplete coverage, limited range and limited detection ability in clear air, also the parameters that can be used are radial velocity and reflectivity with nontrivial for QC

There are two methods to assimilate radar data in WRFDA

1. Direct assimilation of reflectivity (Xiao et al. 2007)
   - Requires an observation operator to link the reflectivity with microphysics
   - No cloud control variables
   - Vertical velocity is diagnosed using the Richardson equation
   - Microphysics are diagnosed using a warm rain partition scheme
2. Indirect assimilation of reflectivity (Wang et al. 2013)
   - Diagnose microphysics (qr, qs, qg) and humidity from reflectivity
   - Assimilate the diagnosed quantities
   - Cloud control variables and vertical velocity control variable

To prepare the observation data from radar (in BMKG), see 2 figures below:

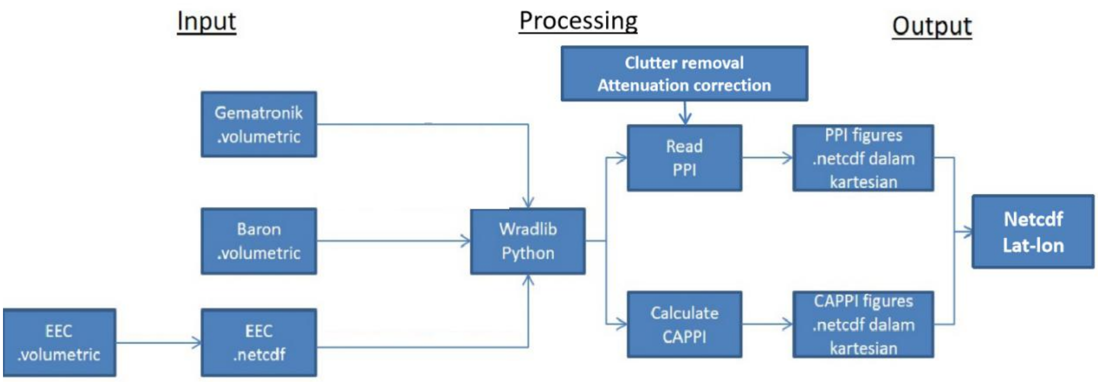


Figure 4. Wradlib processiong procedure

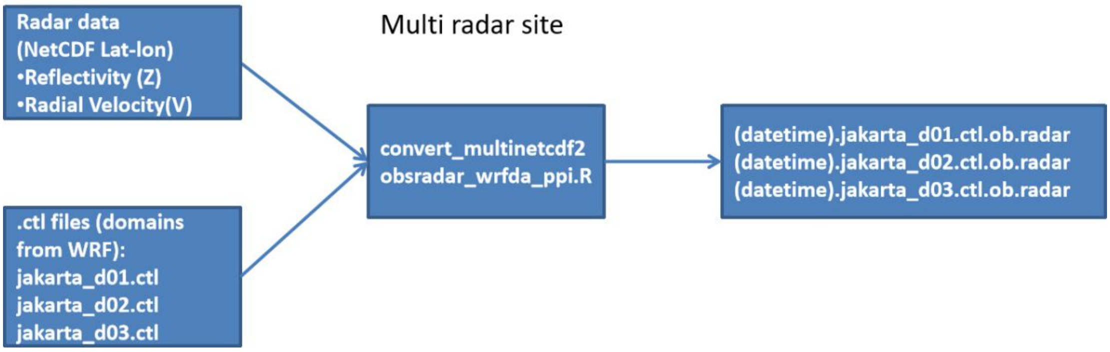


Figure 5. R_Studio processing procedur

C-Band Doppler radar data processing use the wradlib software which has been applied to all radar data BMKG. Wradlib is a language pack (library) Python programming developed by Potsdam University and the University of Stuttgart, German. Furthermore, the radar data is processed use R_Studio software for yields the obs.radar format which becomes then input data assimilation on WRFDA model.

2.5. WRFDA Runing Procedure

Installing WRFDA for 3DVAR Run

1. Obtaining WRFDA Souce Code from :
   http://www2.mmm.ucar.edu/wrf/users/wrfda/download/get_source.html

   Note: Beginning with Version 4.0, the source code for WRFDA and WRFPLUS are contained in the Basic WRF tar file. Although the WRF and WRFDA source code is packaged together, they cannot be built together. To build basic WRF, one would need to download the tar file again and build it separately (in a separate directory). For instructions to build WRFDA versions prior to V4.0, please see:

   http://www2.mmm.ucar.edu/wrf/users/docs/user_guide_V3.9/users_guide_chap6.htm#_Installing_WRFDA_for_1

   After the tar file is unzipped (gunzip WRFV4.1.TAR.gz) and untarred (tar -xf WRFV4.1.TAR), the directory WRF should be created. This directory contains the WRFDA source, the WRFPLUS source (now fully integrated into WRF and TLM/ADJ code, located under the 'wrftladj' directory), external libraries, and fixed files. The following is a list of the system components and content for each subdirectory:

Table 3. Description of the directory files in WRFDA

Directory Name	Content
var/da	WRFDA source code
var/run	Fixed input files required by WRFDA, such as background error covariance, radiance-related files, CRTM coefficients and VARBC.in
var/external	Libraries needed by WRFDA, includes CRTM, BUFR, LAPACK, BLAS
var/obsproc	OBSPROC source code, namelist, and observation error files
var/gen_be	Source code of gen_be, the utility to create background error statistics files
var/build	Builds all .exe files.

2. Compile WRFDA and Libraries. You should set an environment variable NETCDF to point to the directory where your netCDF library is installed

```
> setenv NETCDF your_netcdf_path
```

To use HDF5 in WRFDA, you should set the environment variable “HDF5” to the parent path of your HDF5 build:

```
> setenv HDF5 your_hdf5_path
```

The HDF5 path should contain the directories “include” and “lib”.For some platforms, you may have to also add the HDF5 “lib” directory to your environment variable LD_LIBRARY_PATH:

```
> setenv LD_LIBRARY_PATH

${LD_LIBRARY_PATH}:your_hdf5_path/lib
```

If satellite radiance data are to be used, a Radiative Transfer Model (RTM) is required. The current RTM versions that WRFDA supports are CRTM V2.3.0 and RTTOV V12.1. All the data can be downloaded

http://www2.mmm.ucar.edu/wrf/users/wrfda/download/crtm_coeffs.html.

http://nwpsaf.eu/deliverables/rtm/index.html.

The RTTOV libraries must be compiled with the "emis_atlas" option in order to work with WRFDA; see the RTTOV "readme.txt" for instructions on how to do this. After compiling RTTOV (see the RTTOV documentation for detailed instructions), set the "RTTOV" environment variable to the path where the `lib` directory resides. For example, if the library files can be found in `/usr/local/rttov11/gfortran/lib/librttov11.*.a`, you should set RTTOV as:

```
> setenv RTTOV /usr/local/rttov11/gfortran
```

Note: **Make sure the required libraries were all compiled using the same compiler that will be used to build WRFDA.** Assuming all required libraries are available and the WRFDA source code is ready. For organization and consistency, rename the WRF directory:

```
> mv WRF WRFDA
```

Enter the `WRFDA` directory and run the configure script:

```
> cd WRFDA
> ./configure wrfda
```

To compile WRFDA, type

```
> ./compile all_wrfvar >& compile.out
```

Note: Successful compilation will produce 44 executables: 43 of which are in the `var/build` directory and linked in the `var/da` directory, with the 44th, `obsproc.exe`, found in the `var/obsproc/src` directory. The main executable for running WRFDA is `da_wrfvar.exe`. Make sure it has been created after the compilation: it is fairly common that all the executables will be successfully compiled except this main executable. If this occurs, please check the compilation log file carefully for any errors.

3. Clean old compilation. To remove all object files and executables, type:

```
./clean
```

To remove all build files, including `configure.wrf`, type:

```
./clean -a
```

The `clean –a` command is recommended if your compilation fails, or if the configuration file has been changed and you wish to restore the default settings.

2.6. Installing WRFPLUS and WRFDA for 4DVAR Run

WRFPLUS contains the adjoint and tangent linear models based on a simplified WRF model, which includes a few simplified physics packages, such as surface drag, large scale condensation and precipitation, and cumulus parameterization.

To install WRFPLUS (for versions prior to V4.0, see:

http://www2.mmm.ucar.edu/wrf/users/docs/user_guide_V3.9/users_guide_chap6.htm#_Installing_WRFPLUS_and_2)

1. Beginning with V4.0, the WRFPLUS code is now fully integrated into WRF and TLM/ADJ code, located under the 'wrftladj' directory. To obtain the WRF packaged file, see:

   http://www2.mmm.ucar.edu/wrf/users/download/get_source.html.

2. Unzip and untar the WRF file, rename it appropriately to WRFPLUS, and then run the `configure` script

```
> gunzip WRFV4.1.tar.gz

> tar -xf WRFV4.1.tar

> mv WRF WRFPLUS

> cd WRFPLUS

> ./configure wrfplus
```

   As with 3D-Var, "serial" means single-processor, and "dmpar" means Distributed Memory Parallel (MPI). Be sure to select the same option for WRFPLUS as you will use for WRFDA.

3. Compile WRFPLUS

```
> ./compile wrfplus >& compile.out

> ls -ls main/*.exe
```

   If compilation was successful, you should see the WRFPLUS executable (named `wrfplus.exe`):

4. Finally, set the environment variable `WRFPLUS_DIR` to the appropriate directory:

```
>setenv WRFPLUS_DIR ${your_source_code_dir}/WRFPLUS
```

5. To install WRFDA for the 4D-Var run:

   You should see the same 44 executables as are listed in the above 3DVAR section, including `da_wrfvar.exe`

```
>./configure 4dvar

>./compile all_wrfvar >& compile.out

>ls -ls var/build/*.exe var/obsproc/*.exe
```

2.7. Running Observation Preprocessor (OBSPROC)

The OBSPROC program reads observations in LITTLE_R format (a text-based format, in use since the MM5 era). WRF have provided observations for the tutorial case, see http://www2.mmm.ucar.edu/wrf/users/wrfda/download/free_data.html forthe sources of some freely-available observations.

The purpose of OBSPROC is to:

- Remove observations outside the specified temporal and spatial domains
- Re-order and merge duplicate (in time and location) data reports
- Retrieve pressure or height based on observed information using the hydrostatic assumption
- Check multi-level observations for vertical consistency and super adiabatic conditions
- Assign observation errors based on a pre-specified error file
- Write out the observation file to be used by WRFDA in ASCII or BUFR format

The OBSPROC program (obsproc.exe) should be found under the directory $WRFDA_DIR/var/obsproc/src if "compile all_wrfvar" completed successfully. If you haven't already, you should download the tutorial case, which contains example files for all the exercises in this User's Guide.

(http://www2.mmm.ucar.edu/wrf/users/wrfda/download/testdata.html).

a. OBSPROC for 3D-Var

As an example, to prepare the observation file at the analysis time, all the observations in the range ±1h will be processed, which means that (in the example case) the observations between 23h and 1h are treated as the observations at 0h. This is illustrated in the following figure:

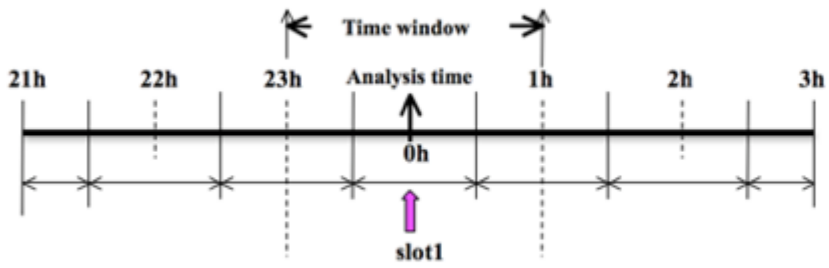


Figure 6. Obsproc for 3D-Var in the system of WRFDA
(sources: http://www2.mmm.ucar.edu/wrf/users/wrfda/index.html)

OBSPROC requires at least 3 files to run successfully:

- A namelist file (namelist.obsproc)
- An observation error file (obserr.txt)
- One or more observation files
- Optionally, a table for specifying the elevation information for marine observations over the US Great Lakes (msfc.tbl)

The files obserr.txt and msfc.tbl are included in the source code under var/obsproc. The following step to run obsporoc

1. To create the required namelist file in the var/obsproc directory. Thus, proceed as follows.

   > cd $WRFDA_DIR/var/obsproc

   > cp namelist.obsproc.3dvar.wrfvar-tut namelist.obsproc

2. Edit the namelist file, namelist.obsproc
3. To run OBSPROC, type

   > ./obsproc.exe >& obsproc.out

Once obsproc.exe has completed successfully, you will see an observation data file, with the name obs_gts_YYYY-MM-DD_HH:NN:SS.3DVAR, in the obsproc directory.

b. **OBSPROC for 3D-Var**

To prepare the observation file, for example, at the analysis time 0h for 4D-Var, all observations from 0h to 6h will be processed and grouped in 7 sub-windows (slot1 through slot7) as illustrated in the following figure:

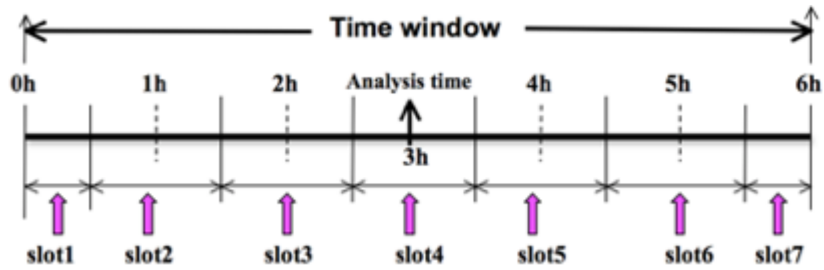


Figure 7. Figure 7. Obsproc for 4D-Var in the system of WRFDA
(sources: http://www2.mmm.ucar.edu/wrf/users/wrfda/index.html)

NOTE: The "Analysis time" in the above figure is not the actual analysis time (0h). It indicates the time_analysis setting in the namelist file, which in this example is three hours later than the actual analysis time. The actual analysis time is still 0h.

1. To create the required namelist file in the var/obsproc directory. Thus, proceed as follows.

   > cd $WRFDA_DIR/var/obsproc

   > cp namelist.obsproc.4dvar.wrfvar-tut namelist.obsproc

2. Edit namelist file, you need to change the following variables to accommodate your experiments. The different values of time_analysis, num_slots_past, and time_slots_ahead contribute to the actual times analyzed. For example, if you set time_analysis = 2008-02-05_16:00:00, and set the num_slots_past = 4 and time_slots_ahead=2, the final results will be the same as before.
3. To run OBSPROC, type

   > obsproc.exe >& obsproc.out

   Once obsproc.exe has completed successfully, you will see 7 observation data files

2.8. Running WRFDA

1. Preaparing Data

   The WRFDA system requires three input files to run:

a) WRF first guess file, output from either WPS/real.exe (`wrfinput`) or a WRF forecast (`wrfout`)
b) Observations (in ASCII format, PREPBUFR or BUFR for radiance)
c) A background error statistics file (containing background error covariance)

Table 4. Summarize of input data using on WRFDA process

Input Data	*Format*	*Created By*
First Guess	NETCDF	WRF Preprocessing System (WPS) and real.exe or WRF
Observations	ASCII (PREPBUFR also possible)	Observation Preprocessor (OBSPROC)
Background Error Statistics	Binary	WRFDA gen_be utility or generic CV3

For example of data:

`ob/2008020512/ob.2008020512` (Observation data in “little_r” format)
`rc/2008020512/wrfinput_d01` (First guess file)
`rc/2008020512/wrfbdy_d01` (lateral boundary file)
`be/be.dat` (Background error file)

2. Run 3D-Var
   a) To run WRF 3D-Var, first create and enter into a working directory (for example, `$WRFDA_DIR/workdir`), and set the environment variable `WORK_DIR` to this directory (e.g., `setenv WORK_DIR $WRFDA_DIR/workdir`). Then follow the steps below:

```
> cd $WORK_DIR

> cp $DAT_DIR/namelist.input.3dvar namelist.input

> ln -sf $WRFDA_DIR/run/LANDUSE.TBL .

> ln -sf $DAT_DIR/rc/2008020512/wrfinput_d01 ./fg
```

```
> ln -sf $DAT_DIR/ob/2008020512/obs_gts_2008-02-05_12:00:00.3DVAR ./ob.ascii
```
(note the different name!)

```
> ln -sf $DAT_DIR/be/be.dat .
```

```
> ln -sf $WRFDA_DIR/var/da/da_wrfvar.exe .
```

b) Edit the file namelist.input, which is a very basic namelist for the tutorial test case,

c) No edits should be needed if you are running the tutorial case without radiance data. If you plan to use the PREPBUFR-format data, change the `ob_format=1` in `&wrfvar3` in `namelist.input` and link the data as `ob.bufr`,

```
> ln -fs

$DAT_DIR/ob/2008020512/gdas1.t12z.prepbufr.nr   o
b.bufr
```

d) Run WRFDA 3D-Var

```
> da_wrfvar.exe >& wrfda.log
```

The file `wrfda.log` (or `rsl.out.0000`, if run in distributed-memory mode) contains important WRFDA runtime log information. Always check the log after a WRFDA run

The file `namelist.output.da` (which contains the complete namelist settings) will be generated after a successful run of `da_wrfvar.exe`. The settings appearing in `namelist.output.da`, but not specified in your `namelist.input`, are the default values from `$WRFDA_DIR/Registry/registry.var`.

After successful completion, `wrfvar_output` (the WRFDA analysis file, i.e. the new initial condition for WRF) should appear in the working directory along with a number of diagnostic files.

3. Run 3D-Var

To run WRF 4D-Var, first create and enter a working directory, such as `$WRFDA_DIR/workdir`. Set the `WORK_DIR` environment variable (e.g. `setenv WORK_DIR $WRFDA_DIR/workdir`)

***Note*: WRFDA 4D-Var is able to assimilate conventional observational data, satellite radiance BUFR data, and precipitation data. The input data format can be PREPBUFR format data or ASCII observation data, processed by OBSPROC.**

a) Link the executable file

```
> cd $WORK_DIR

> ln -fs $WRFDA_DIR/var/da/da_wrfvar.exe .
```

b) Link the observational data, first guess, BE and `LANDUSE.TBL`, `etc`.

```
> ln -fs $DAT_DIR/ob/2008020512/ob01.ascii ob01.ascii

> ln -fs $DAT_DIR/ob/2008020513/ob02.ascii ob02.ascii

> ln -fs $DAT_DIR/ob/2008020514/ob03.ascii ob03.ascii

> ln -fs $DAT_DIR/ob/2008020515/ob04.ascii ob04.ascii

> ln -fs $DAT_DIR/ob/2008020516/ob05.ascii ob05.ascii

> ln -fs $DAT_DIR/ob/2008020517/ob06.ascii ob06.ascii

> ln -fs $DAT_DIR/ob/2008020518/ob07.ascii ob07.ascii

> ln -fs $DAT_DIR/rc/2008020512/wrfinput_d01 .

> ln -fs $DAT_DIR/rc/2008020512/wrfbdy_d01 .

> ln -fs wrfinput_d01 fg

> ln -fs $DAT_DIR/be/be.dat .

> ln -fs $WRFDA_DIR/run/LANDUSE.TBL .

> ln -fs $WRFDA_DIR/run/GENPARM.TBL .
```

```
> ln -fs $WRFDA_DIR/run/SOILPARM.TBL .
> ln -fs $WRFDA_DIR/run/VEGPARM.TBL .
> ln –fs $WRFDA_DIR/run/RRTM_DATA_DBL RRTM_DATA
```

c) Copy the sample namelist

```
> cp $DAT_DIR/namelist.input.4dvar namelist.input
```

d) Edit necessary namelist variables. WRFDA 4D-Var has the capability to consider lateral boundary conditions as control variables as well during minimization. The namelist variable `var4d_lbc=true` turns on this capability. To enable this option, WRF 4D-Var needs not only the first guess at the beginning of the time window, but also the first guess at the end of the time window.

```
> ln -fs $DAT_DIR/rc/2008020518/wrfinput_d01 fg02
```

Please note: WRFDA beginners should not use this option until you have a good understanding of the 4D-Var lateral boundary conditions control. To disable this feature, make sure `var4d_lbc` in namelist.input is set to false.

If you use PREPBUFR format data, set `ob_format=1` in `&wrfvar3` in `namelist.input`. Because 12UTC PREPBUFR data only includes the data from 9UTC to 15UTC, for 4D-Var you should include 18UTC PREPBUFR data as well:

```
> ln -fs
$DAT_DIR/ob/2008020512/gdas1.t12z.prepbufr.nr  ob01.
bufr

> ln -fs
$DAT_DIR/ob/2008020518/gdas1.t18z.prepbufr.nr  ob02.
bufr
```

e) Run WRF 4D-Var

```
> cd $WORK_DIR
> ./da_wrfvar.exe >& wrfda.log
```

4DVAR is much more computationally expensive than 3DVAR, so running may take a while; you can set `ntmax` to a lower value so that WRFDA uses fewer minimization steps. You can also MPI with multiple processors to speed up the execution:

```
> mpirun –np 4 ./da_wrfvar.exe >& wrfda.log &
```

The "mpirun" command may be different depending on your machine. The output logs will be found in files named `rsl.out.####` and `rsl.error.####` for MPI runs.

Please note: If you utilize the lateral boundary conditions option (`var4d_lbc=true`), in addition to the analysis at the beginning of the time window (`wrfvar_output`), the analysis at the end of the time window will also be generated as `ana02`, which will be used in subsequent updating of boundary conditions before the forecast.

Excercise

1. What are the main features of the WRF-DA model?
2. Why WRF-DA is so important to increasing the skill of forecasting?
3. Please explain the simple procedure to running WRF-DA?

CHAPTER III. ADVANCE WRF : ENSEMBLE FORECAST

Objective : Participants are able to apply the skills in developing WRF modeling mainly in the ensemble forecast.

Ensemble methods employ a set of predictions instead of using a single prediction. A common application, for example, is the multimodel ensemble, which combines different models to improve on a single-model forecast. Another approach uses a single model but, to produce multiple results, perturbs its initial conditions or parameter settings or schemes. Various input data which are used as boundary conditions in some cases may also be used to create an ensemble. With data assimilation methods, more complex approaches are used, in which observations are inserted into models to nudge them closer to actual conditions. The famous Ensemble Kalman Filter (EnKF), for example, is an approximation of the Kalman filter from Monte Carlo and is especially useful for highly nonlinear models with uncertain initial states (Mandel, 2009)

3.1. Principles

The Ensemble Forecast reflects uncertainty in the initial conditions by generating a series of predictions starting from slightly different states which are similar, but not identical, to our best approximation of the initial atmosphere. Each forecast is based on a model close to our best approximation of the model equations, thus also reflecting the effect of model uncertainties on predictive error.

The main sources of uncertainty in numerical weather forecast arise from our incomplete knowledge of the exact state of the atmosphere (the initial conditions) and inevitable simplifications in the representation of nature's complexity in numerical weather models. Despite huge developments in the observational network, which involves all sorts of observations ranging from satellite measurements to traditional terrestrial observations, it will still be difficult to characterize the atmosphere without any ambiguity. Similarly, a computational model cannot capture the entire range of all the atmospheric processes and their interactions with the ocean and land surface. For example, the complex processes of vegetation and soil moisture can only be defined by assuming a simplistic definition of vegetation and soil types and the processes associated with them.

The basic principle of ensemble-based probabilistic forecasting is not only to make one forecast from our best guessing initial conditions, but also to carry out a number of additional forecasts starting from slightly disturbed initial conditions, with each forecast being created with a slightly disturbed model. This technique offers an estimation of the uncertainty associated with predictions consistent with measurement errors from a given set of initial conditions. The spread will remain small if the atmosphere is in a stable state; if the atmosphere is less stable the spread will be greater. Reality would fall somewhere in the range expected in a successful ensemble prediction framework (Figure 0). This means users get details about the actual predictability of the weather, i.e., whether a given prediction can be assumed to be certain or not. Also, they get details about the range from which they may expect reality to fall.

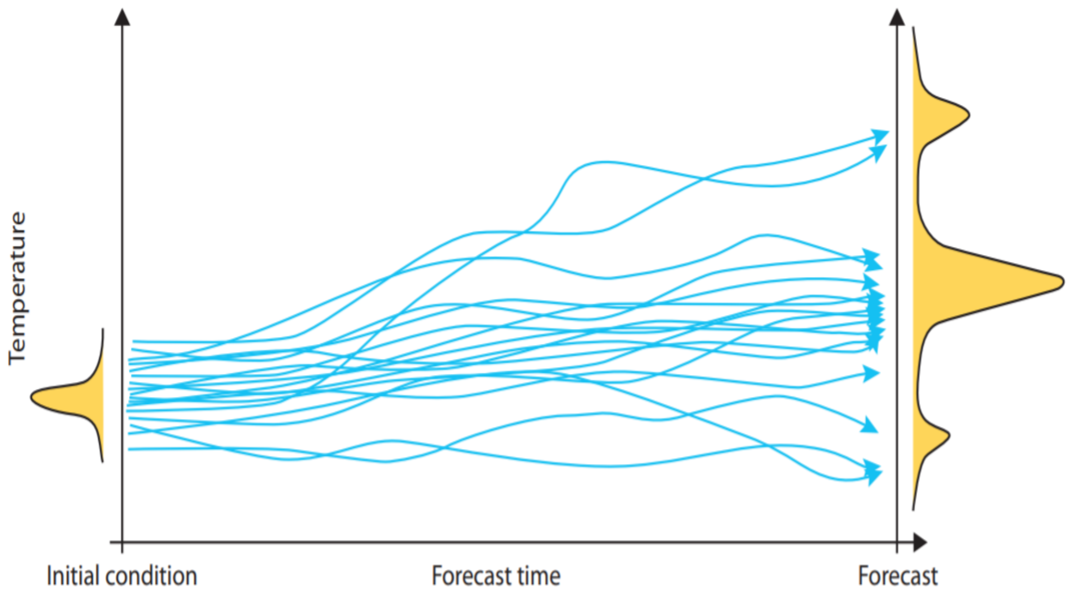


Figure 8. The basic principle of Ensemble Forecast

3.2. Techniques

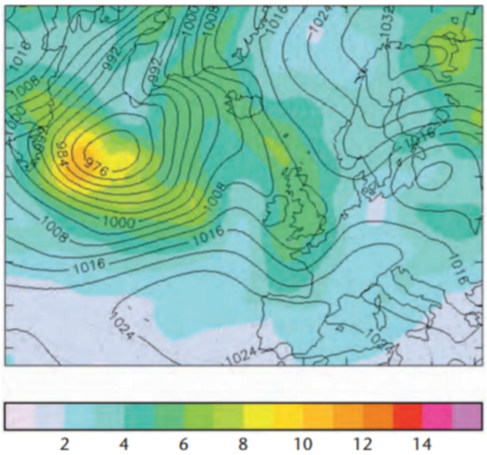
This section describes some of the standard Ensemble Forecast products techniques that are developed from most Ensemble Forecast and briefly how to use them. From most Ensemble Forecast a variety of basic products are derived directly from the fields of model output. This involves usually the following.

1. Ensemble mean

    That is a simple mean of the value of the parameter between all members of the ensemble. The ensemble mean usually verifies by most standard verification scores (root mean squared error, mean absolute error, temporal anomaly correlation coefficient, etc.) better than the control forecast, since it smoothest out unpredictable information and simply introduces the more reliable elements of the forecast. It can provide a reasonable guide to the aspect of the forecast that can be trustfully predicted, but must not be relied upon alone, as it cannot capture the risk of extreme events.

2. Ensemble spread

This is measured as the normal (non-biased) variance of a model output variable, and provides a measure of the degree of uncertainty in a forecast parameter. Often it is plotted overlaid with the ensemble mean on maps. Figure 1 shows both ensemble mean sea level pressure (PMSL) as black contours, and PMSL spread as color shading. The strong colored areas indicate greater spread and therefore less predictability.

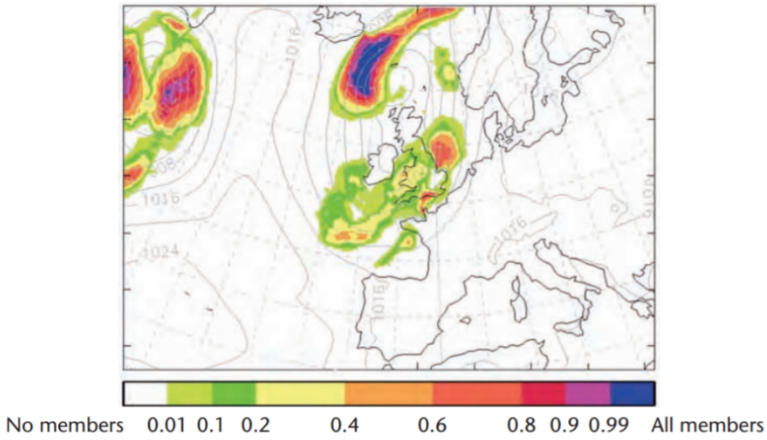


Source: UK Met Office, © British Crown Copyright

Figure 9. Mean (black contours) and spread (color shading) for PMSL forecast (T + 72)

3. Basic Probability

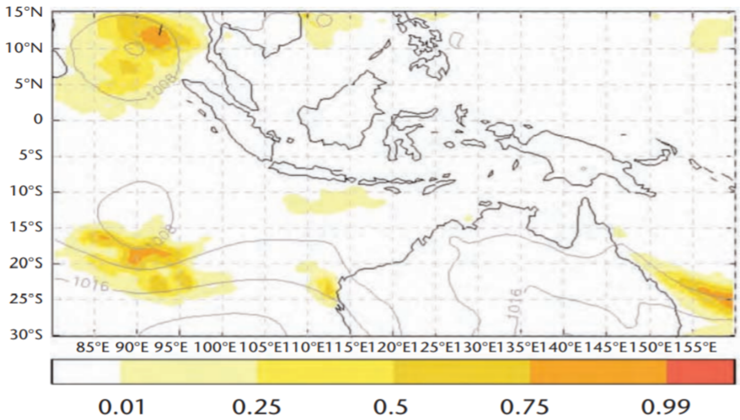
Probability is often estimated as a simple proportion of the members of the ensemble predicting an event to occur at a particular location or grid point, such as 2-m temperature below 0 ° Celsius, or more than one standard deviation below normal. The contoured probability of wind gusts exceeding 40 kt is shown in figure 2. The PMSL mean ensemble is also included as grey contours.



Source: UK Met Office, © British Crown Copyright

Figure 10. Regional MOGREPS probability map for gust speed > 40 kt for 16 July 2010 at 0300 UTC (T + 21 from 15 July 2010 at 0600 UTC); ensemble mean PMSL plotted as faint background

It should be noted that this concept of probability is not a true Bayesian probability as would be determined by a statistician but for practical purposes offers a useful approximation. It presumes that the model correctly represents the climate distribution of an event's occurrence. In order to determine the degree to which forecast probabilities relate to observed frequencies, probability forecasts provided in this way should always be checked over large samples of cases. The example for the project in the South Pacific, part of the larger Severe Weather Forecasting Demonstration Project (SWFDP), was provided in Figure 3.



Source: UK Met Office, © British Crown Copyright

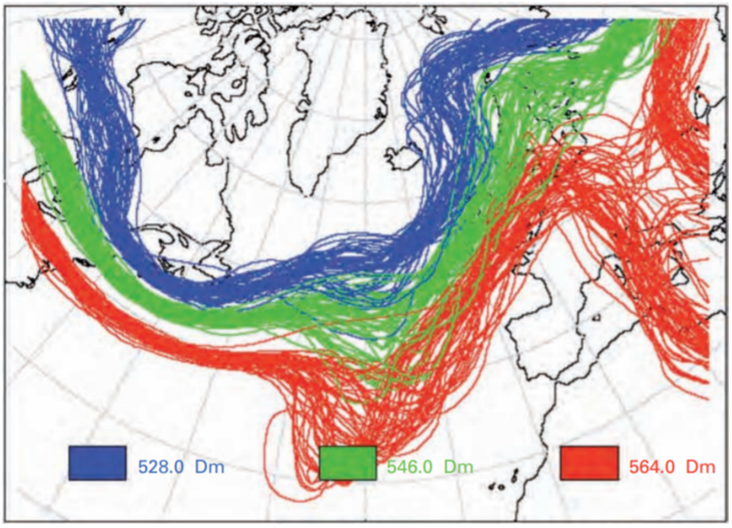
Figure 11. Global MOGREPS probability map for 10-m wind speed > 20 kt for 5 November 2010 at 0000 UTC (T + 48 from 3 November 2010 at 0000 UTC); ensemble mean PMSL plotted as faint background

4. Quantiles

    A collection of ensemble distribution quantiles can give a brief overview of the uncertainty. Commonly used quantiles are maximum and minimum distribution of the ensemble, and percentiles 25th, 50th (median) and 75th. Certain percentiles widely used include the 5th, 10th, 90th and 95th.

5. Spaghetti Map

    Charts showing a few selected contours of variables (e.g., 528, 546 and 564 Dm contours of 500 hPa geopotential height) from all members of the ensemble will provide a useful picture of field predictability. The predictability is higher when all ensemble member contours lie close together; when they look like spaghetti on a plate, there is less predictability (Figure 4).

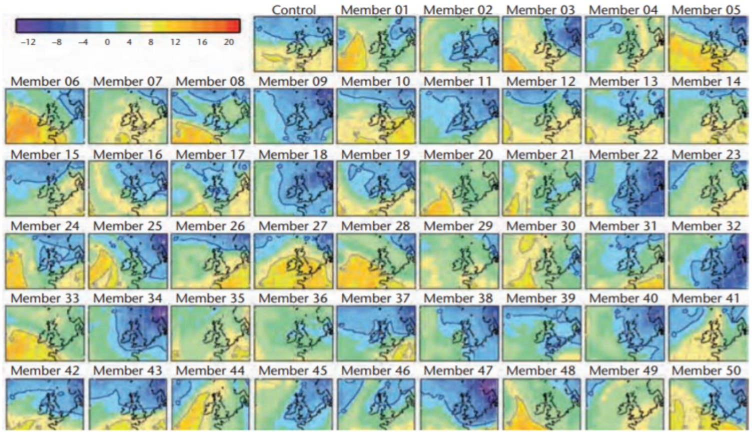


Source: UK Met Office using data from ECMWF, © British Crown Copyright

Figure 12. Ensemble 500 hPa forecast spaghetti charts for 11 February 2001 at 1200 UTC (T + 96 from 7 February 2001 at 1200 UTC) Ensemble 500 hPa forecast spaghetti charts for 11 February 2001 at 1200 UTC (T + 96 from 7 February 2001 at 1200 UTC)

6. Postage

    A series of small maps showing contoured plots of each individual member ensemble (Figure 5) allows the forecaster to display the scenarios in each member's forecast and determine the potential risks from extreme events. But this provides a lot of knowledge that can be hard to assimilate.

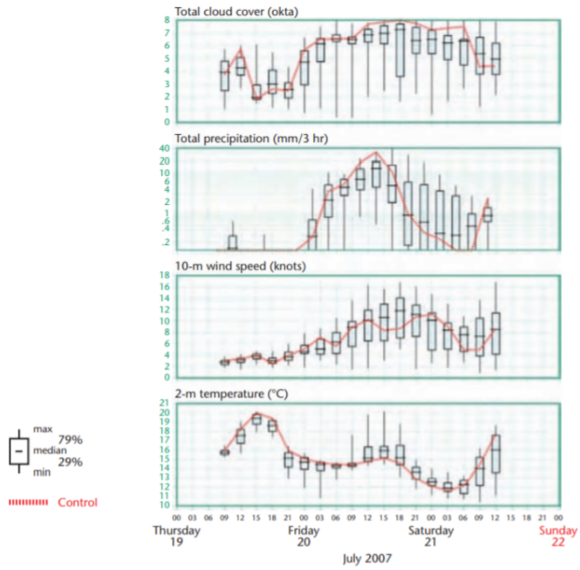


Source: UK Met Office using data from ECMWF, © British Crown Copyright

Figure 13. Postage stamp map for 7 February 2009 at 1200 UTC (850 hPa wet bulb potential temperature, in degrees Celsius; T + 300 from 26 January at 0000 UTC)

7. Site Specific Meteograms

Component performance variables for unique locations can be extracted from grid. There are several displays that can be used at locations to reflect the forecast, such as plume charts and precipitation probabilities. One of the most widely used is the ensemble meteogram that uses a box and whisker plot to show the forecast distribution's principal percentile points for one or more variables (Figure 6).



Source: UK Met Office, © British Crown Copyright

Figure 14. MOGREPS European EPS Meteogram for Brize Norton (51.8°N 1.6°W) from 19 July 2007 at 0900 UTC to 21 July 2007 at 1200 UTC

3.3. Types

There are three fundamental types of Ensemble Forecast for use in weather forecasting – global, regional, and convective scale – and, as with deterministic NWP models, they address different weather forecast timescales. These are briefly described below. There are several variations within each of these categories, such as the way perturbations are produced and the differences in the models used in the models; however, the concepts of how the ensembles are used remain the same, and these specifics are not discussed here.

1. Global Ensemble Forecast

   Global Ensemble Forecast is usually planned and used for a 3–15 day forward medium-range forecast. They use global NWP models which operate at relatively low resolution with average grid lengths ranging from 30 to 70 km. While mainly intended for medium range use, their global coverage means that they can also be used to provide short term Ensemble Forecasts in regions of the globe where no other Ensemble Forecast is available, and may be the only alternative open to many WMO members. Forecasters using global Ensemble Forecast should always remember that the relatively low grid resolutions will limit the details of the forecasts they can expect. Global Ensemble Forecast often cannot solve details like the full strength of wind speed in a storm.

2. Regional Ensemble Forecast

   Regional or Restricted Area Model (LAM) Ensemble Forecast is uses regional models over smaller areas and is based more on a 1–3 day forward short-range forecast. They use higher grid length resolution than the global EPS, usually between 7 and 30 km, enabling them to predict more local weather information and even help solve extreme weather systems. Nevertheless, the forecaster should note the limitations of resolution; for example, the prediction of specifics of small-scale systems such as thunderstorms should not be expected from a regional Ensemble Forecast.

   A regional Ensemble Forecast may take its lateral boundary conditions from a global Ensemble Forecast (the weather systems that come into the area from outside the domain). Some regional Ensemble Forecast use a high-resolution regional analysis and measure corresponding high-resolution disturbances, but others simply take the initial conditions and disturbances from the same global Ensemble Forecast which provides the boundary conditions – this is typically called downscaling. In a downscaling Ensemble Forecast, the forecast must run for a number of hours before the model can "spin up" the detail of the higher resolution.

3. Convective-scale Ensemble Forecast

Convective-scale NWP is now available in a range of more advanced NWP centers, with model grid lengths of 1–4 km operating over relatively small domains. These models, also referred to as convection-permitting, can solve some of the specifics of large convective systems and can therefore attempt to predict information such as thunderstorm position and strength. Although this provides great potential for improved forecasts, convective systems are developing very rapidly and have limited timescales for predictability, so uncertainty can quickly affect the forecasts. Therefore, Ensemble Forecasts are highly important for convective NWP because convective instability introduces a new dimension of forecast uncertainty not addressed by models of lower resolution and with much shorter timescales.

Besides convection itself, models on this resolution have greatly enhanced the ability to forecast other aspects of local weather, such as low cloud and visibility of aviation interest. Many of these phenomena are significantly affected by topographic forcing, which can give greater predictability when the forcing (e.g., slopes, coastlines, vegetation, albedo) can be solved by the models (e.g., convective initiation or valley fog). Convective-scale Ensemble Forecast has the potential to provide predictability information for all of these weather elements.

The much higher convective-scale Ensemble Forecast resolution is expected to enable better resolution of many weather phenomena than is possible with global and regional Ensemble Forecast, for example local topography-forced winds and possibly elements such as low cloud and visibility, especially where such phenomena are forced by local topography or land surface details.

The models are likely to resolve the intensity and spatial scales of local precipitation, especially in convective precipitation, better for precipitation. However, it will take very large ensembles of hundreds or thousands of participants to study the full spectrum of uncertainty in convective precipitation, which would not be feasible in the foreseen future. Therefore, convective-scale Ensemble Forecast is strongly recommended to be post-processed using techniques such as neighborhood processing (If a feature such as a convective shower is assumed to be true but may be mistaken and occur anywhere in the neighborhood within, say, 10 grid lengths of where it appears in the model) to provide a more accurate spatial distribution of probabilities.

3.4. Utilizations

Use of Ensemble Forecast in Deterministic Forecasting

In general, probabilistic forecasts are strongly recommended to provide the best and most accurate weather forecast for customers, and should be encouraged, particularly at longer lead times. However, it is known that many customers need a simple deterministic forecast, and where there is to be a deterministic forecast, the use of an Ensemble Forecast will always provide a more accurate forecast than a single deterministic NWP run. This is particularly true for forecasts more than 1–3 days ahead, and at any time range can help to reduce jumpiness from run-to-run forecast system.

To optimize the deterministic forecast, various Ensemble Forecast indicators can be used. The ensemble means that many standard verification scores will score the best on average, but it should be remembered that it will tend to smooth out the unpredictable detail of the smaller scale, and will rarely capture the intensity of important high impact weather systems. Thus, if the forecast is used for future extreme weather effects, the ensemble estimate should not be used by itself. The median or mode can be another useful guide to the most likely forecast. For single weather parameter these are easier to identify than for the full forecast image.

If a deterministic forecast is to be given, this forecast may sometimes be improved by a declaration of confidence to take advantage of the available knowledge about uncertainty. Trust on all elements of the same prediction will not always be the same. If used, confidence indexes are better supplied separately for each element. The level of confidence should be focused on the distribution of the ensemble but also taking into account the established weaknesses of forecasting skills.

The best approach to providing a deterministic forecast depends on the predictability as shown by the distribution of the ensemble. The distribution could be analyzed using different products such as spaghetti plots and a synoptic-scale map depicting variance and then using meteograms, quantiles, cluster analysis, and so on at the lower scales.

a) Small Small spread in the ensemble (good predictability)
    i. In this case it may be reasonable to offer more detail in the forecast.
    ii. Take the control, the high-resolution control, the ensemble means or the median as a guide (with due regard for the need for calibration or bias correction).
    iii. Spread may often differ between model variables so small spread in one parameter does not guarantee confidence in all aspects of the forecast.

- Good synoptic scale predictability does not always mean predictability in surface weather variables such as temperature or convective precipitation.
- Forecaster should still take account of uncertainty in parameters not resolved by the model.

b) Large spread in the ensemble (poor predictability)
   i. Avoid giving too much detail in the forecast.
   ii. Ensemble mean should be considered, but if the ensemble covers a range of scenarios the ensemble mean will not provide a realistic scenario.
   iii. In that situation, take the most representative member of the ensemble (for example, most populated cluster or mode of pdf) as a guide to the most probable outcome.
      - Note that the most representative ensemble member may not give the most probable value for each weather element (for example, most probable temperature at a location may not be correlated with the most probable precipitation amount).
   iv. The uncertainty assessment.
      - Encourage users to follow forecast updates.
   v. Take into account extremes of the EPS and of the high-resolution control.
      - Make a careful evaluation of the possible evolutions of the synoptic situation and their potential impacts.
      - Take into account the behavior of models.
      - The high-resolution control may be better able to represent certain high-impact events.

c) In the short range (12–18 hours), it may be possible to take into account the latest observations (3–6 hours into the forecast) in order to choose a scenario or a member of the ensemble.
   i. For example, a rapidly evolving cyclone may be best predicted by the member with the best position after a few hours but ONLY in the very short range.
   ii. Be aware that future evolution will be influenced by features coming from upstream.
   iii. This makes member selection for forecasts beyond ~24 hours impossible.
   iv. Also, the consistency of the latest runs with respect to the previous is a factor to take into account.

d) In the longer range, while probabilistic forecasts are best suited, if a deterministic forecast is to be produced, the use of the ensemble mean or median could yield more reliable forecasts, with less jumpiness between runs of the forecast.

3.5. Use of Ensemble Forecast in Prediction of Severe Weather and Issue of Warnings

Severe or high-impact weather events occur on a wide range of space and time scales, from tropical cyclones, extra-tropical cyclones, monsoons, winter storms, and other large-scale systems, to smaller-scale systems such as local severe storms, orographic precipitation, thunderstorms, and tornadoes. Forecasters must take into account the inherent predictability of various types of events (for example, don't try to predict a three-day thunderstorm ahead of time).

A well-structured extreme weather alert system of the National Meteorological and Hydrological Service (NMHS) should have acceptable thresholds, lead times and service level agreed with the users. Thresholds should usually reflect the degree of effect that weather is likely to have on society, including life and property threat, and daily life disturbance. Attributes that should be taken into account in a warning system include:

1. Types of warnings; regions; thresholds (severity/impact and probability).
   Risk = probability x impact
2. A successful warning system is one which users can readily understand, with standard thresholds being adhered to by forecasters.
   Many countries now use a four-color traffic light system (green, yellow, amber and red) indicating different levels of risk and corresponding levels of action that users should take.
3. A successful warning system would need user input on NMHSs. NMHSs, in turn, should provide feedback to producers that would allow them to develop suitable products.

Ensemble Forecasting is an efficient instrument for predicting extreme weather events. The Ensemble Forecast may be used for impact-based alert systems to help estimate the likelihood of weather hazards for use in risk estimation = chance x impact. Nevertheless, Ensemble Forecast will precisely forecast the extreme weather the model or models will resolve for. Otherwise, the following applies:

a) The numerical weather forecast has limitations in explicitly resolving smaller-scale phenomena, leading to underestimation of the likelihood of extreme events within Ensemble Forecast;
b) Ensemble Forecast can sometimes classify prerequisite conditions for extreme innovations or beneficial large-scale environments such as convective indices;
c) Lower resolution Ensemble Forecast (global) is less likely to be able to solve extreme event details;
d) Regional Ensemble Forecast, typically of a higher resolution, can provide more accurate estimates of uncertainty at smaller scales.

Hazard thresholds used in the Ensemble Forecast may need to be calibrated to take account of the above limitations.

Early indications of some extreme events will be predicted in the tail of the ensemble distribution.

Therefore, forecasters and users should not ignore low probability events, especially when those events are very rare.

a) For example, ignoring probabilities below 20 % or even 10 % could result in missing the most important events signaled by the Ensemble Forecast.
b) To be able to use low probabilities, forecasters need verification information.
c) “False alarms” are actually correct features of low probabilities. However, low probabilities may be required in potential high-impact situations.
d) It is expected that the probability will increase closer to the event – usually but not always.

An extreme event may also be forecast essentially correctly, but with errors or uncertainties in location or timing.

Synoptic interpretation (for example, weather feature tracking, and use of analogues) or statistical downscaling tools are ways to add skill to the basic Ensemble Forecast.

a) Note that some statistical methods require large data samples for training, and may not be well suited to rare or extreme events.
b) Cyclone tracking products (for both tropical and extra-tropical cyclones) can provide a useful summary of the development of high-impact storms.
c) There is potential for development of more feature-based diagnostics for poorly resolved severe weather systems.

The Extreme Forecast Index (EFI) can be a useful tool in alerting forecasters to a potential severe event.

a) EFI does not provide explicit probabilities of specific events, and should be interpreted in conjunction with other tools.
b) Currently only a small number of systems can provide an EFI because of the need for a model climatology.

Consideration of input from multiple forecasting systems (Ensemble Forecast and deterministic) may give additional information on the probability of extreme events.

a) Production of verification highlighting the skill and limitations of Ensemble Forecast is important.
    i. Users of Ensemble Forecast should be aware of those limitations and strengths.
    ii. However, because of the rarity of most extreme events it is often impossible to provide reliable (or statistically valid) verification of probabilistic performance. It may be possible to gain some estimate of skill for extreme events by extrapolating from the verification of less-severe events.
b) Given the diminishing of the Ensemble Forecast skill with increasing lead time, the latest available products are generally given higher credibility. However, previous runs of an Ensemble Forecast may still provide useful information about a rare event because of the lack of spread (e.g., limitation in the sample size).

Excercise

1. What are the important of ensemble forecasting?
2. How many techniques of ensemble forecast? Explain!
3. Describe the usability of ensemble forecast to extreme events (severe weather)?

CHAPTER IV. SUMMARY

5.1. Summary

This chapter builds on the information presented in the preceding chapters to provide guidance for main topic to be learn on previous chapter. More comprehensive techniques are used for data assimilation processes, in which findings are introduced into models to nudge them closer to real conditions. The entire range of all atmospheric processes and their interactions with the surface of the ocean and land cannot be captured by a theoretical model. Complex vegetation and soil moisture processes, for instance, can only be identified by assuming a simplified description of the types of vegetation and soil and the processes associated with them.

The fundamental principle of ensemble-based probabilistic forecasting is not only to make one forecast from our best initial conditions, but also to make a number of additional forecasts starting from slightly disturbed initial conditions, with a slightly disturbed model being developed for each forecast. This approach provides an estimate of the uncertainty associated with forecasts consistent with measurement errors from a given set of original conditions.

A number of basic products are taken directly from the fields of model performance from most Ensemble Forecasts, i.e., Ensemble mean, Ensemble spread, Basic probability, Quantiles, Spaghetti maps, Postage stamp maps, and Site-specific meteograms. There are three basic types of Ensemble Forecast for weather forecasting use: global, regional, and convective scale, and they address different weather forecast timescales, as with deterministic NWP models.

5.2. Recommendation

In general, to provide the best and most reliable weather forecast for clients, probabilistic forecasts are strongly recommended and should be encouraged, particularly at longer lead times. It is understood, however, that many clients need a simple deterministic forecast, and where a deterministic forecast is to be made, the use of an Ensemble Forecast will often provide a more precise forecast than a single NWP deterministic run.

REFERENCES

Mandel, J. (2009). A brief tutorial on the ensemble Kalman filter, ArXiv e-prints: https://arxiv.org/abs/0901.3725.