Remote sensing for assimilation and validation of dust forecasts

Lecturer: Dr Stavros Solomos
Post-Doc Researcher IAASARS/NOA

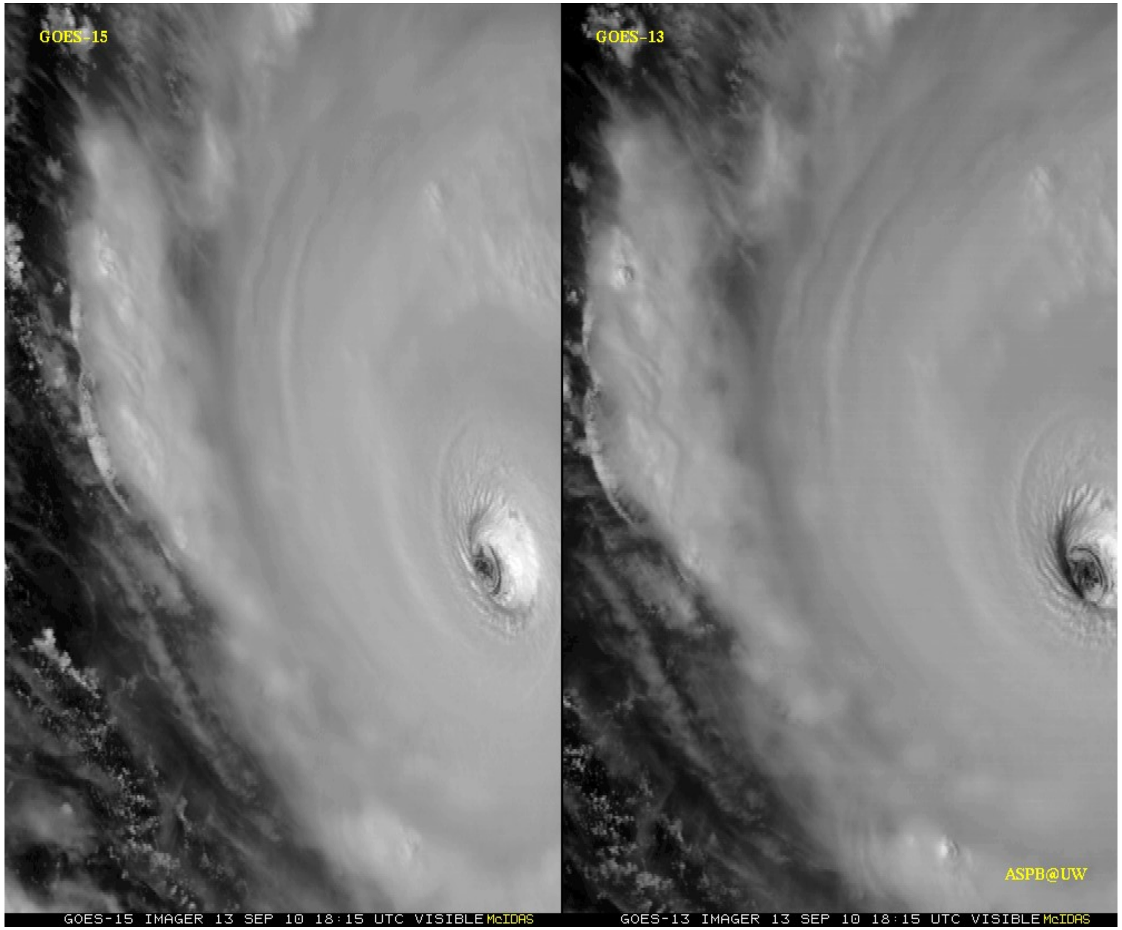
Agios Nikolaos Crete 4 April 2017

Scope of this lecture

Basic understanding of :

- Atmospheric Modeling Principles
- Dust Models
- Model - Remote Sensing Synergies
- Evaluation
- Assimilation

Storms – Hurricanes – Storm surge



Temporal resolution
Advanced Baseline Imager (ABI)

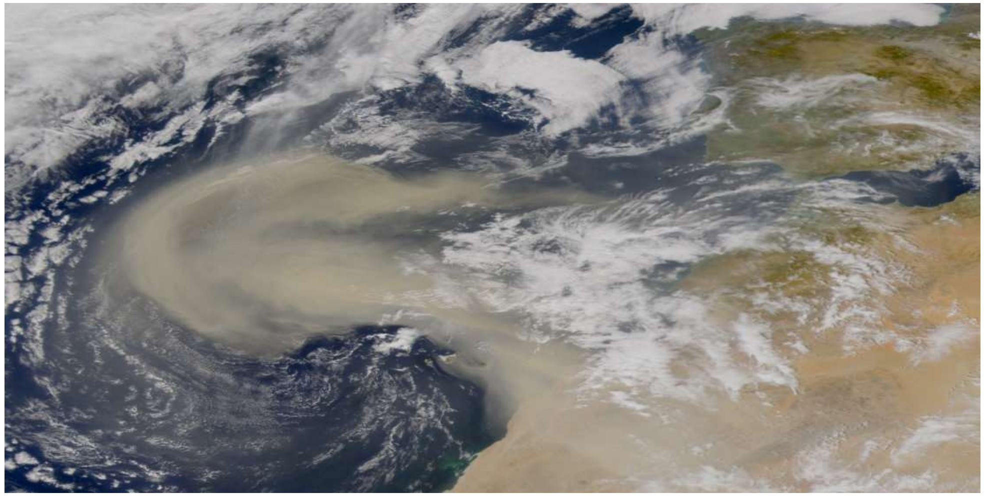
Hurricane Igor, 2010

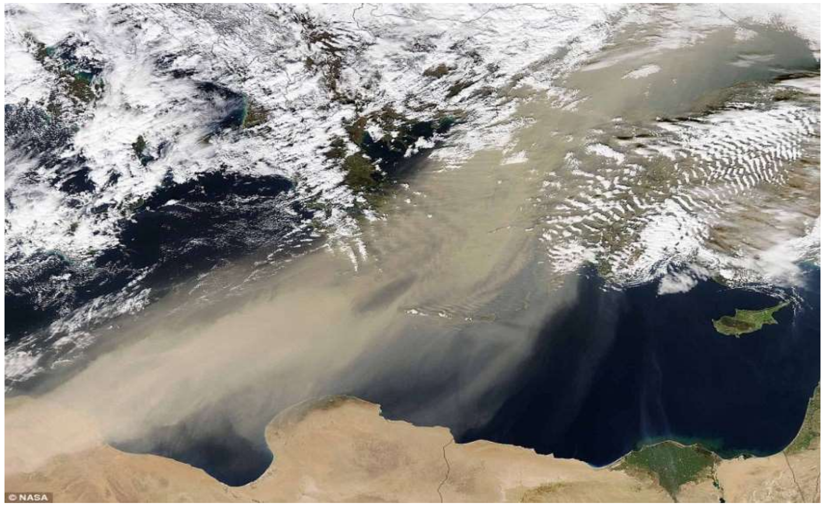
- 1min GOES-15
- 15min GOES-13

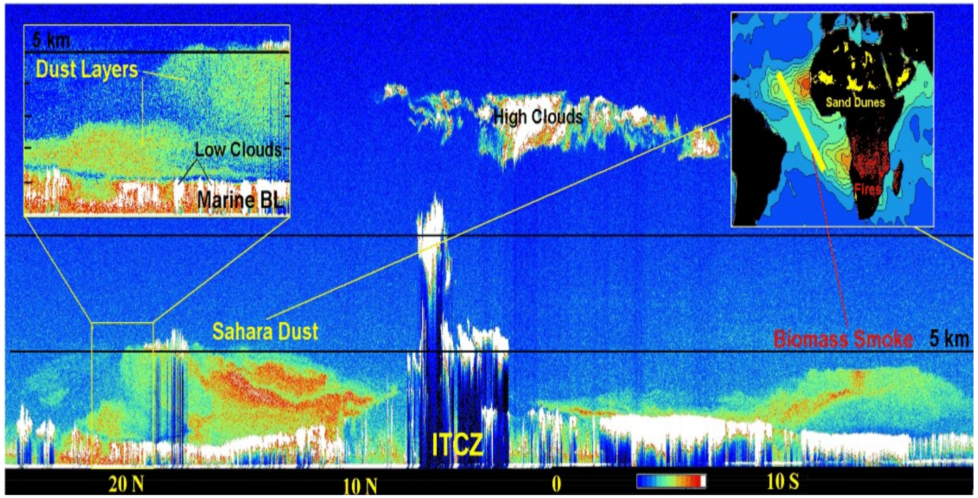
Hurricane Igor (2010) Imagery curtesy of ASPB http://www.goes-r.gov/users/comet/EUMETSAT/at_dust/media/flash/aeolian.swf

Desert Dust Aerosol

Mobilization of dust (Saltation & Bombardement mechanism)



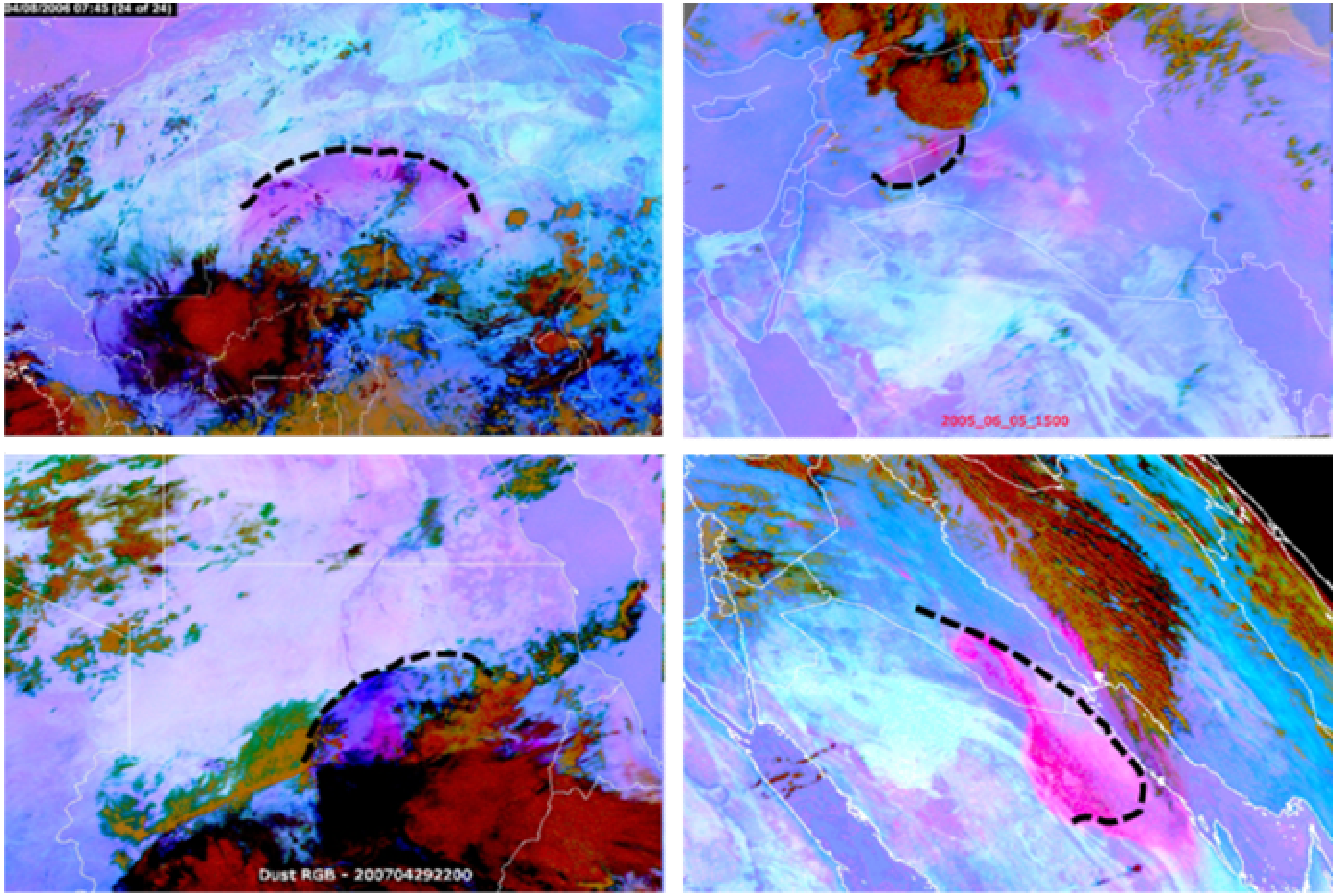




Passive & active space-borne observations of dust

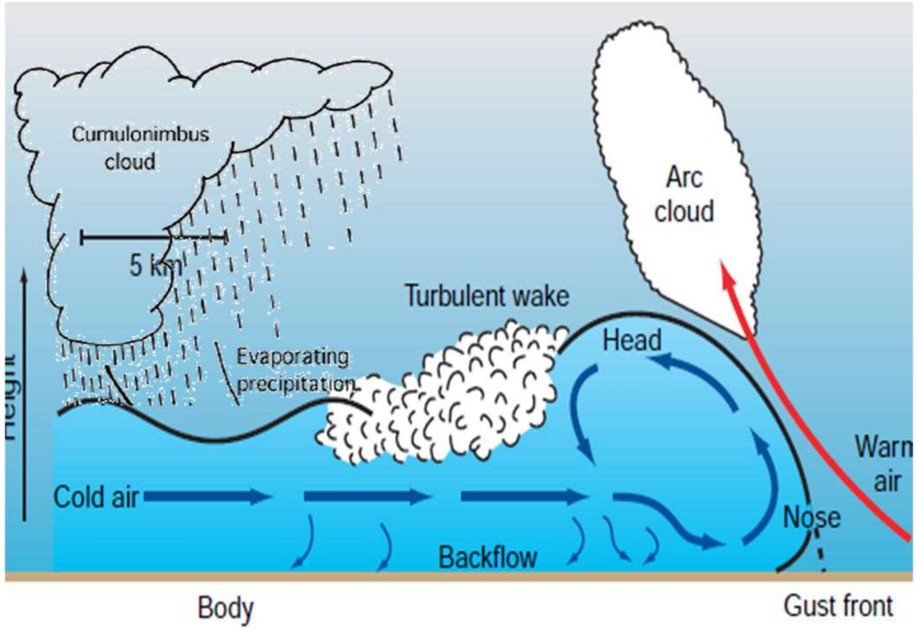
Dust - Haboobs

Dust - Haboobs



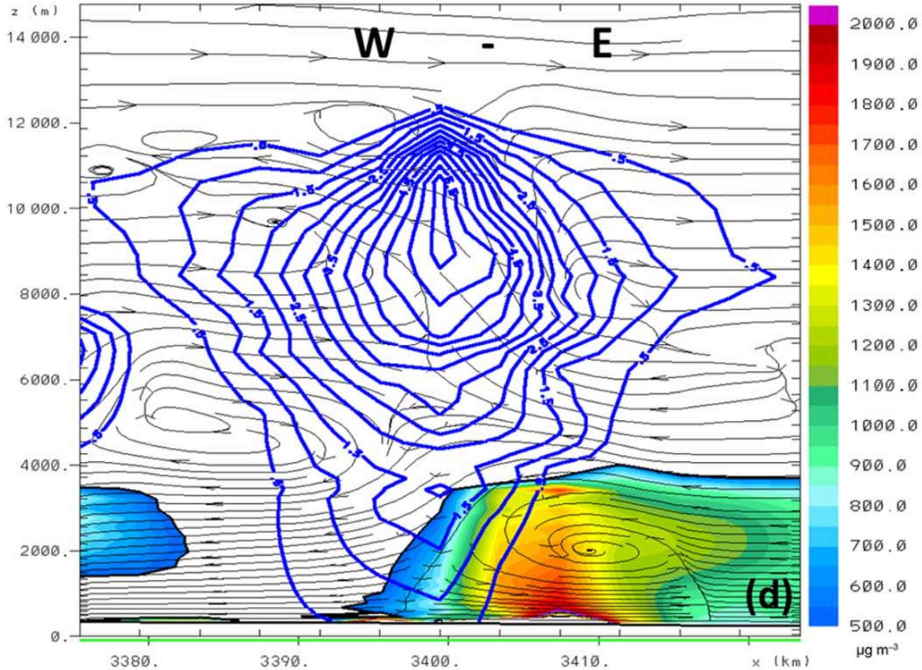
Generation of haboobs by Mesoscale Convective Systems (MCS)
MSG-SEVIRI dust product

Dust - Haboobs



Schematic diagram of a density current formation

Adopted from Knippertz et al., 2007, JGR



Model reproduction of a density current formation and elevated dust concentration

Adopted from Solomos et al., ACP, 2017

Introduction to Numerical Weather Prediction (NWP)

Navier–Stokes Equations
3 – dimensional – unsteady

Glenn Research Center

Coordinates: (x,y,z)
Velocity Components: (u,v,w)

Time : t
Density: ρ
Total Energy: Et

Pressure: p
Stress: τ

Heat Flux: q
Reynolds Number: Re
Prandtl Number: Pr

Continuity:

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} + \frac{\partial(\rho w)}{\partial z} = 0$$

X – Momentum:

$$\frac{\partial(\rho u)}{\partial t} + \frac{\partial(\rho u^2)}{\partial x} + \frac{\partial(\rho uv)}{\partial y} + \frac{\partial(\rho uw)}{\partial z} = -\frac{\partial p}{\partial x} + \frac{1}{Re_r}\left[\frac{\partial \tau_{xx}}{\partial x} + \frac{\partial \tau_{xy}}{\partial y} + \frac{\partial \tau_{xz}}{\partial z}\right]$$

Y – Momentum:

$$\frac{\partial(\rho v)}{\partial t} + \frac{\partial(\rho uv)}{\partial x} + \frac{\partial(\rho v^2)}{\partial y} + \frac{\partial(\rho vw)}{\partial z} = -\frac{\partial p}{\partial y} + \frac{1}{Re_r}\left[\frac{\partial \tau_{xy}}{\partial x} + \frac{\partial \tau_{yy}}{\partial y} + \frac{\partial \tau_{yz}}{\partial z}\right]$$

Z – Momentum

$$\frac{\partial(\rho w)}{\partial t} + \frac{\partial(\rho uw)}{\partial x} + \frac{\partial(\rho vw)}{\partial y} + \frac{\partial(\rho w^2)}{\partial z} = -\frac{\partial p}{\partial z} + \frac{1}{Re_r}\left[\frac{\partial \tau_{xz}}{\partial x} + \frac{\partial \tau_{yz}}{\partial y} + \frac{\partial \tau_{zz}}{\partial z}\right]$$

Energy:

$$\frac{\partial(E_T)}{\partial t} + \frac{\partial(uE_T)}{\partial x} + \frac{\partial(vE_T)}{\partial y} + \frac{\partial(wE_T)}{\partial z} = -\frac{\partial(up)}{\partial x} - \frac{\partial(vp)}{\partial y} - \frac{\partial(wp)}{\partial z} - \frac{1}{Re_r Pr_r}\left[\frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y} + \frac{\partial q_z}{\partial z}\right]$$

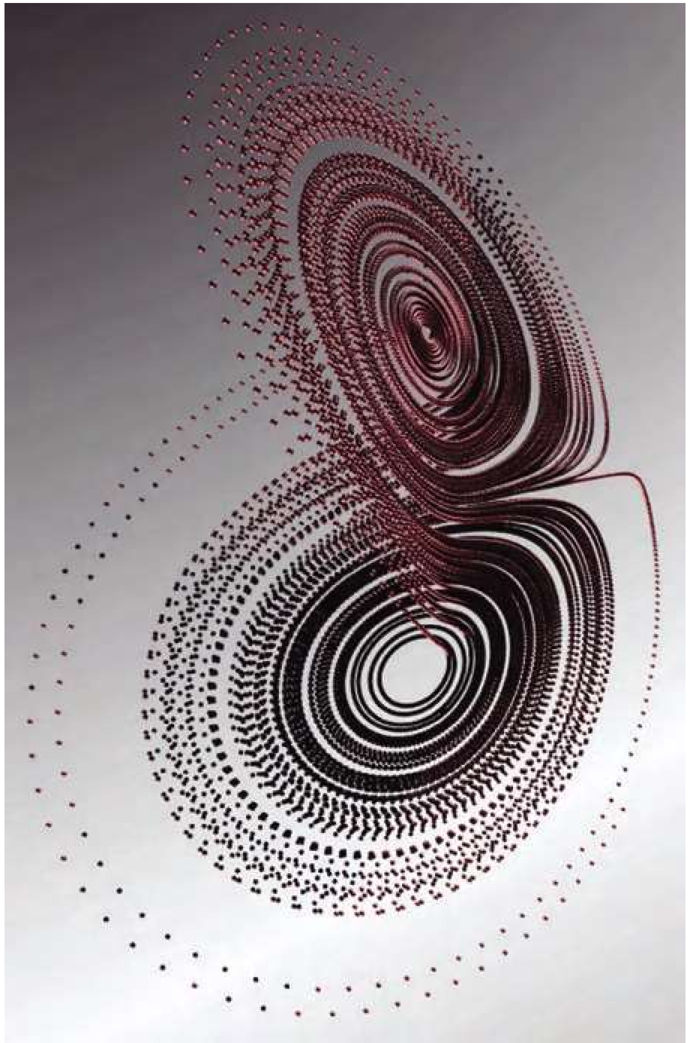
$$+ \frac{1}{Re_r}\left[\frac{\partial}{\partial x}(u\,\tau_{xx} + v\,\tau_{xy} + w\,\tau_{xz}) + \frac{\partial}{\partial y}(u\,\tau_{xy} + v\,\tau_{yy} + w\,\tau_{yz}) + \frac{\partial}{\partial z}(u\,\tau_{xz} + v\,\tau_{yz} + w\,\tau_{zz})\right]$$

Weather Forecast – Numerical Prediction

In 1960, Professor Edward N. Lorenz in the Department of Meteorology at MIT decided to rerun an experiment with a simplified atmospheric model in order to extend his "weather forecast" farther out into the future. To his surprise, he found that he was **unable to duplicate his previous forecast.** Even though the code and the prescribed initial conditions in the two experiments were identical, the states of the model in the two simulations were different.

Atmospheric motions are inherently unpredictable as an initial value problem (i.e., as a system of equations integrated forward in time from specified initial conditions) beyond a few weeks. Beyond that time frame, uncertainties in the forecasts, no matter how small they might be in the initial conditions, become as large as the observed variations in atmospheric flow patterns. Such exquisite sensitivity to initial conditions is characteristic of a broad class of mathematical models of real phenomena, referred to as **chaotic nonlinear systems** .

The history of the state of the model used by Lorenz can be represented as a trajectory in a three-dimensional space defined by the amplitudes of the model's three dependent variables. Regime-like behavior is clearly apparent in this rendition. Oscillations around the two different "climate attractors" correspond to the two, distinctly different sets of spirals, which lie in two different planes in the three-dimensional phase space. Transitions between the two regimes occur relatively infrequently.



Nature, 406, p. 949 (2000 Courtesy of Paul Bourke)

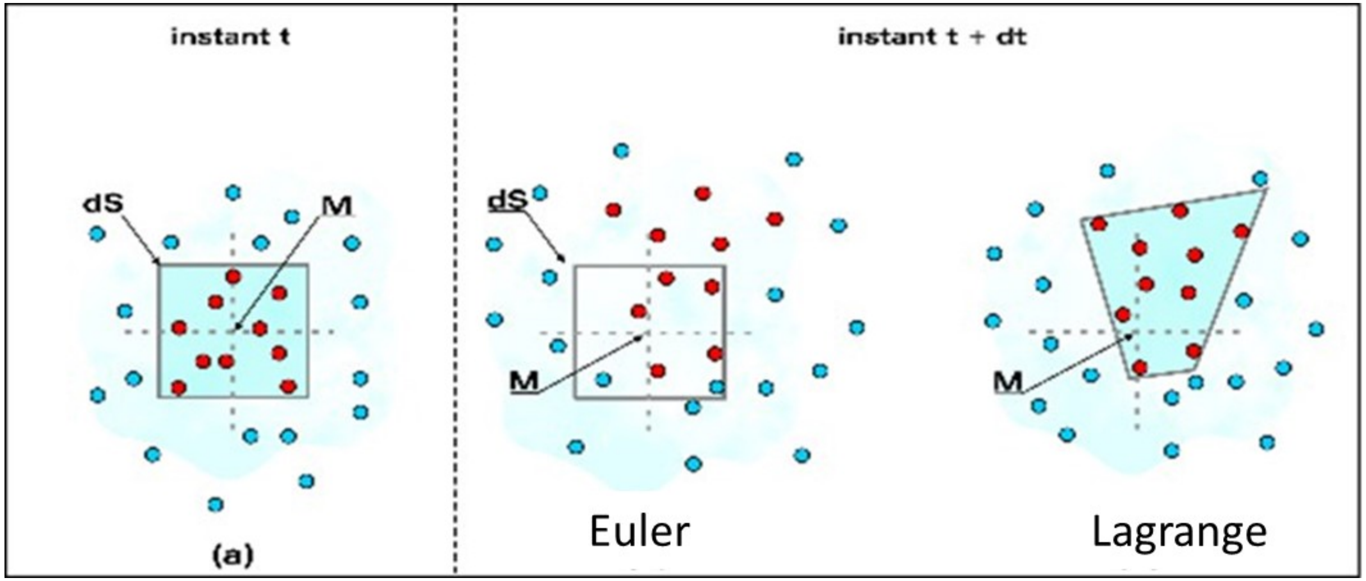
Introduction to Numerical Weather Prediction (NWP)

Lagrangian Description of Flow

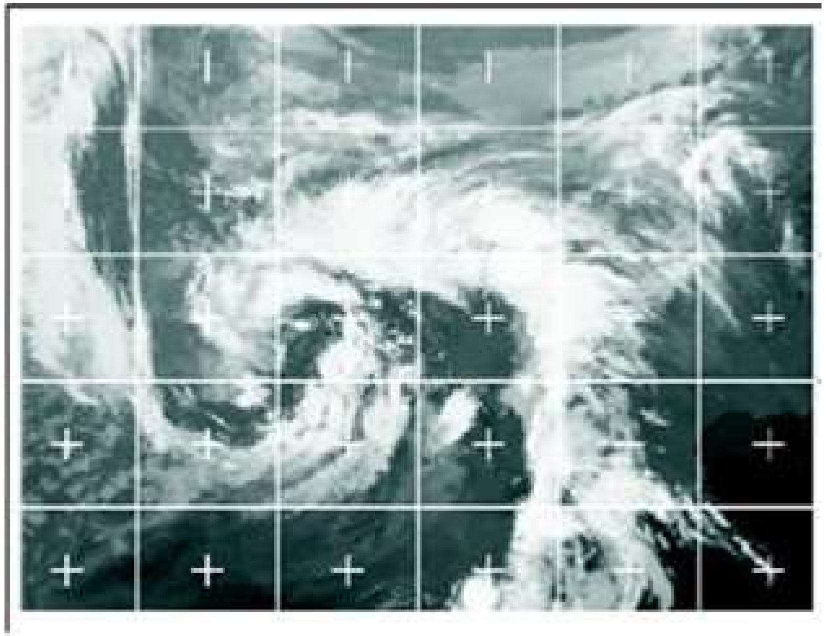
- We follow individual fluid particles (tracers)
- As the particles move their positions and velocities change with time
- The physical laws apply directly to each particle

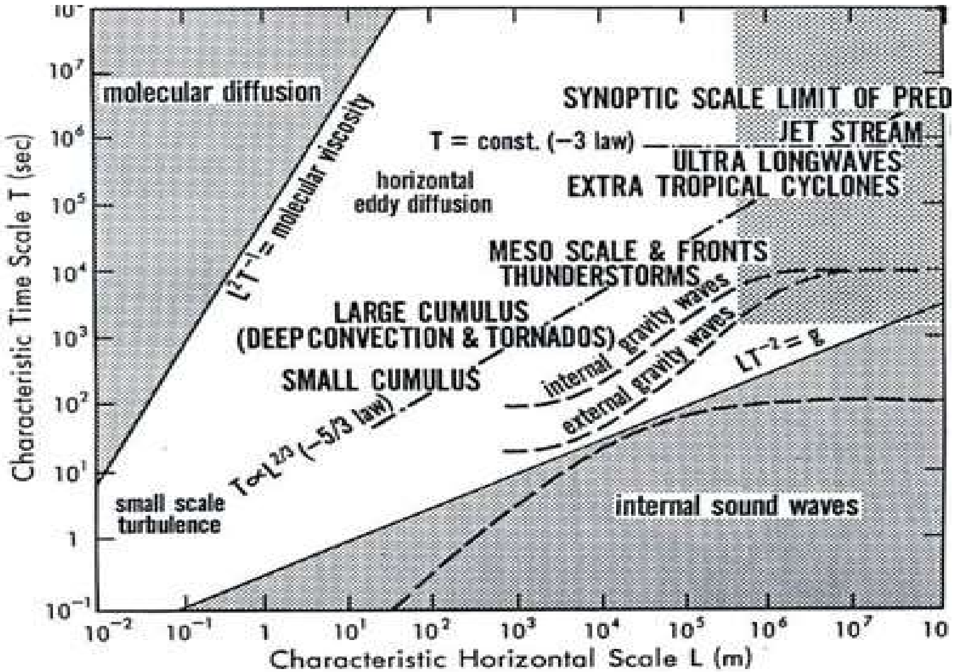
Eulerian Description of Flow

- We define a finite space grid
- The properties of each grid cell change with time
- The physical laws are reformulated to an Eulerian format



Introduction to Numerical Weather Prediction (NWP)





- Practically speaking we need at least 10 grid points to describe a physical phenomenon.
- For example in order to resolve the development of a 20 km diameter convective cloud (Cb) this yields a model grid resolution of 2 × 2 km
- Sub-grid parameterizations for small scale effects
- Convective parameterization remains the biggest problem in atmospheric models

Introduction to Numerical Weather Prediction (NWP)

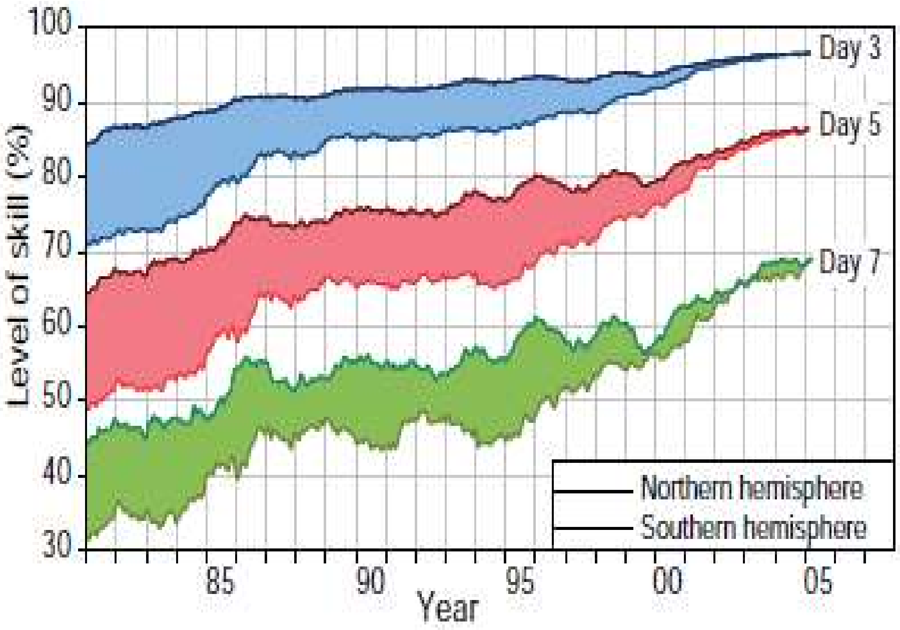
- Most of the important development of primary atmospheric physical processes in NWP models was accomplished by 1990
- Currently we describe everything we know about atmospheric processes (actually, models have mostly caught up with our ability to observe the atmosphere)
- Most important NWP development in past 15-20 years: Cheap computer power (PC, Workstations, Supercomputers) and Multi-processing
- Higher resolution improves model topography, coastlines, treatment of physical processes

Introduction to Numerical Weather Prediction (NWP)

- When using coarse resolution (> 10 km), important weather events (e.g., thunderstorms) are not simulated explicitly
- Need of “parameterizations”
- If a parameterization gives an indication that a forecast thunderstorm occurred in a 10x10 km grid cell, and it actually happened, it was considered a good forecast
- With high resolution (100 m), if a thunderstorm is forecast to occur 200m west of a road, but it actually occurred 200m east of the road:
  - A good forecast?
  - Two bad forecasts?

Forecast Skill

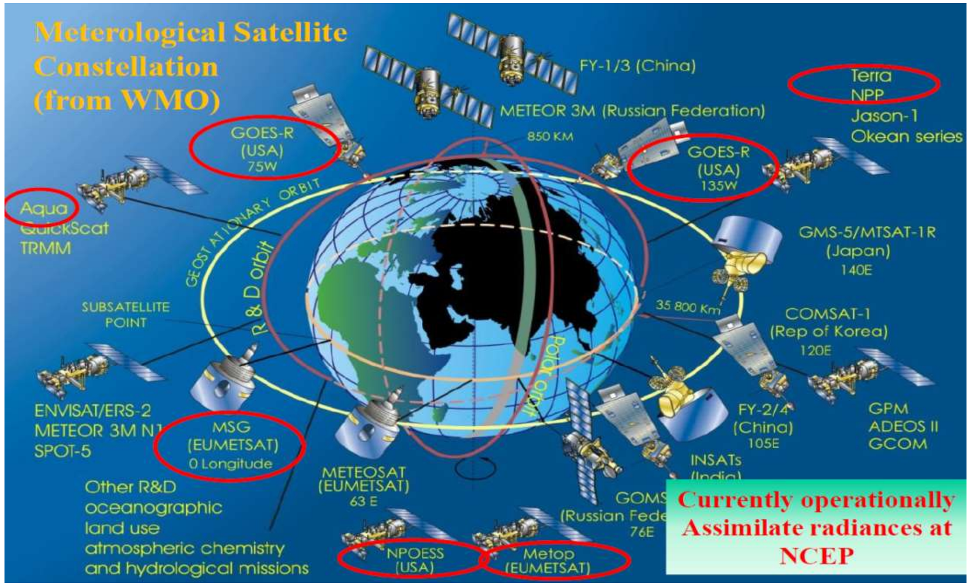
Fig. 1.1 Improvement of forecast skill with time from 1981 to 2003. The ordinate is a measure of forecast skill, where 100% represents a perfect forecast of the hemispheric flow pattern at the 5-km level. The upper pair of curves is for 3-day forecasts, the middle pair for 5-day forecasts, and the lower pair for 7-day forecasts. In each pair, the upper curve that marks the top of the band of shading represents the skill averaged over the northern hemisphere and the lower curve represents the skill averaged over the southern hemisphere. Note the continually improving skill levels (e.g., today's 5-day forecasts of the northern hemisphere flow pattern are nearly as skillful as the 3-day forecasts of 20 years ago). The more rapid increase in skill in the southern hemisphere reflects the progress that has been made in assimilating satellite data into the forecast models. [Updated from *Quart. J. Royal Met. Soc.*, **128**, p. 652 (2002). Courtesy of the European Centre for Medium-Range Weather Forecasting.]



Initial and boundary conditions – Forecast window

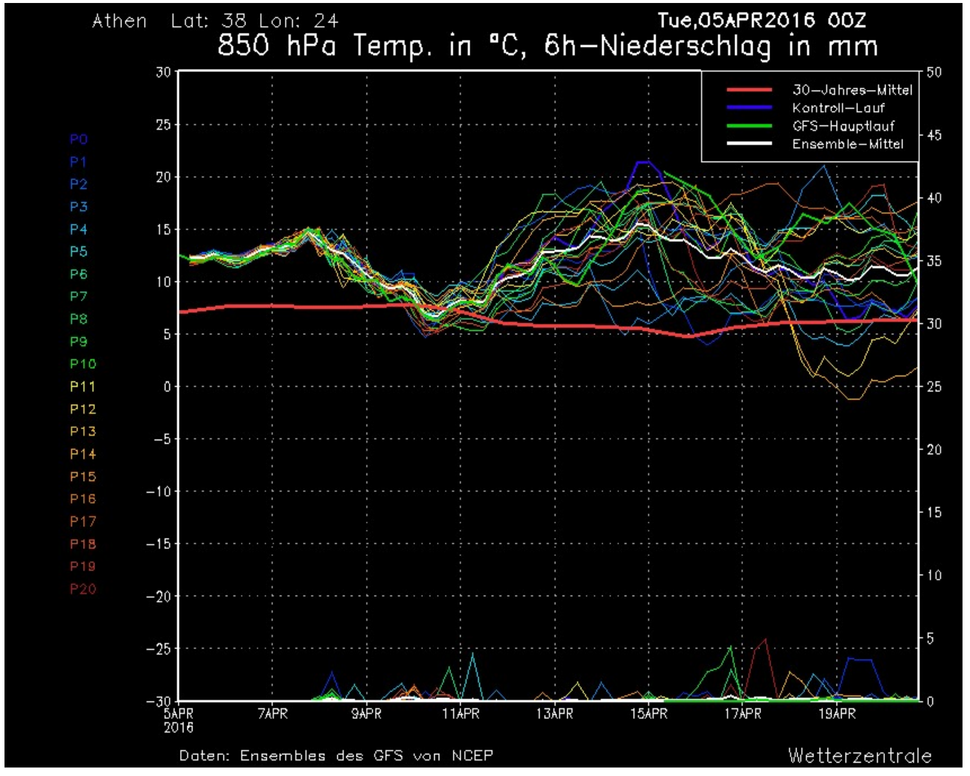
- The initial conditions for modern numerical weather prediction are based on an array of global observations, an increasing fraction of which are remote measurements from radiometers carried on board satellites.

- In situ observations include surface reports, radiosonde data, and flight level data from commercial aircraft. In situ measurements of pressure, wind, temperature, and moisture are combined with satellite-derived radiances in dynamically consistent, multivariate four-dimensional data assimilation systems.



Nevertheless, there will always remain some degree of uncertainty (or errors) in the initial conditions and, due to the nonlinearity of atmospheric motions, these errors inevitably amplify with time. Beyond some threshold forecast interval the forecast fields are, on average, no more like the observed fields against which they are verified than two randomly chosen observed fields for the same time of year are like one another. For the extratropical atmosphere this so-called limit of deterministic predictability is believed to be on the order of 2 weeks.

Ensemble Forecasts



Athens, Greece GFS Ensemble, 05 April 2016, 00UTC Temperature at 850 hPa (in °C) and 6h accumulated precipitation (in mm)

- Forecast models, as well as perturbed initial conditions, are used to generate different members of the ensemble.
- At times when the entire hemispheric circulation is relatively predictable, members of the ensemble do not diverge noticeably from one another until relatively far into the forecast.
- Often the errors grow most rapidly over one particular sector of the hemisphere due to the presence of local instability in the hemispheric flow pattern.
- The rate of divergence of the individual members of the ensemble provides a measure of the credibility of the forecasts in various sectors of the hemisphere and the length of the time interval over which the forecasts can be trusted.

Ensemble Forecasts

- The results are not as easy to interpret as those for the idealized model based on the Lorenz attractor, but they are nonetheless informative.
- As in the idealized experiments, the ensemble forecasts also provide an indication of the range of atmospheric states that could develop out of the observed initial conditions.

- The mean is considerably smoother because it represents an average over 50 individual forecasts.
- Some of the individual forecasts, like the one in the lower left panel, capture the features in the verifying analysis with remarkable fidelity.
- Unfortunately, there is no way of identifying these highly skillful forecasts at the time that the ensemble forecast is made.

ECMWF 7-day ensemble forecasts for a typical winter day.

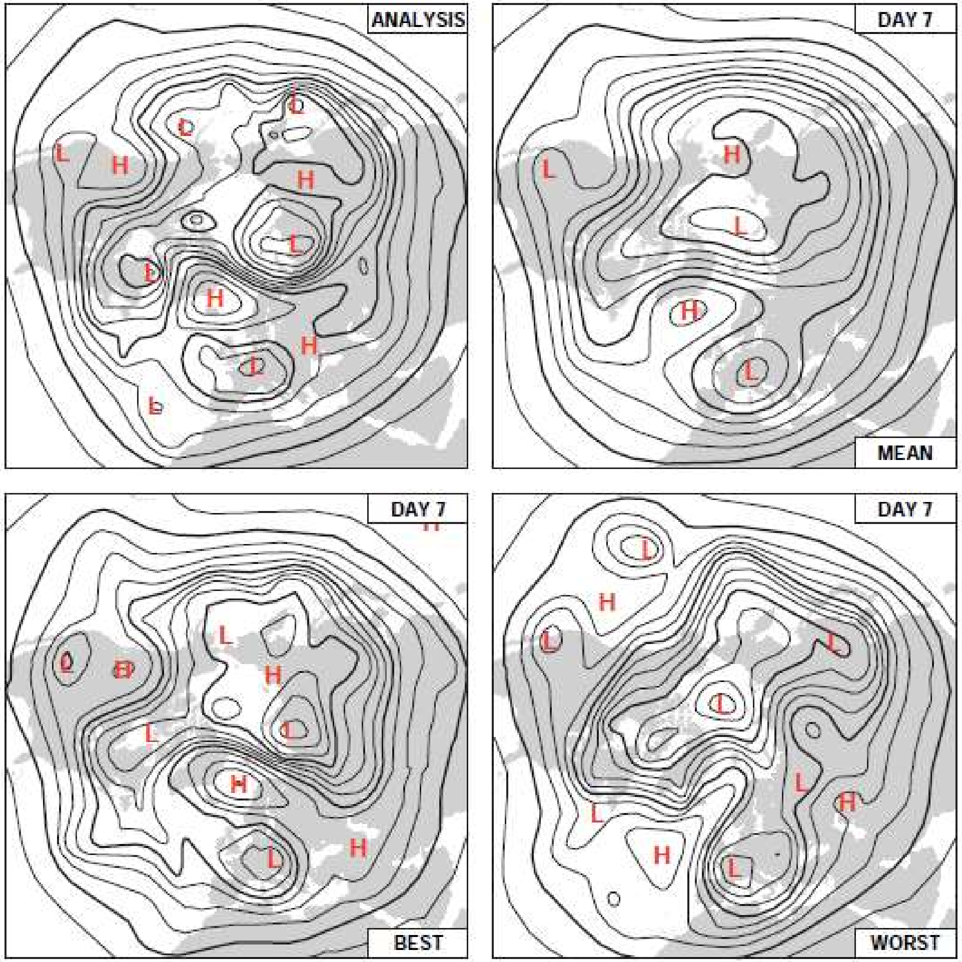
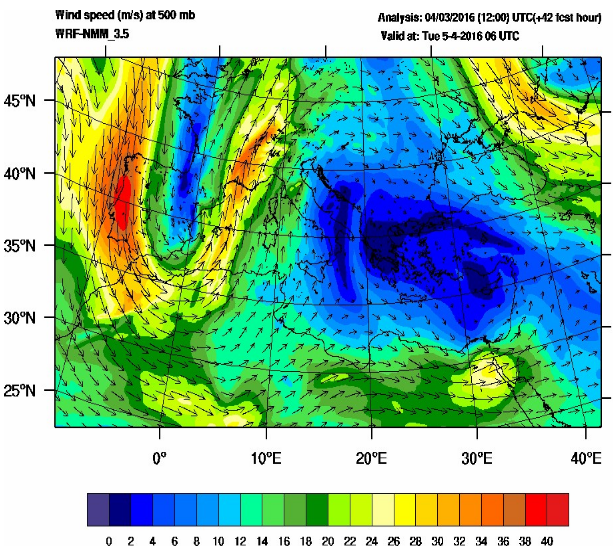
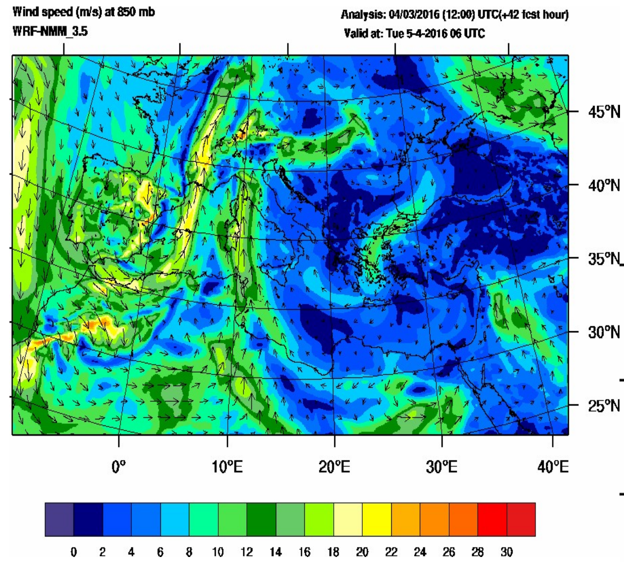
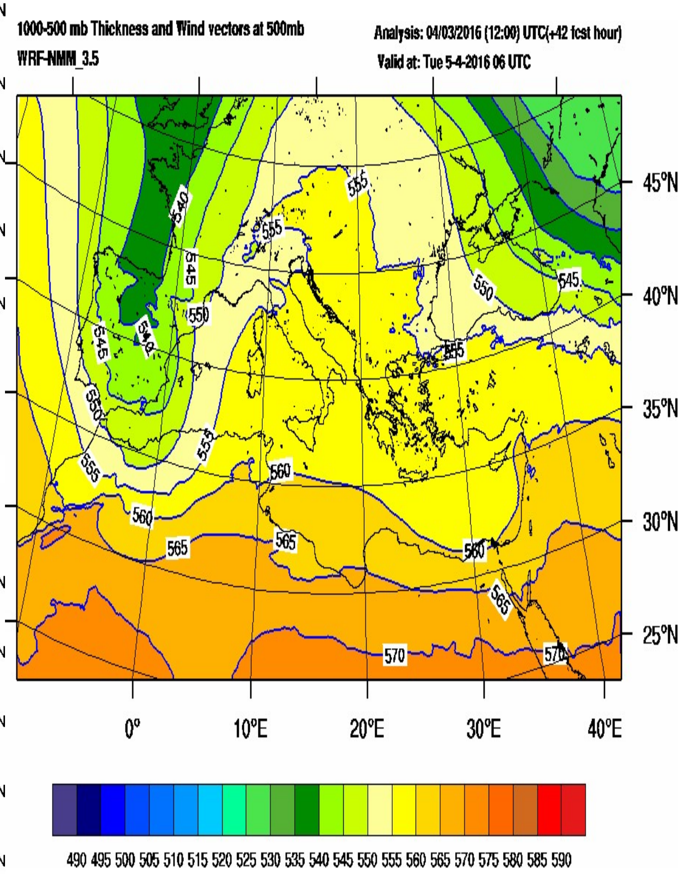


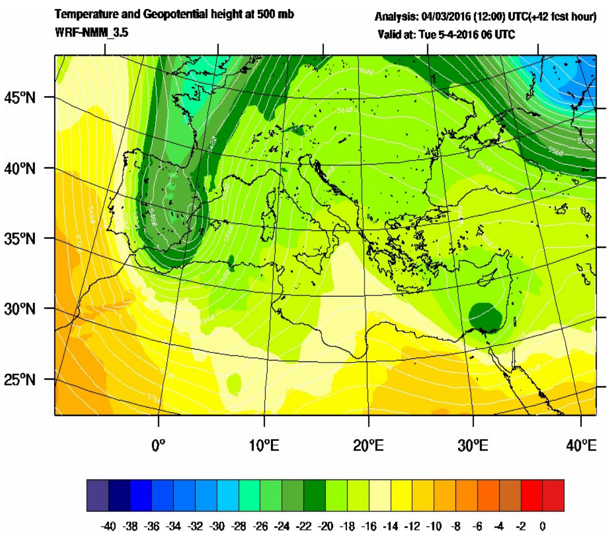
Fig. 7.27 As in Fig. 7.25 but for the 7-day forecasts generated by the ensemble forecasting system in current use at ECMWF. *Mean* is the average of the 50 members of the ensemble *Best* and *Worst* forecasts are selected based on anomaly correlations with the verifying analysis. [Courtesy of Adrian J. Simmons, ECMWF.]

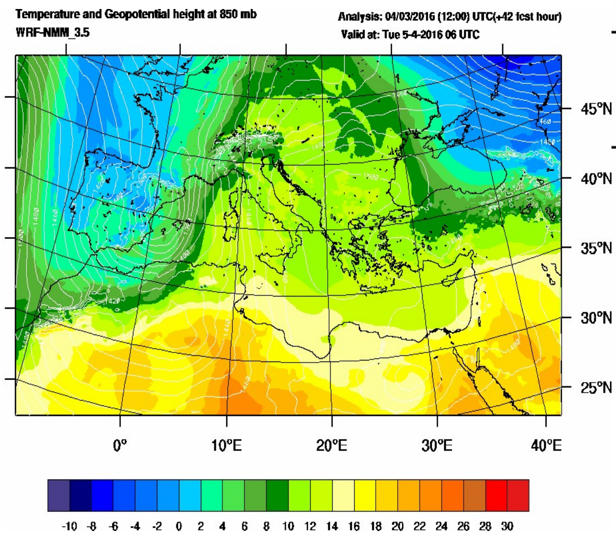
Post processing & Meteorological parameters



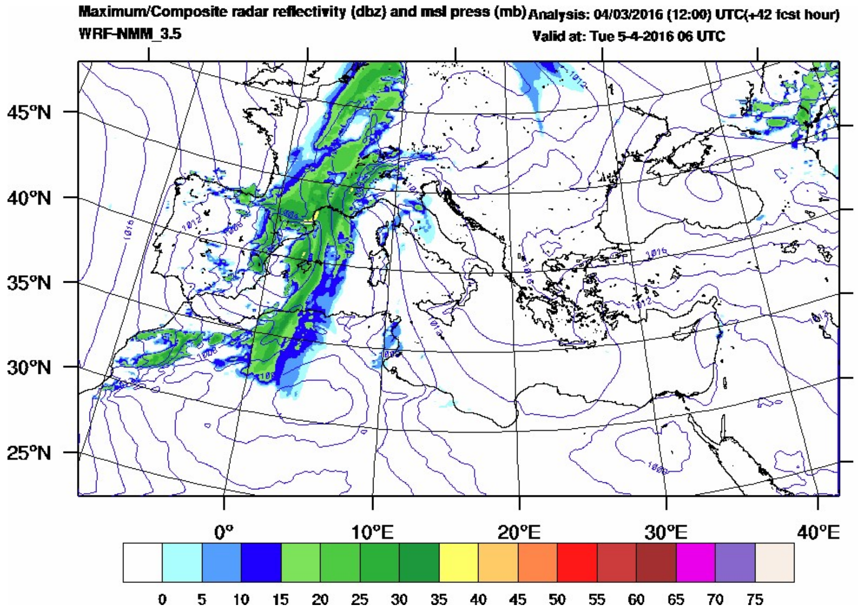


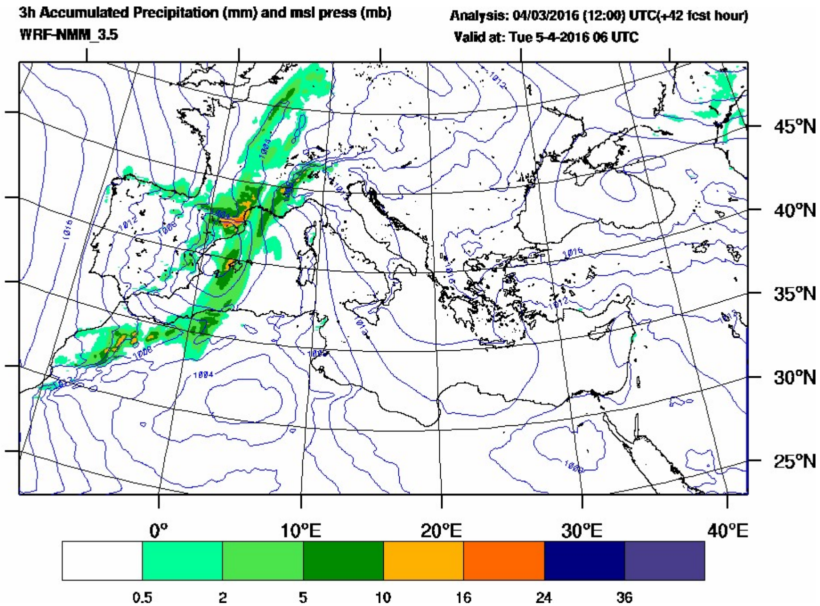


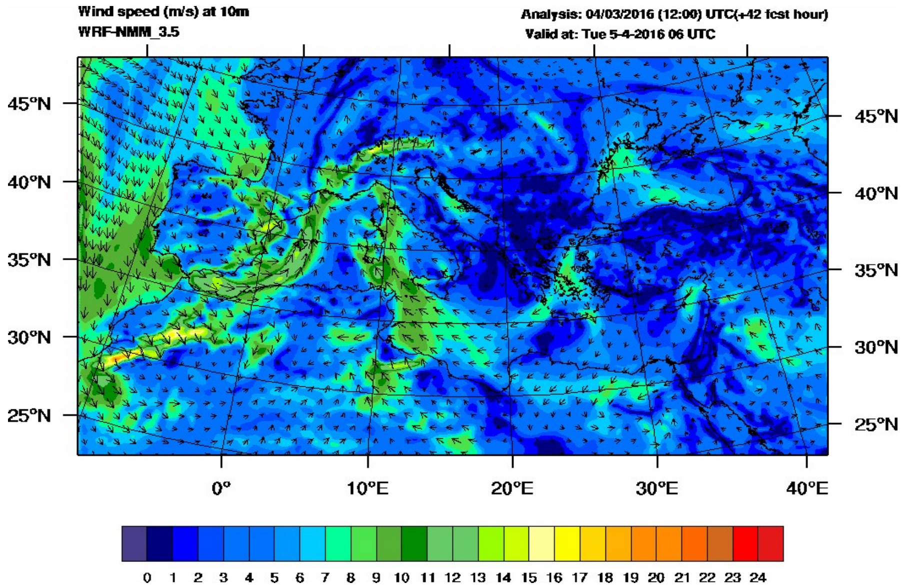


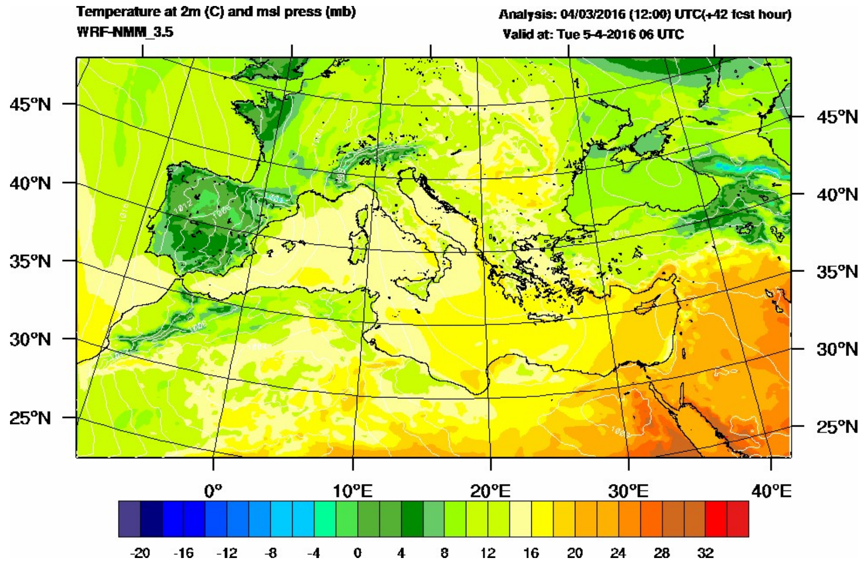


Post processing & Meteorological parameters

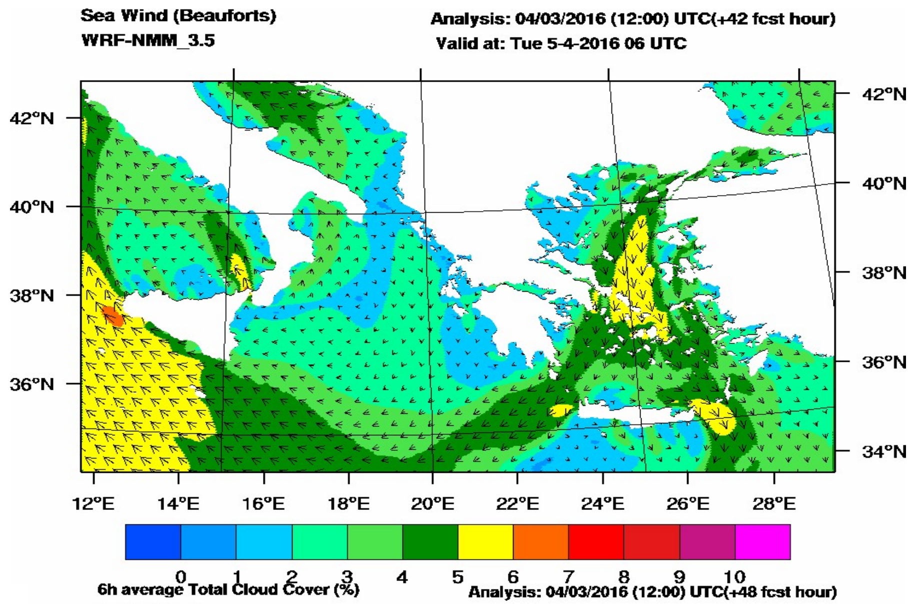


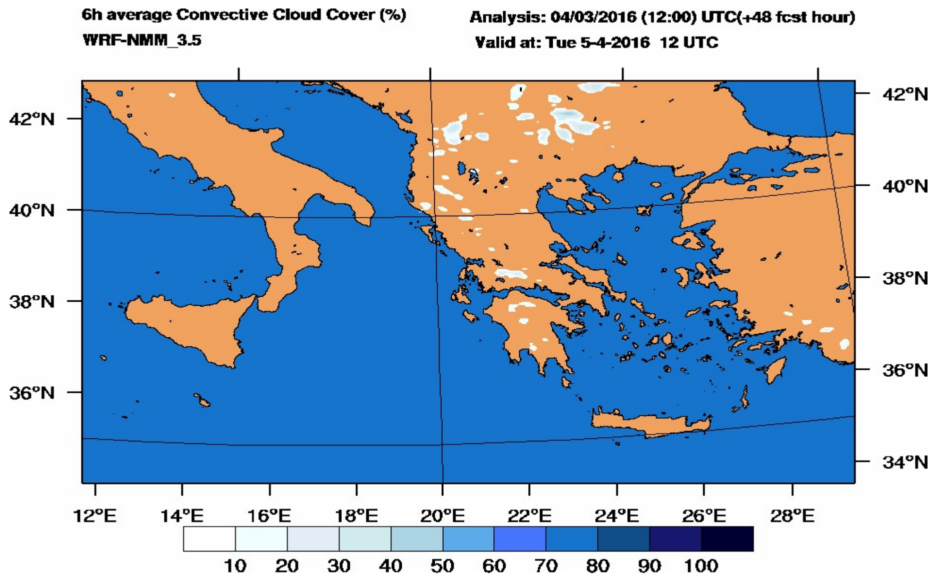


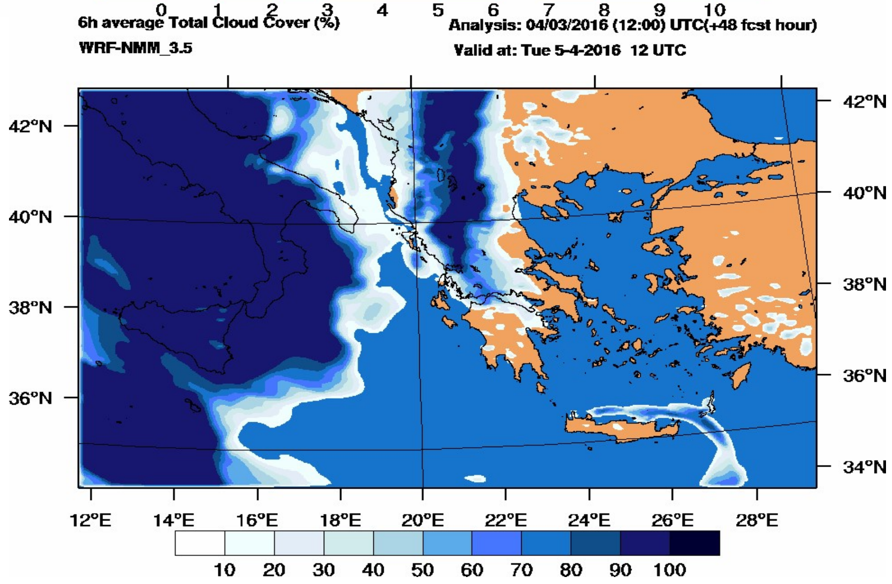


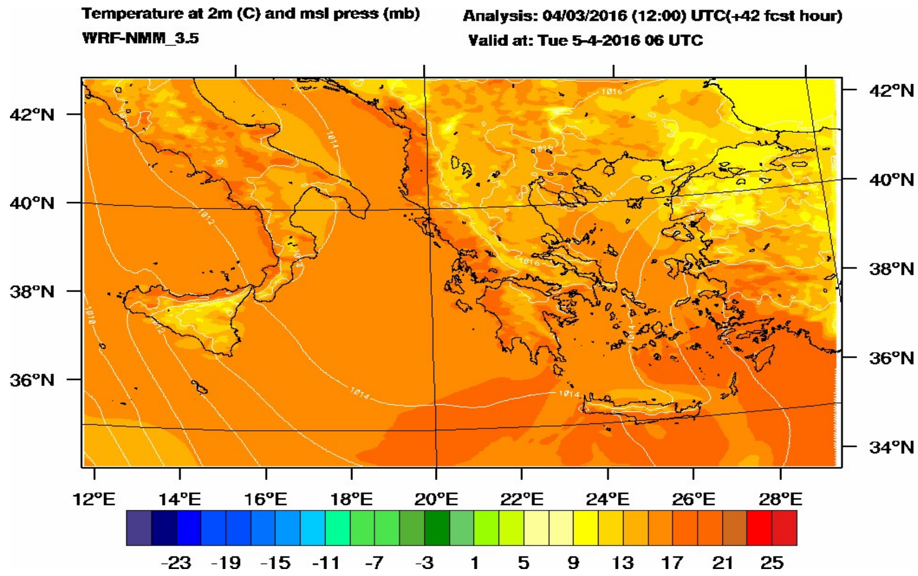


Post processing & Meteorological parameters

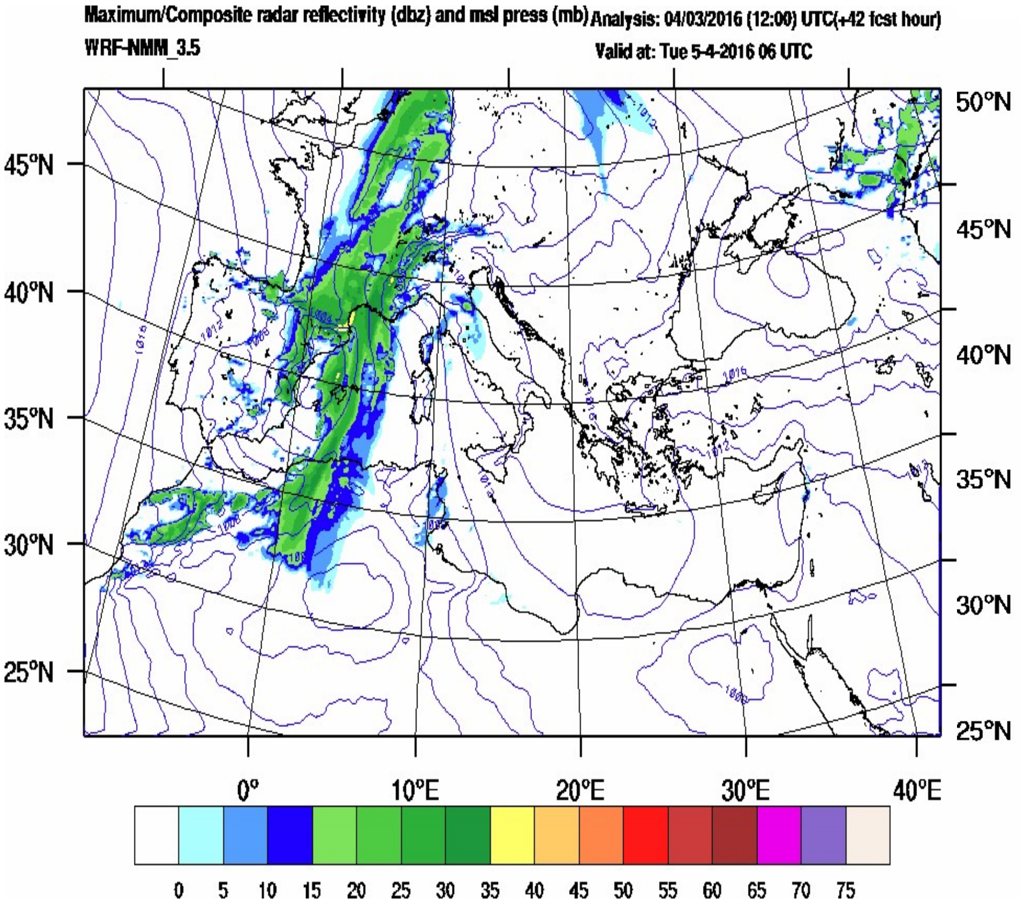




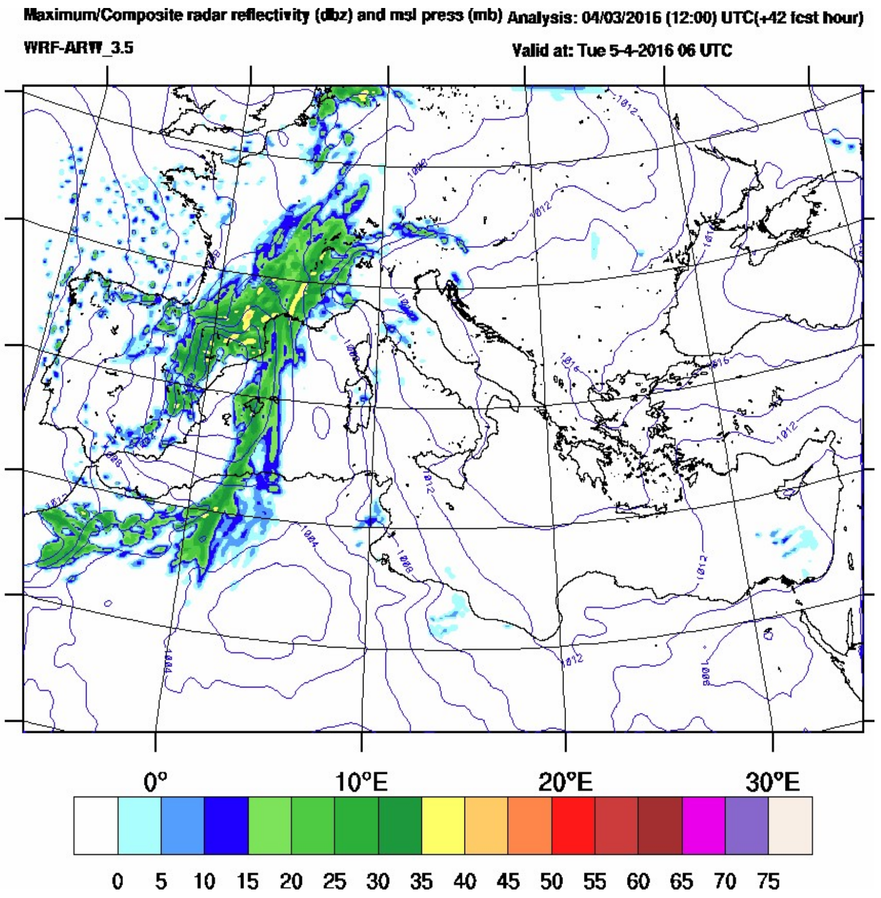




Post processing & Meteorological parameters

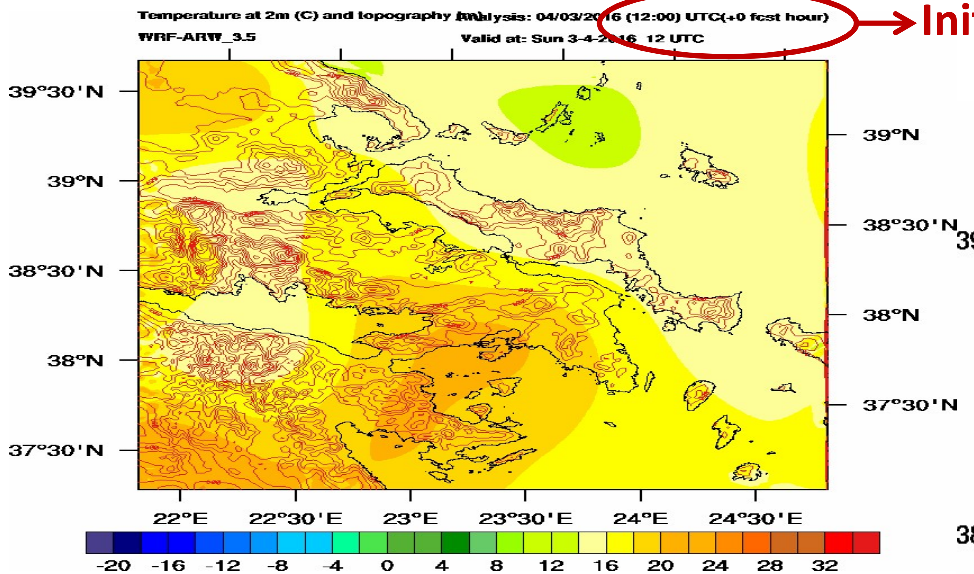


WRF-NMM_3.5

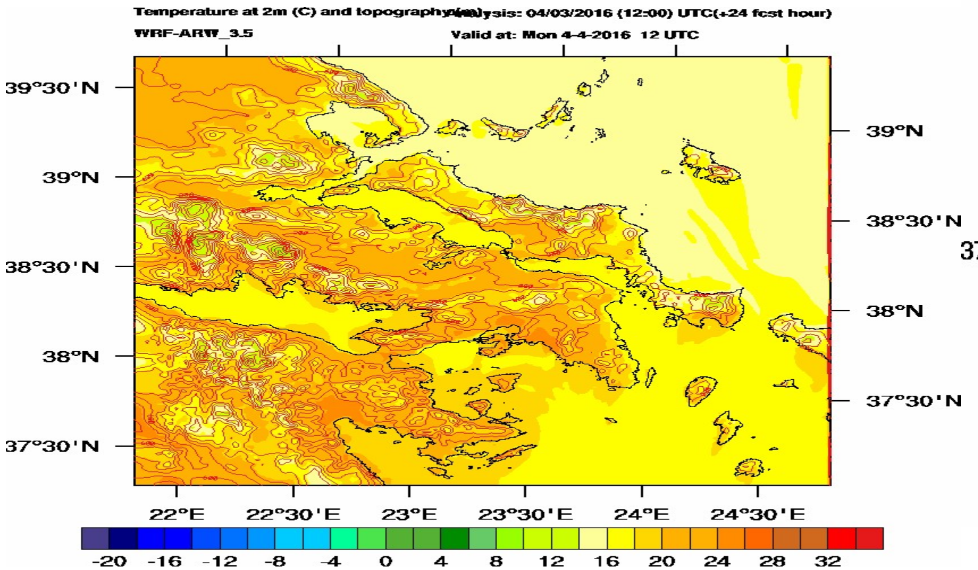


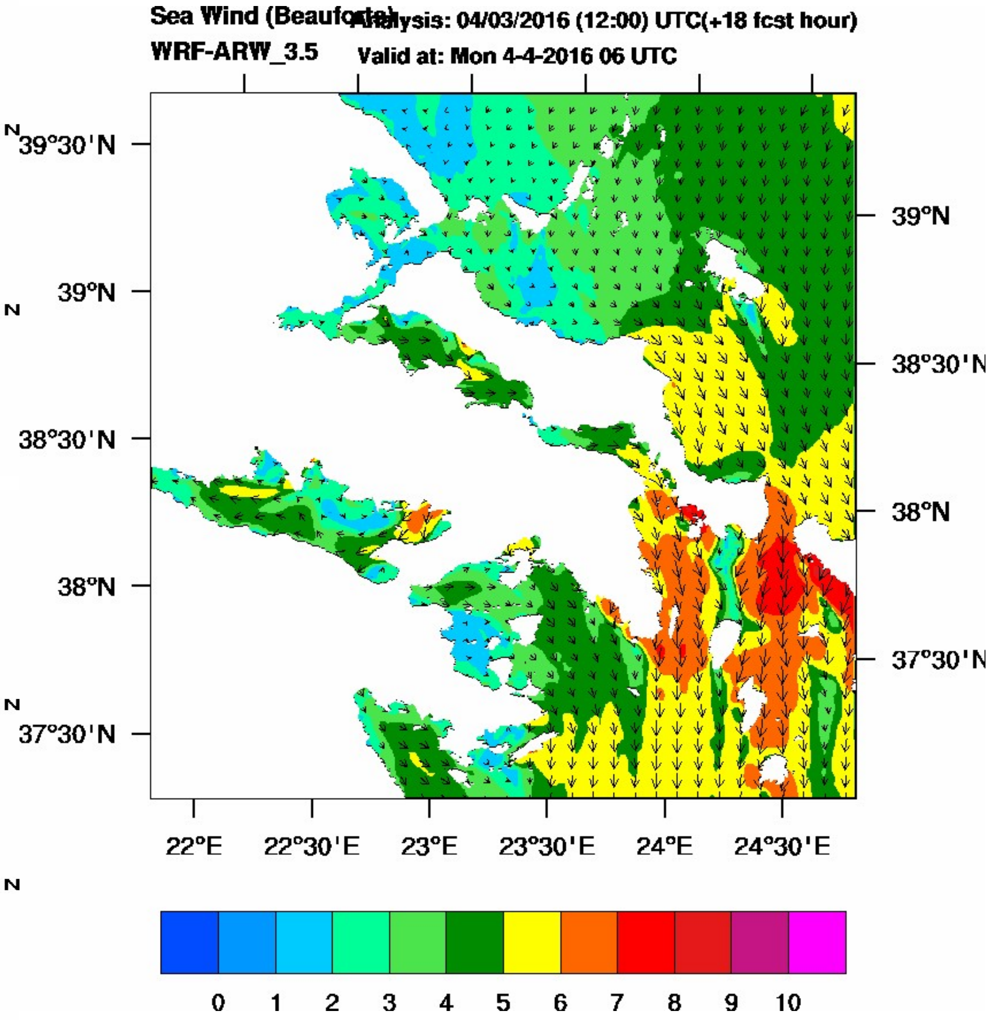
WRF-ARW_3.5

Post processing & Meteorological parameters

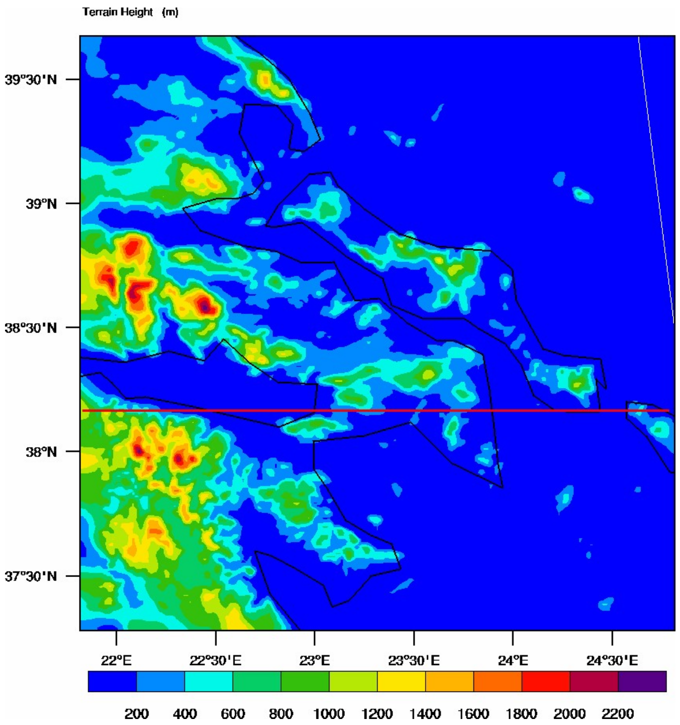


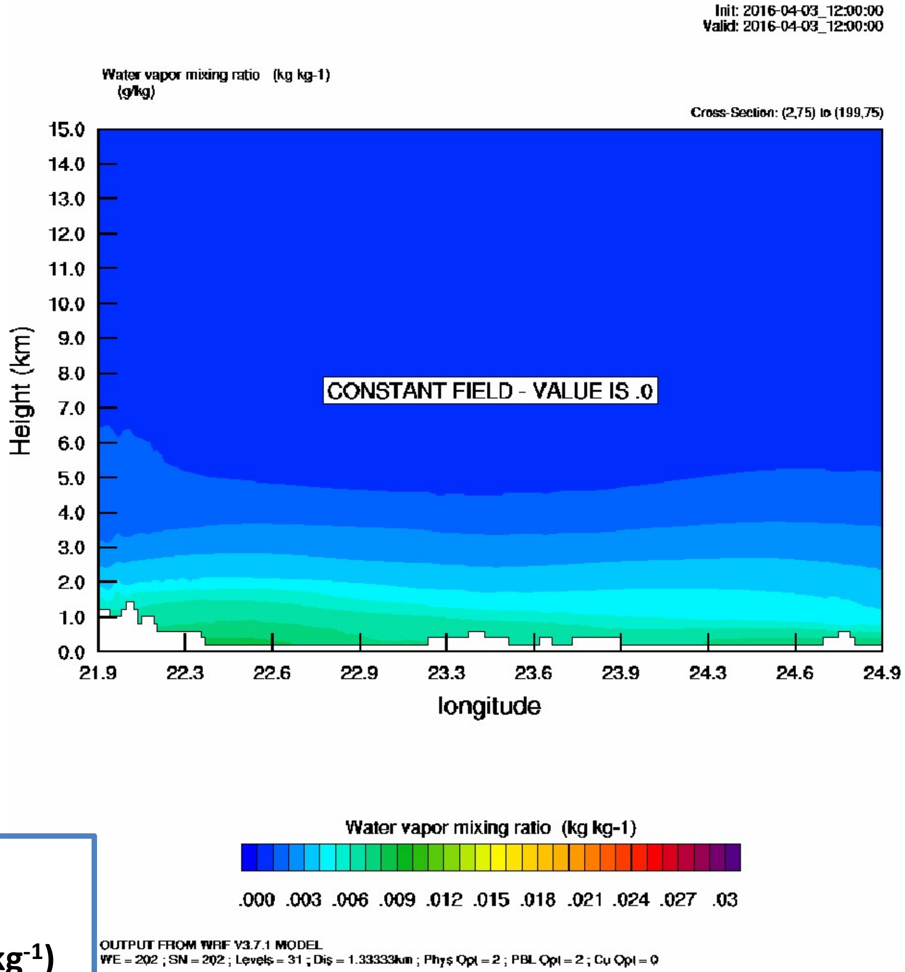
Initial field





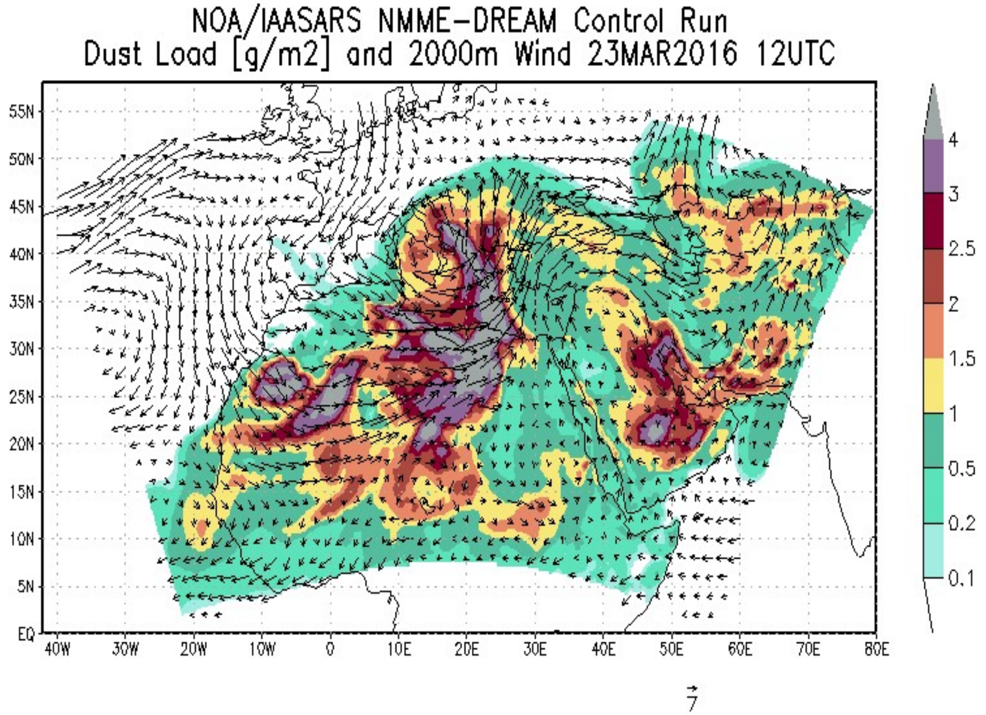
Post processing & Meteorological parameters

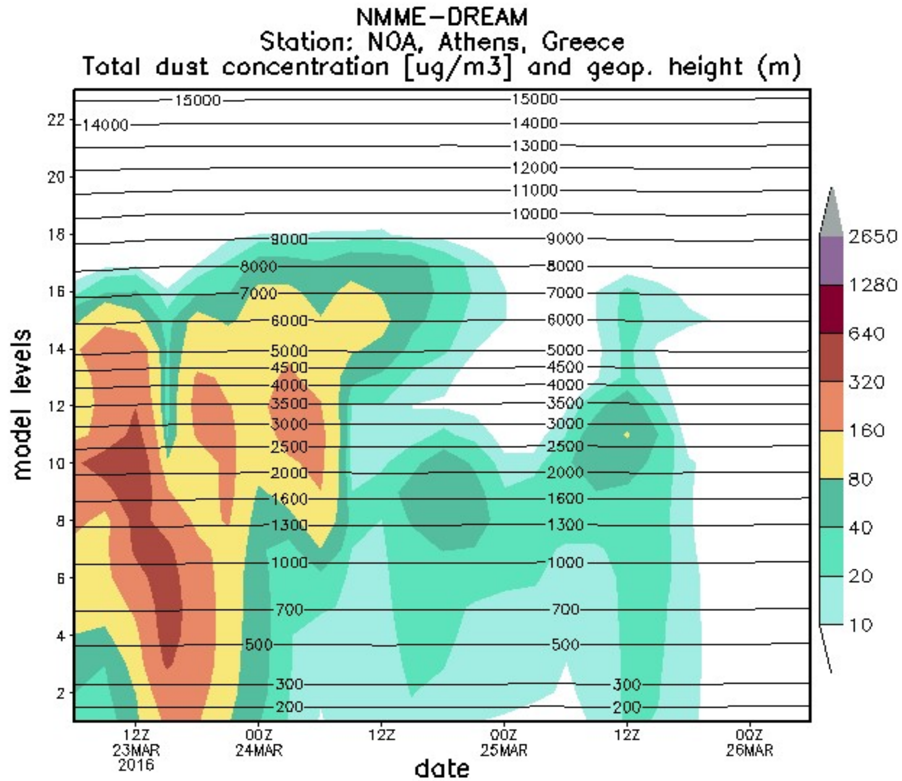




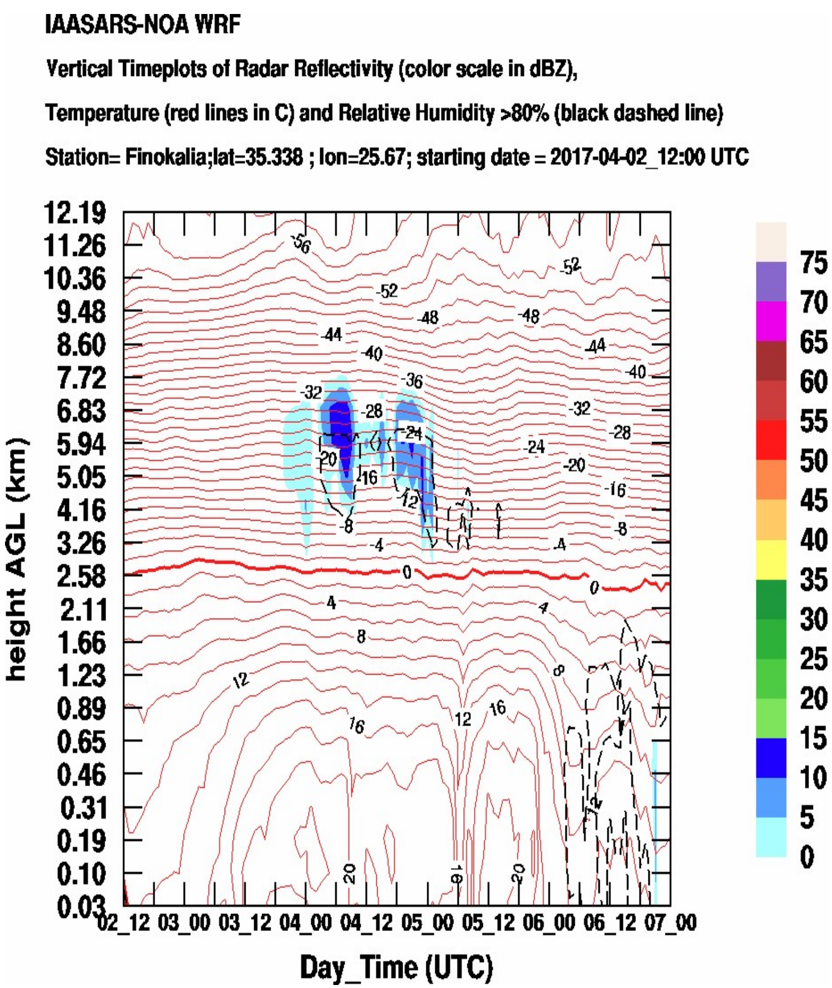
Vertical cross section of
- Water vapor mixing ratio (color scale in kg kg^{-1})
- Liquid condensates mixing ration (red contours in g kg^{-1})
- Ice condensates mixing ration (white contours in g kg^{-1})

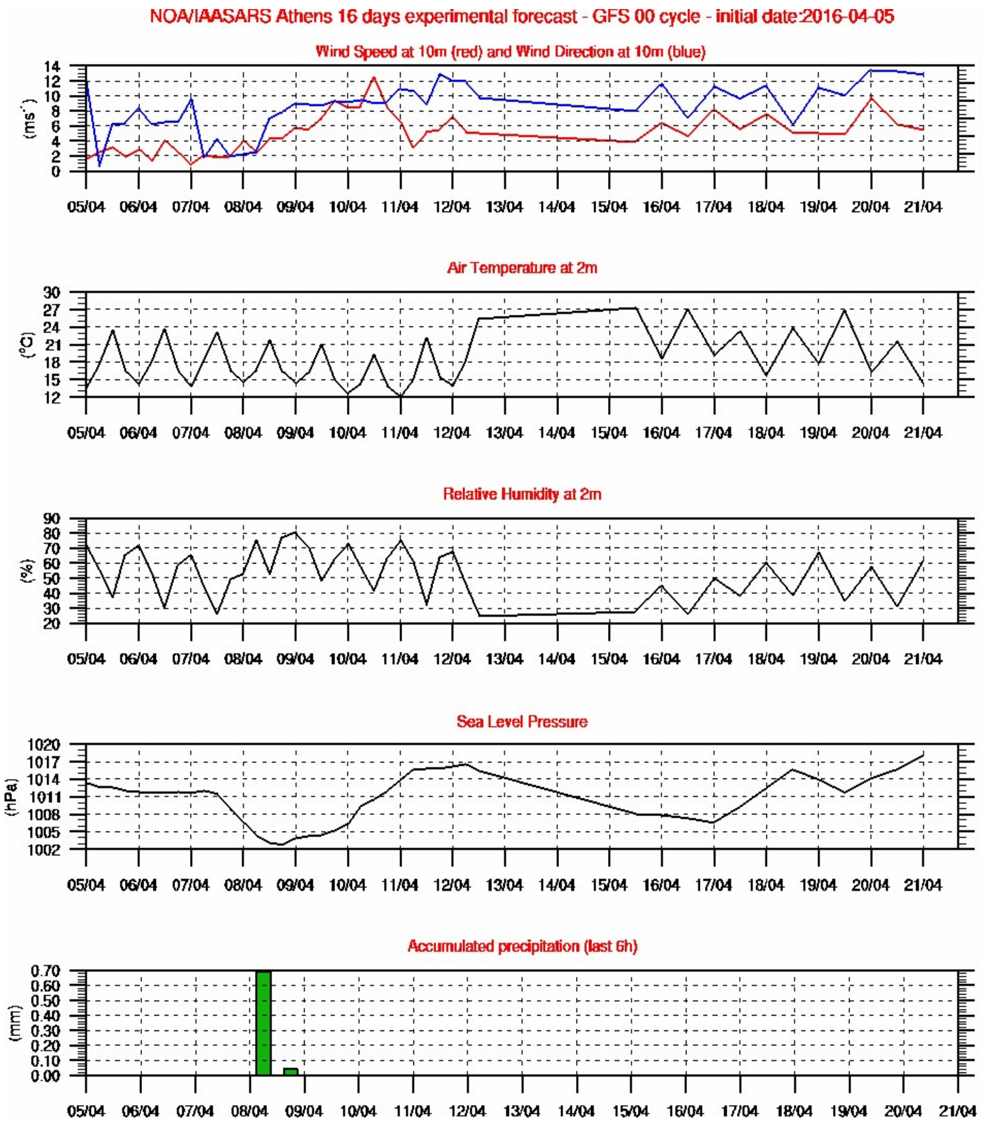
Post processing & Meteorological parameters



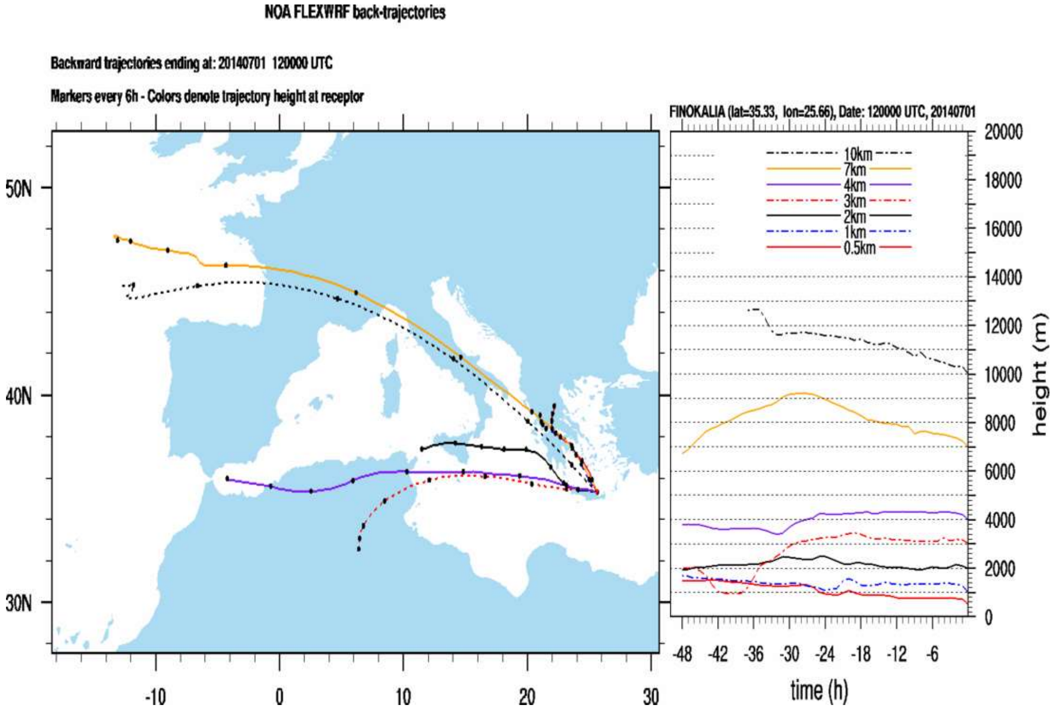


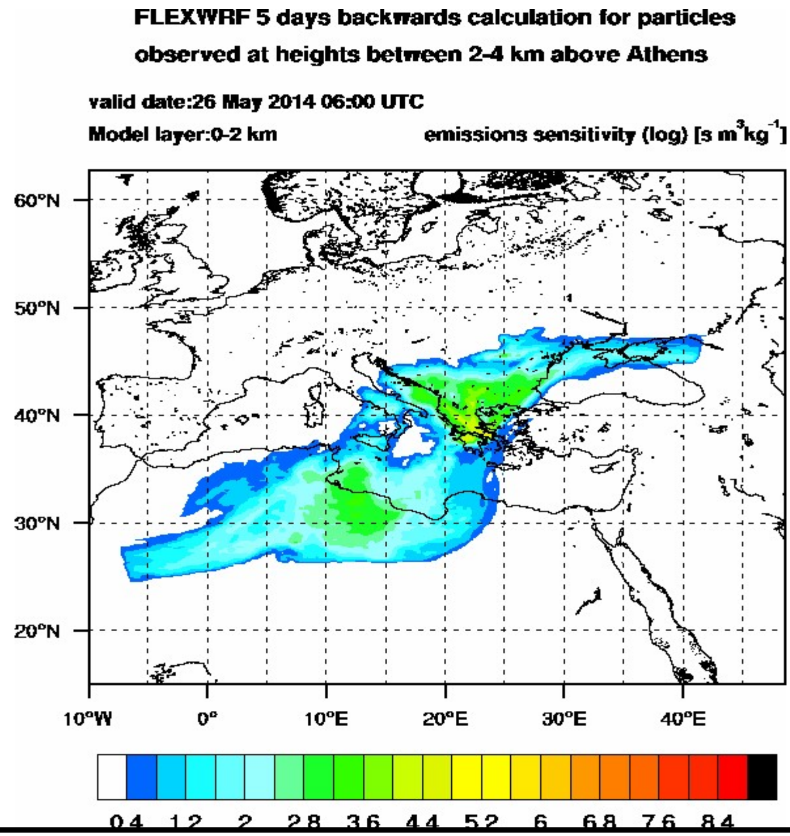
Post processing & Meteorological parameters





Post processing & Meteorological parameters





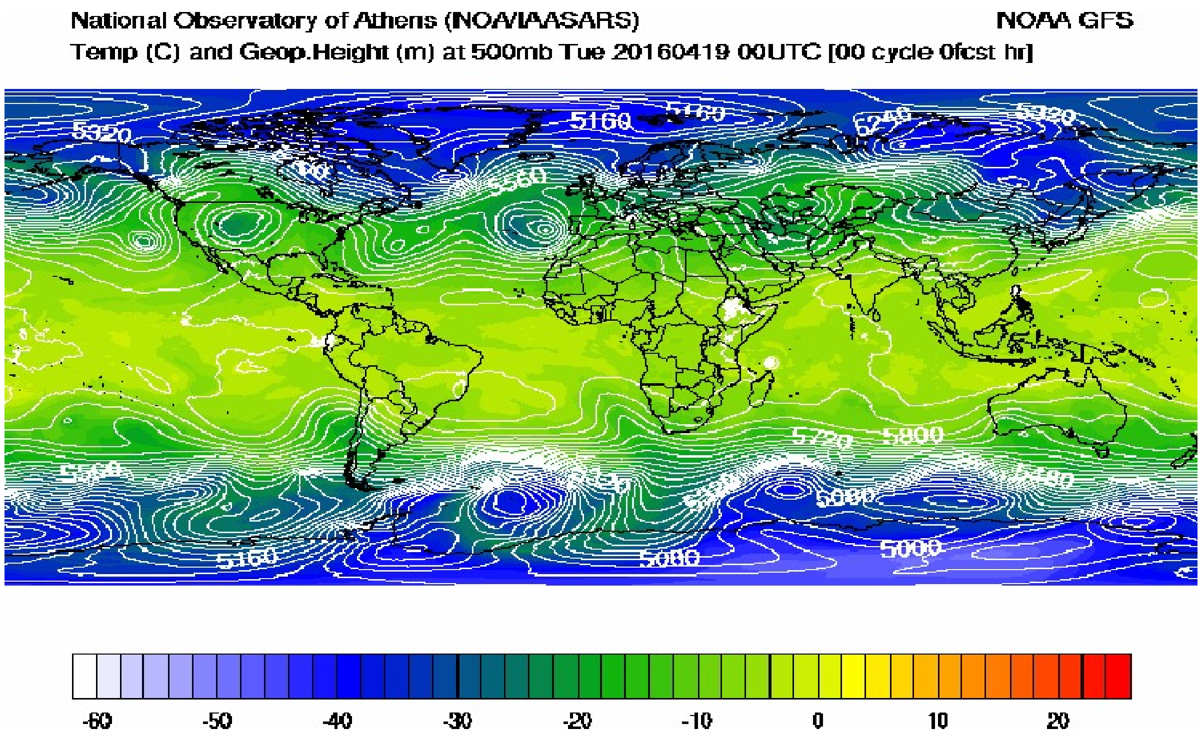
FLEXWRF 48 hours backward-trajectories ending at Finokalia on 01 July 2014, 12:00 UTC. Arrival heights are 0.5 km (solid red), 1 km (dashed blue), 2 km (solid black), 3 km (dashed red), 4 km (magenta), 7 km (yellow) and 10 km (dashed black).

FLEXWRF emissions sensitivity (residence time) calculation for a particle population that was observed at heights between 2 and 4 km above Athens on 26 May 2014, 12:00 UTC. The colored areas indicate particles from that height range that were present at heights below 2 km during the last 5 days thus indicating the possible source areas.

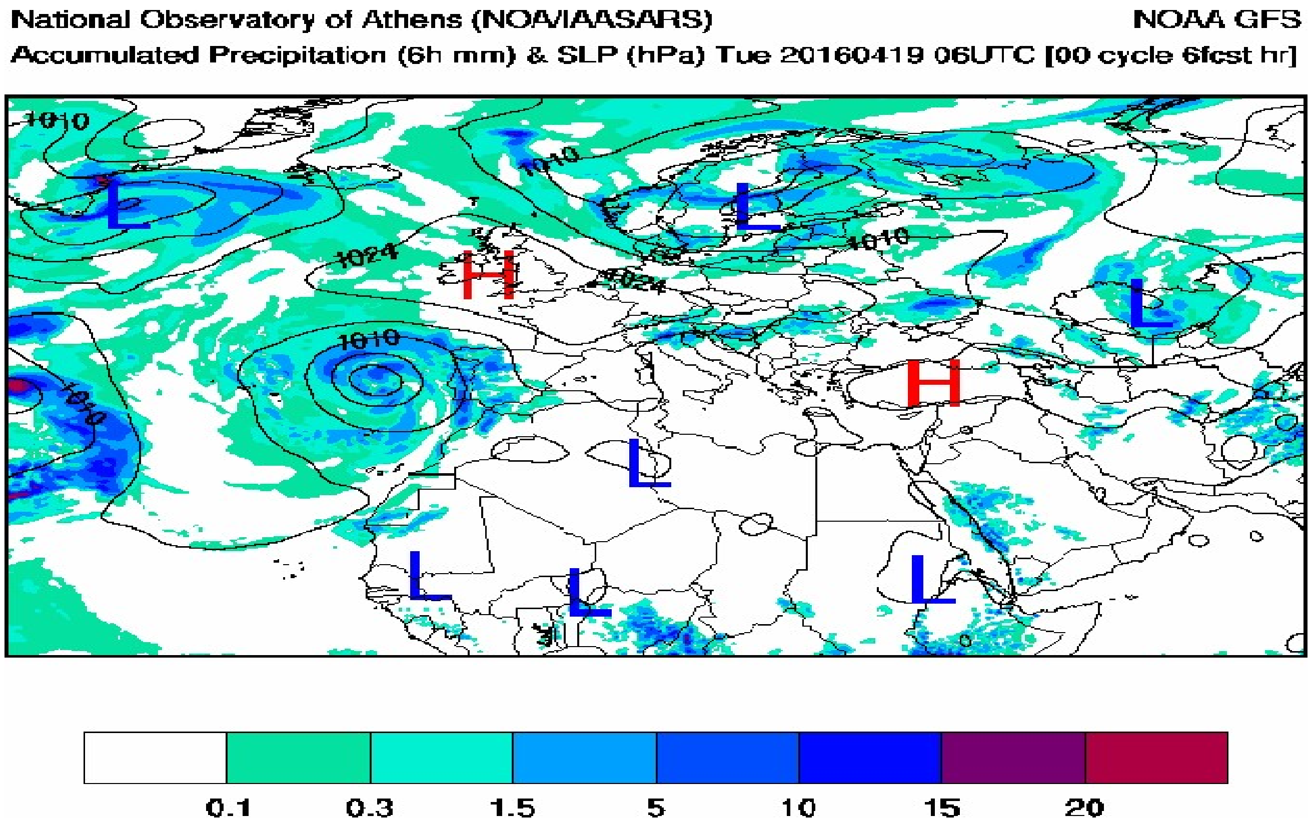
Global Models

- GFS Global Forecast System (previously AVN) – developed by NOAA
- IFS developed by the European Centre for Medium-Range Weather Forecasts
- UM Unified Model developed by the UK Met Office
- GME developed by the German Weather Service, DWD
- ARPEGE developed by the French Weather Service, Météo-France

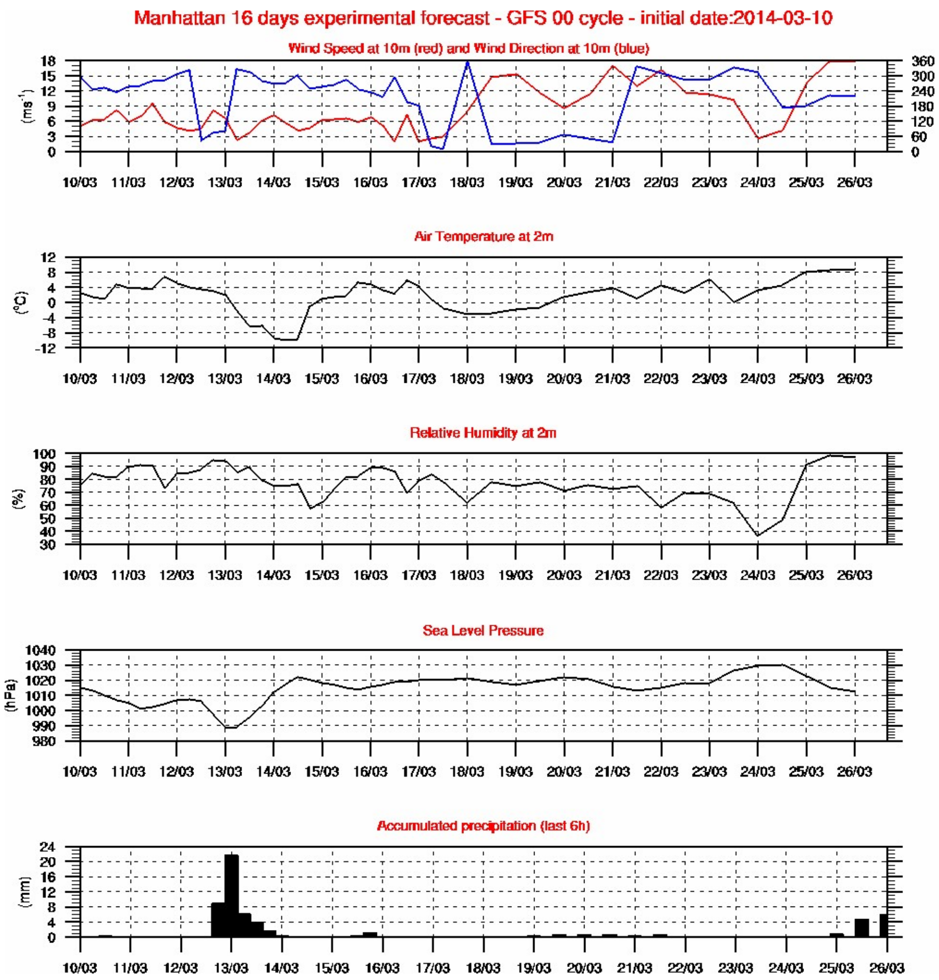
Global Models – GFS 0.25

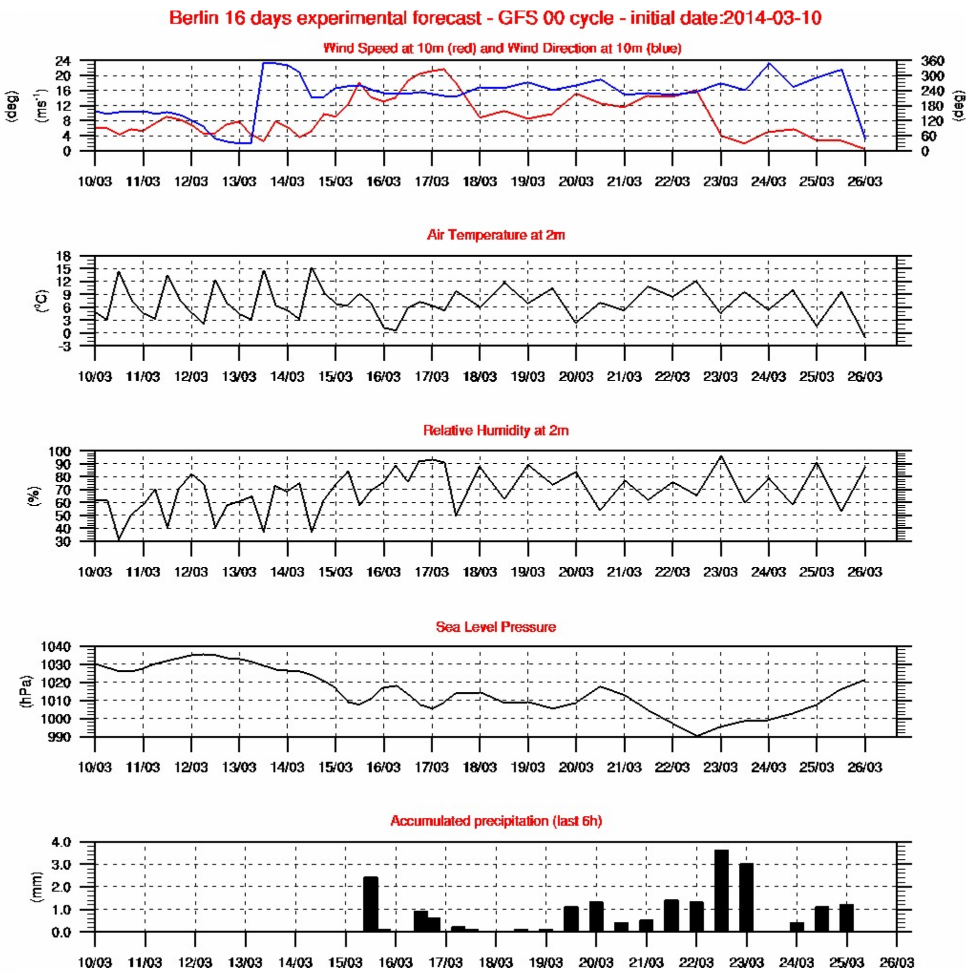


Global Models / GFS 0.25 zoom



Global Models / GFS 0.25 zoom

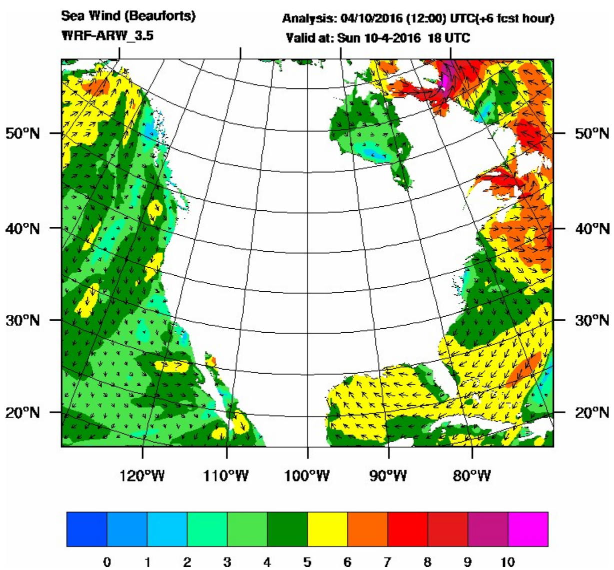


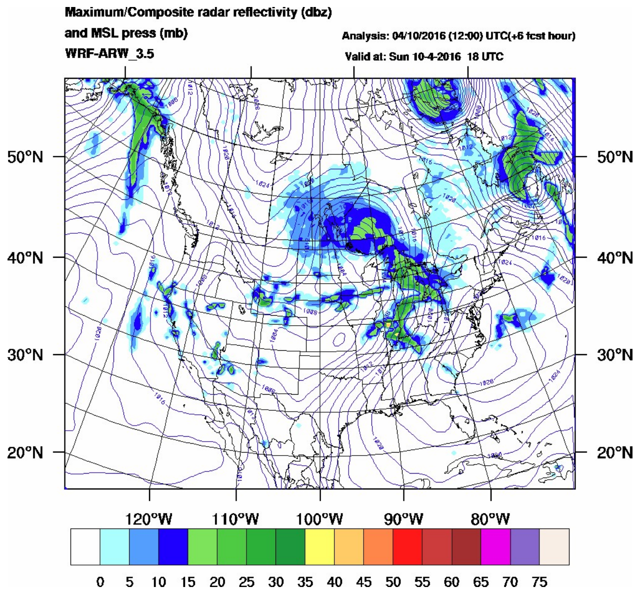


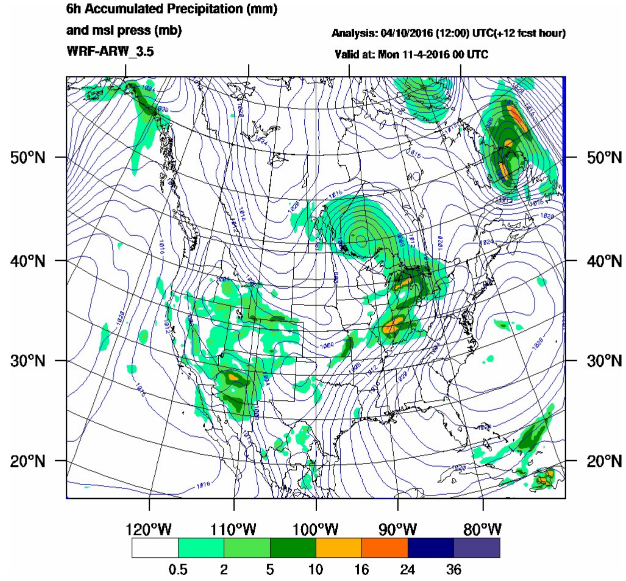
Regional Models

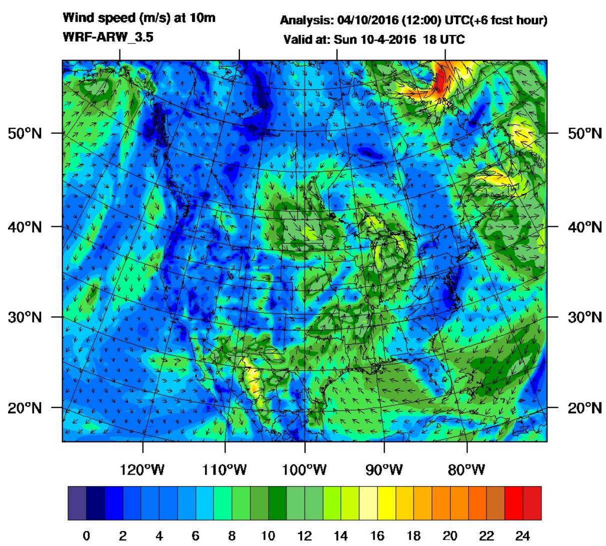
- **WRF** The Weather Research and Forecasting model was developed cooperatively by NCEP, NCAR, and the meteorological research community. WRF has several configurations, including:
  - **WRF-NMM** is the primary short-term weather forecast model for the U.S., replacing the Eta model. Beginning in May 2006, NCEP began to use the WRF-NMM as the operational NAM.
  - **WRF-ARW** Advanced Research WRF developed primarily at the U.S. National Center for Atmospheric Research (NCAR)
- **RAMS** the Regional Atmospheric Modeling System developed at Colorado State University for numerical simulations of atmospheric meteorology and other environmental phenomena on scales from meters to hundreds of kilometers - now supported in the public domain
- **MM5** The Fifth Generation Penn State/NCAR Mesoscale Model
- **ALADIN** The high-resolution limited-area hydrostatic and non-hydrostatic model developed and operated by several European and North African countries under the leadership of Météo-France
- **COSMO** The COSMO Model, formerly known as LM, aLMo or LAMI, is a limited-area non-hydrostatic model developed within the framework of the Consortium for Small-Scale Modelling (Germany, Switzerland, Italy, Greece, Poland, Romania, and Russia).

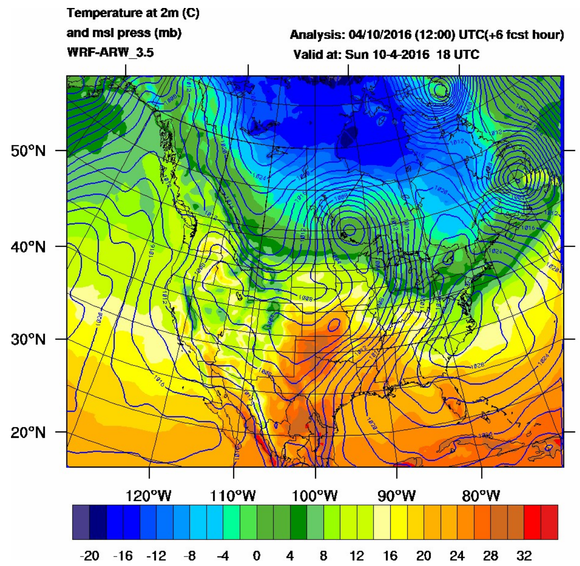
Post Processing Maps

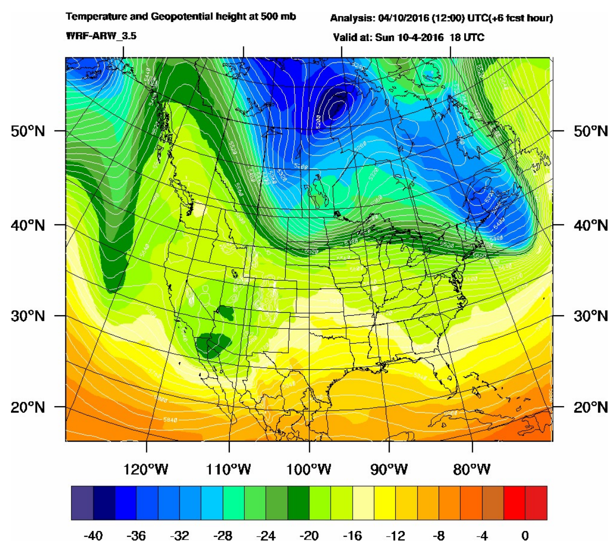




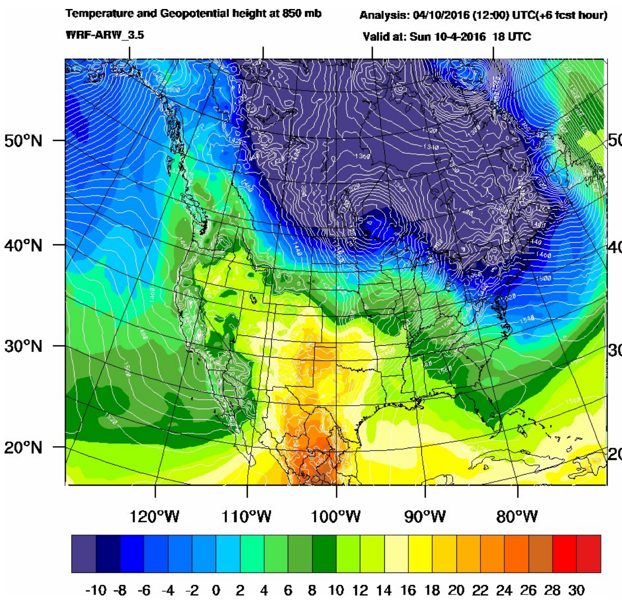


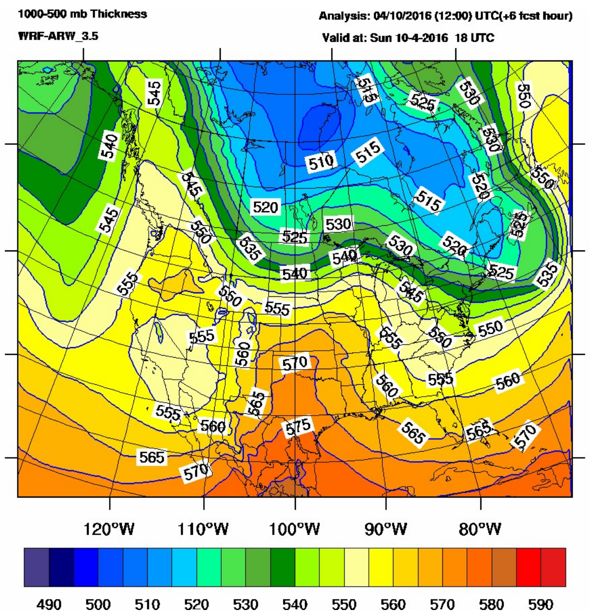


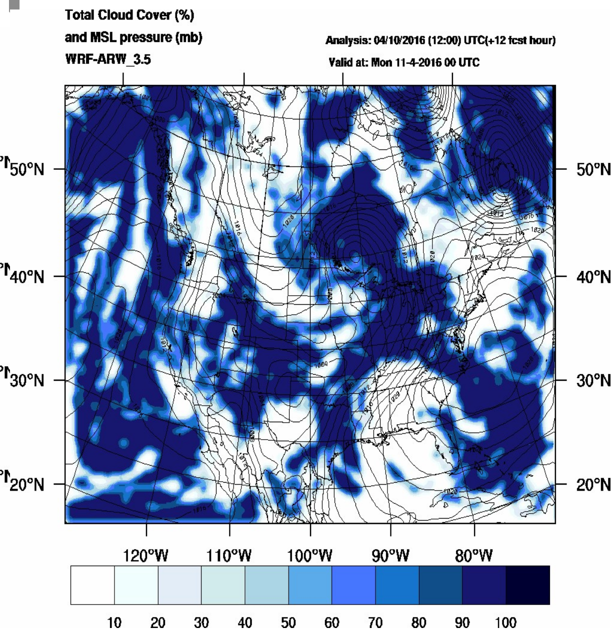


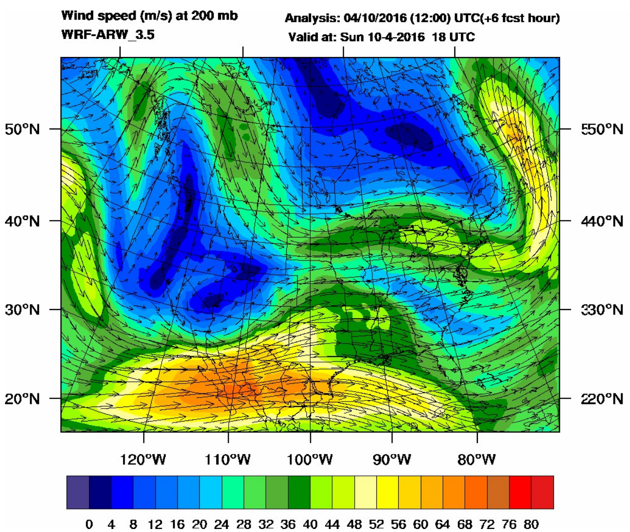


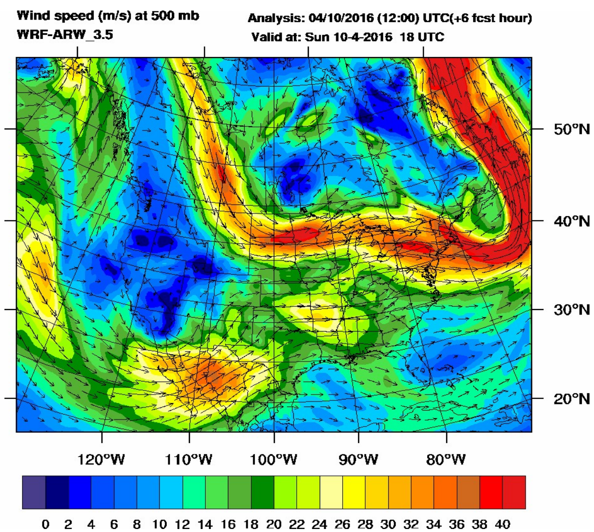
Post Processing Maps

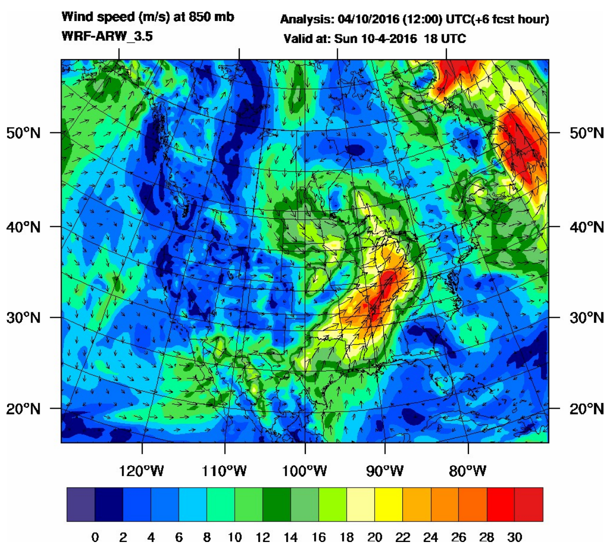




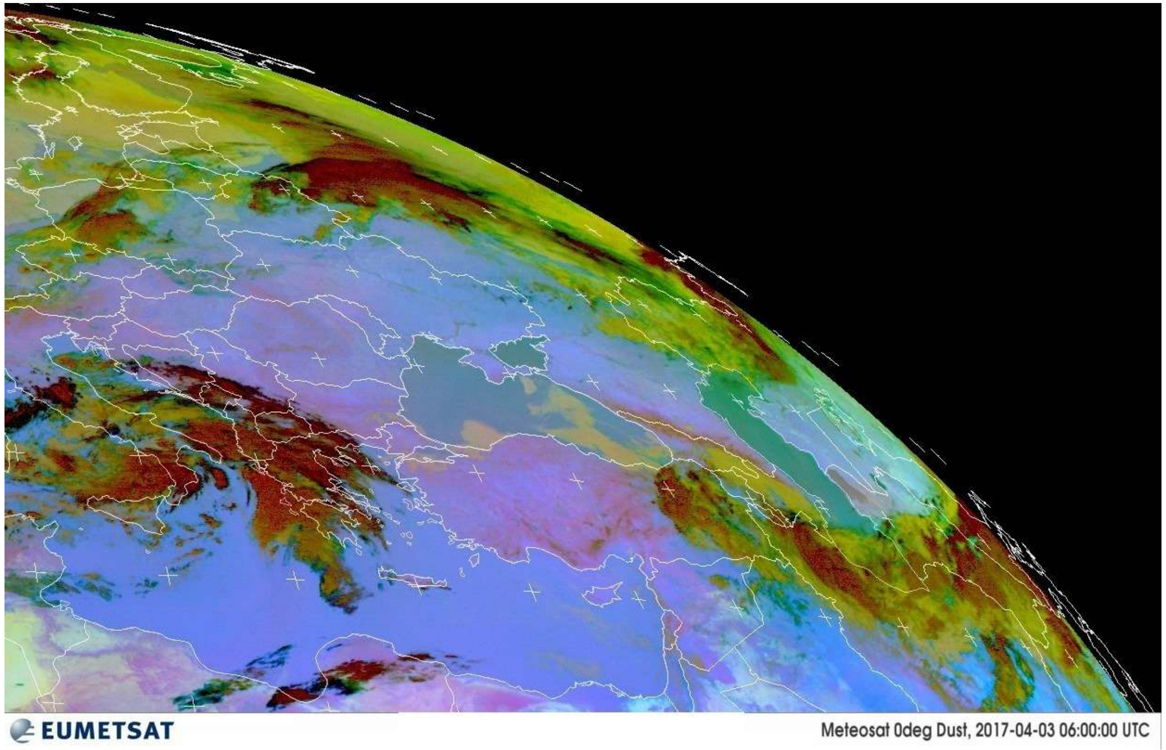








Comparisons with EUMETSAT MSG



http://oiswww.eumetsat.org/IPPS/html/MSG/RGB/DUST/

Comparisons with EUMETSAT MSG and CALIPSO/CALIOP

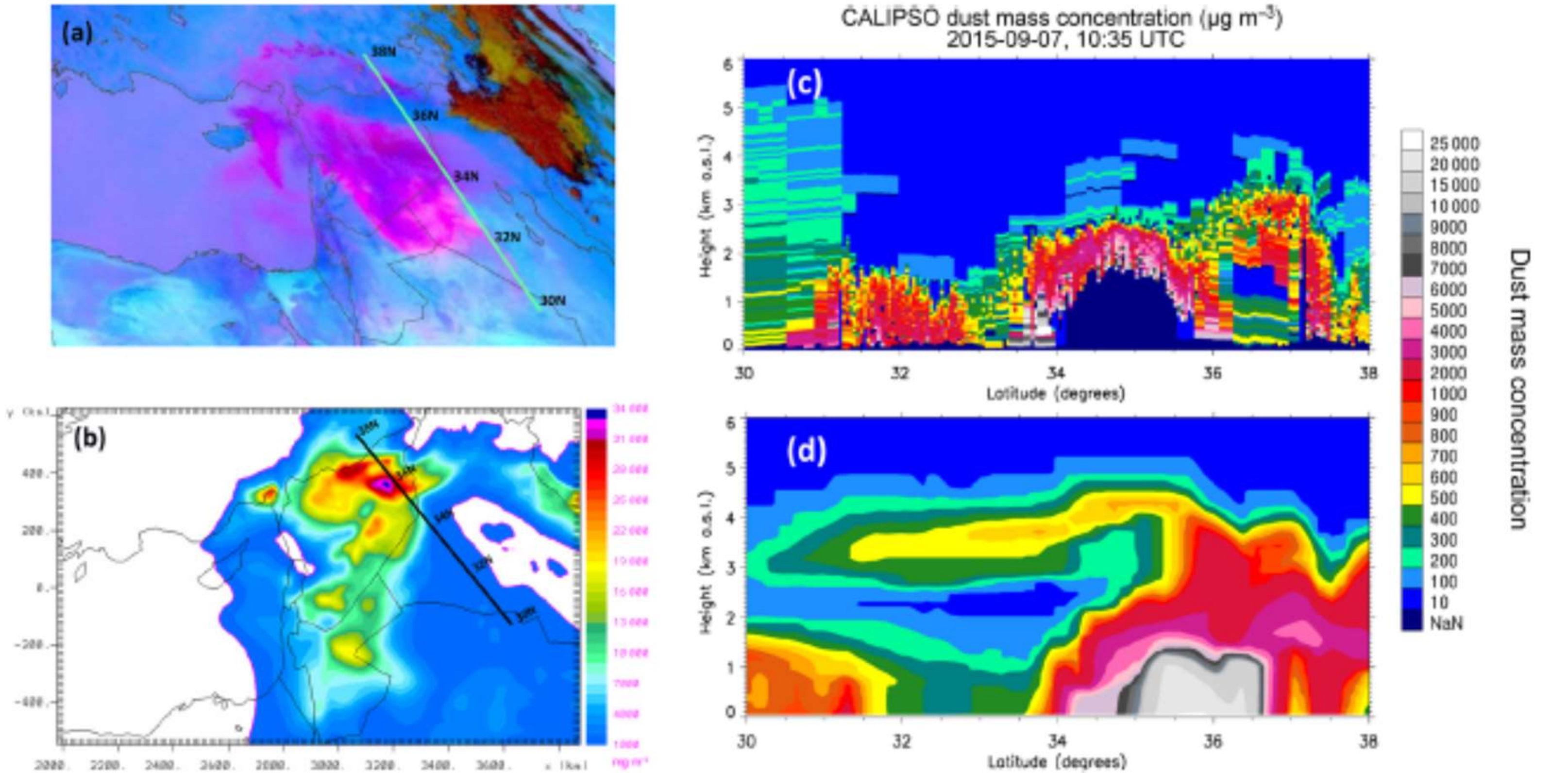


Figure 10. (a) MSG-SEVIRI RGB map and CALIPSO overflight (green line); (b) model dust load ($mg\,m^{-2}$); (c) CALIPSO dust mass concentration ($\mu g\,m^{-3}$); and (d) model dust mass concentration on 7 September 2015, 10:35 UTC. Due to the severity of the event CALIPSO signal is totally attenuated below ~ 1 km a.s.l. in the area between 34 and 36° N (dark blue colour).

Solomos et al., ACP 2017

Comparisons with ground LIDARS and LIRIC retrievals

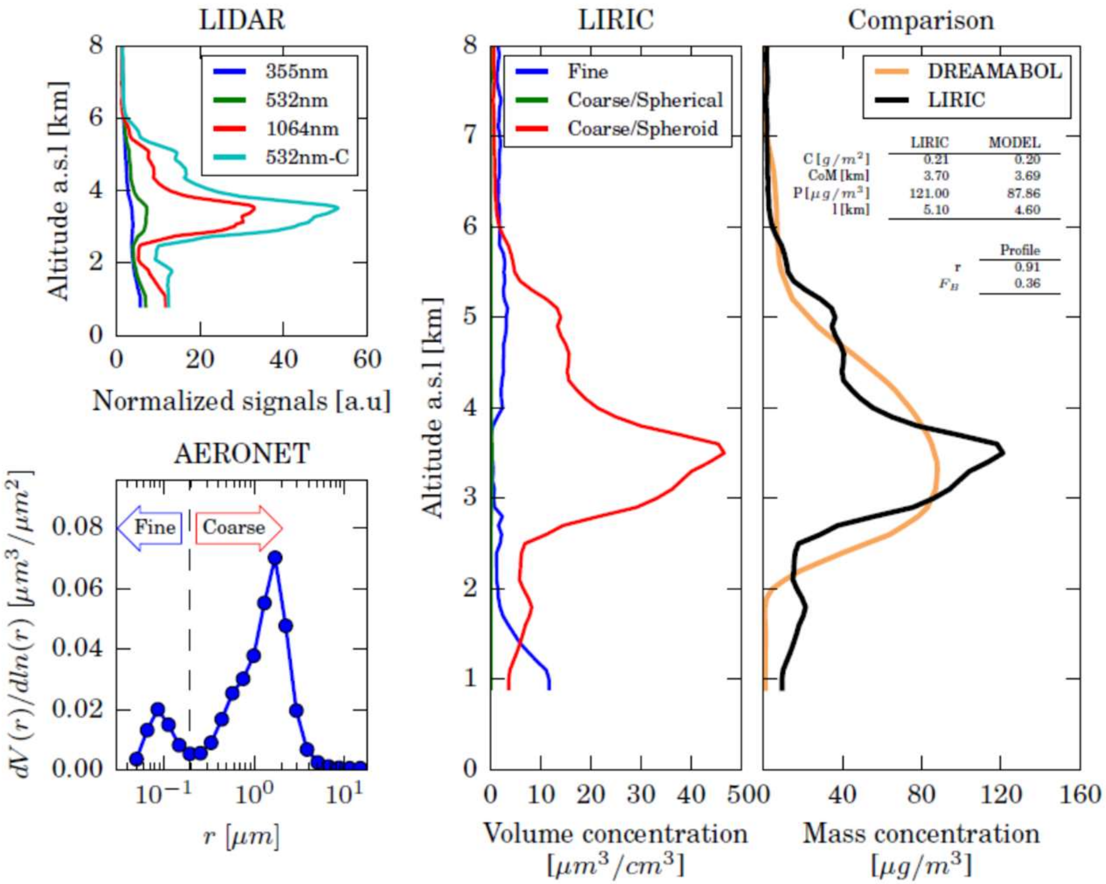
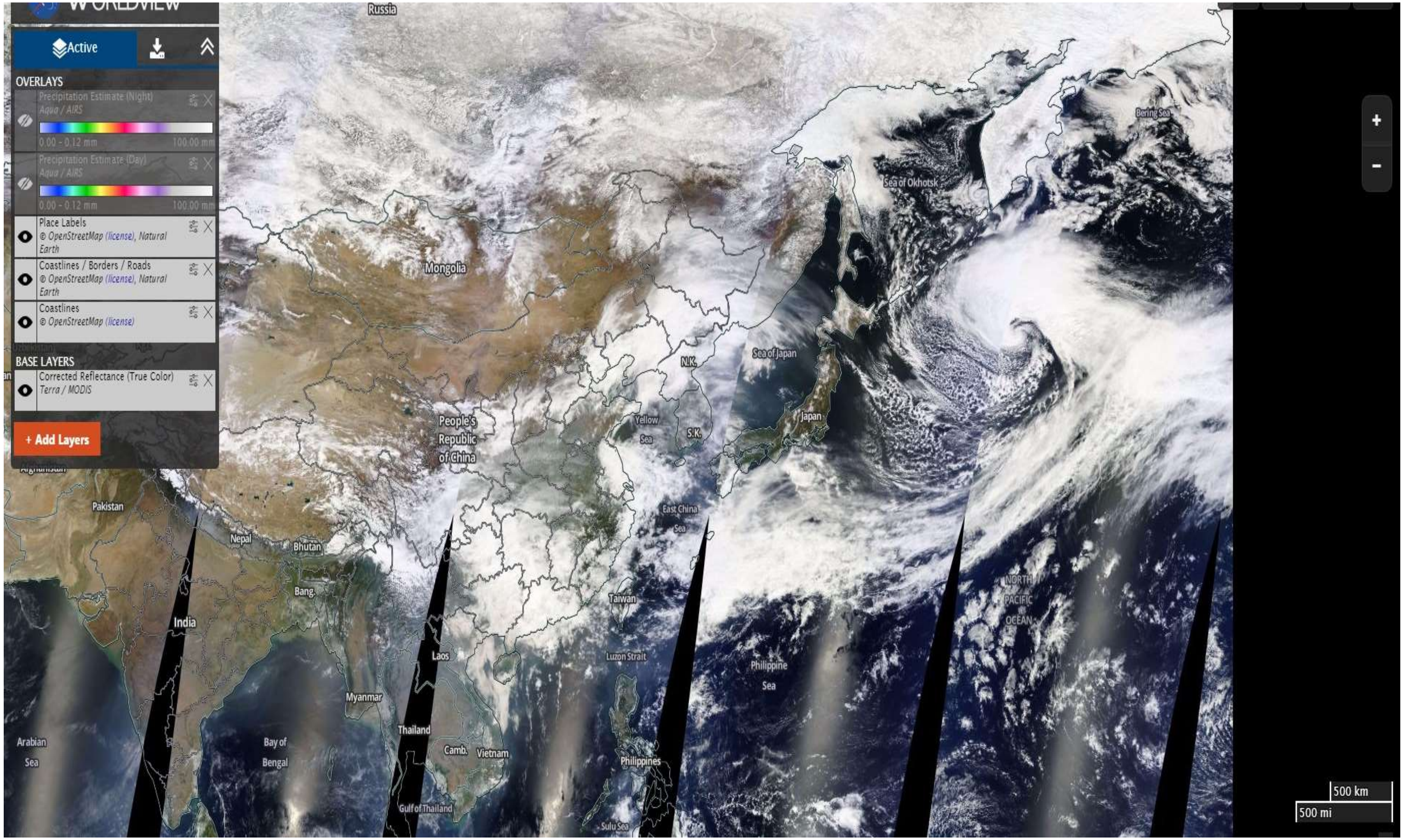


Figure 1. A sketch of the data processing procedure. Data are from Potenza, Italy (40.60° E, 15.72° N) at 11 April 2011. Left plots: LIRIC input i.e., normalized lidar signals (top) and AERONET microphysical inversion (bottom). The vertical line indicates the split between fine and coarse mode. Center plot: volume concentration profiles retrieved by LIRIC. Coarse spherical mode is near zero for all altitudes. Right plot: comparison of the mass concentration profile from LIRIC and DREAMABOL. The embedded tables give the point and profile statistics.

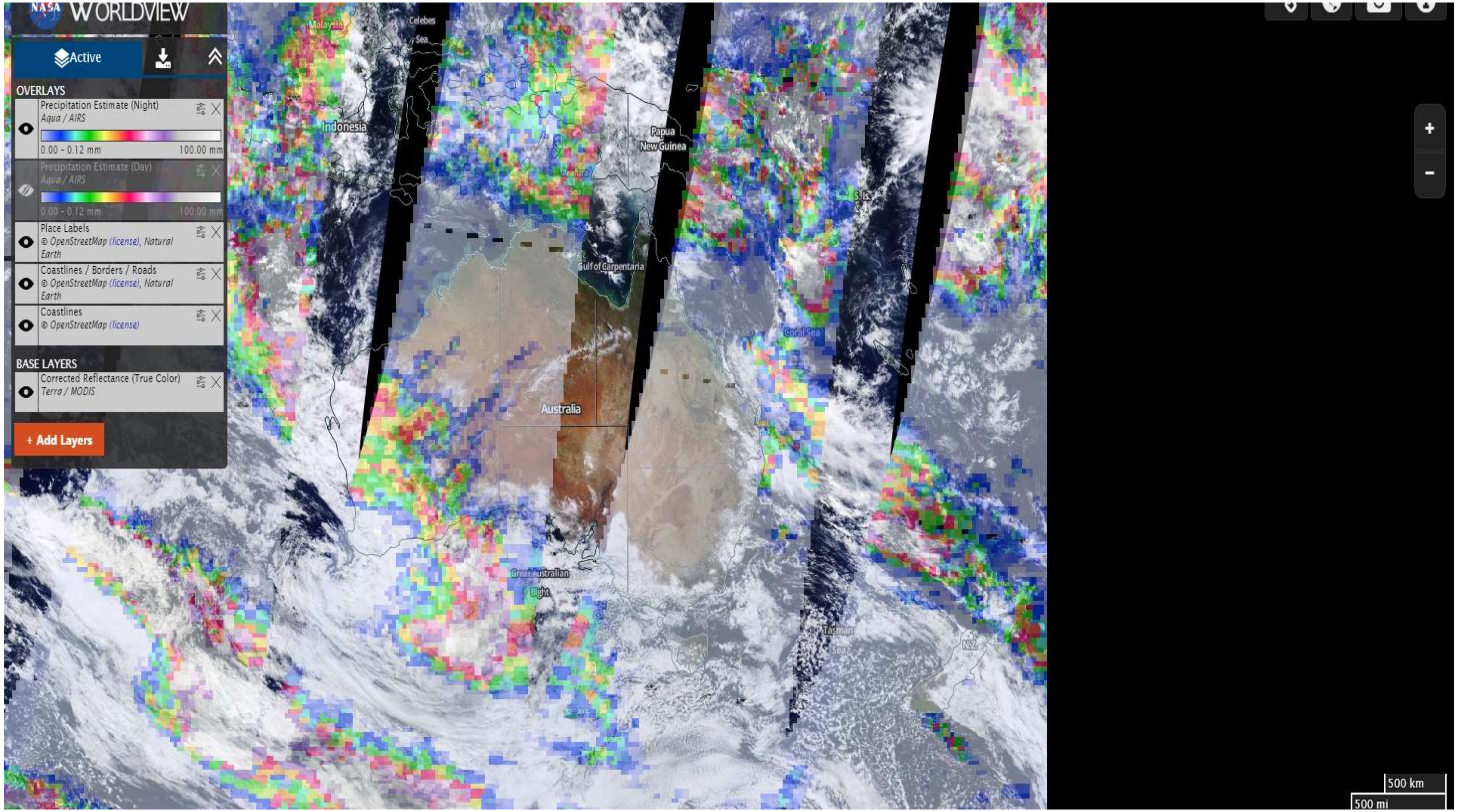
Binietoglou et al., AMT, 2015

Compare with satellite images

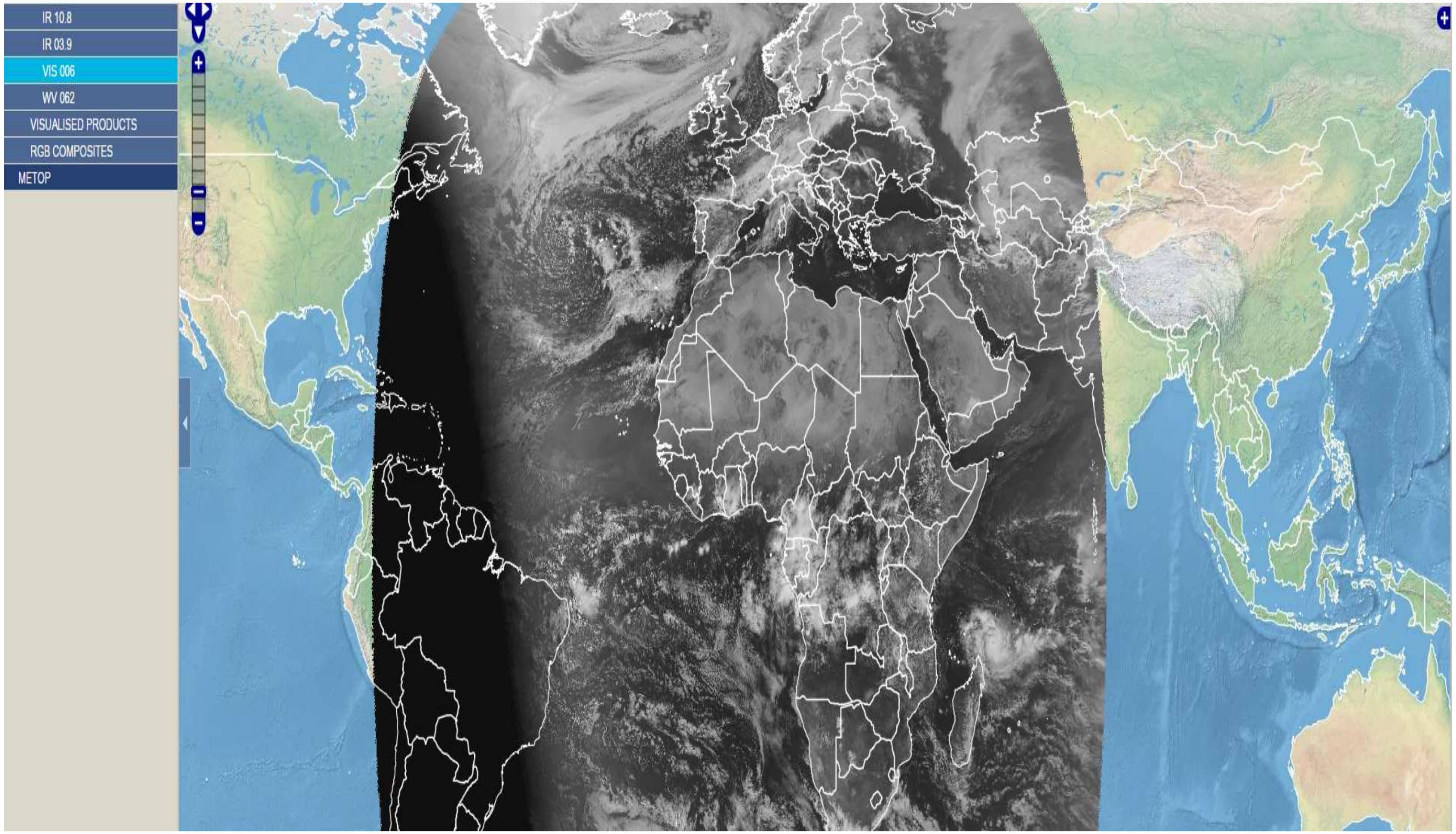


https://worldview.earthdata.nasa.gov/

Compare with satellite images



Compare with satellite images

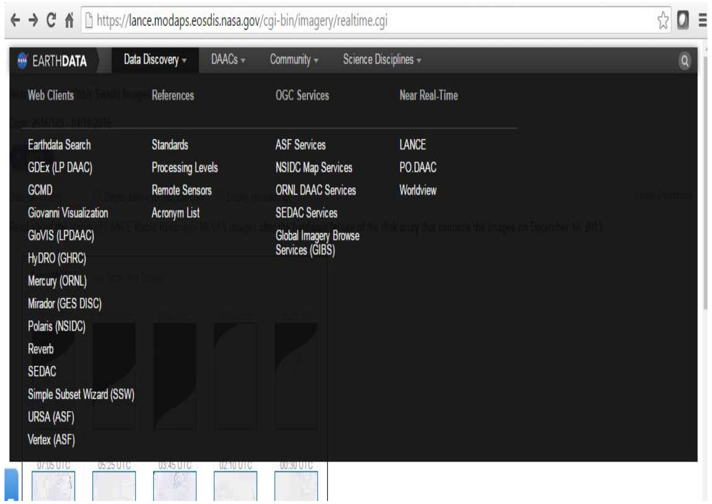


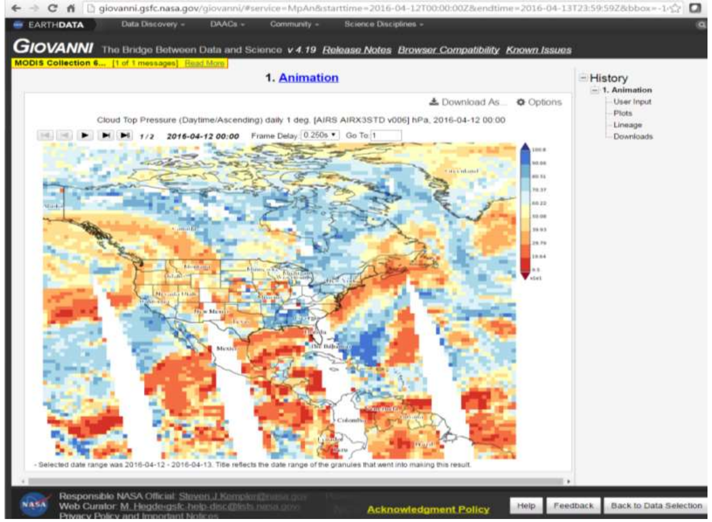
http://eumetview.eumetsat.int/mapviewer/

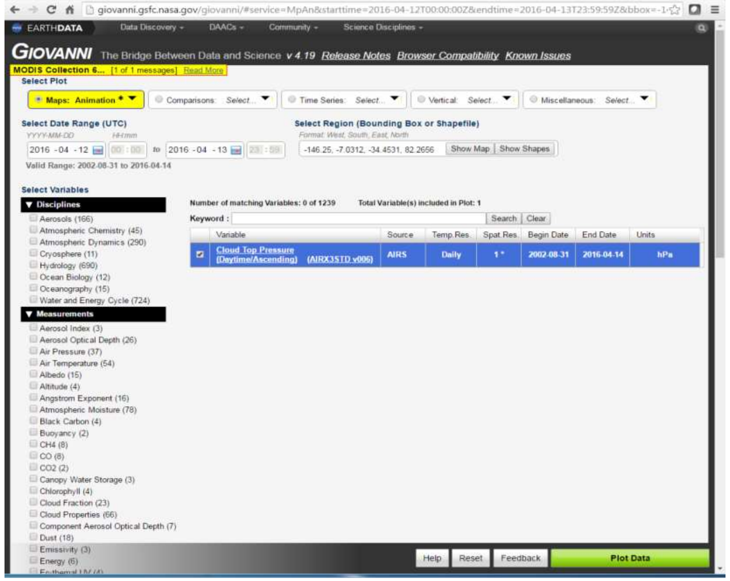
Compare with satellite images

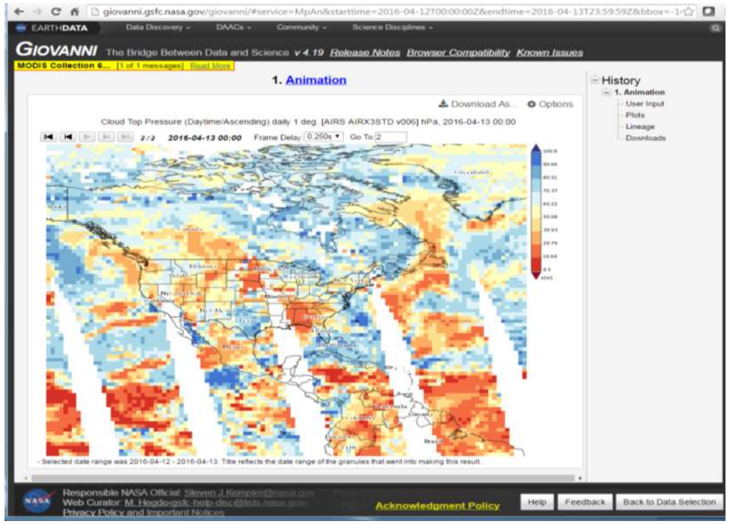
http://www.jma.go.jp/en/gms/

Compare with satellite images



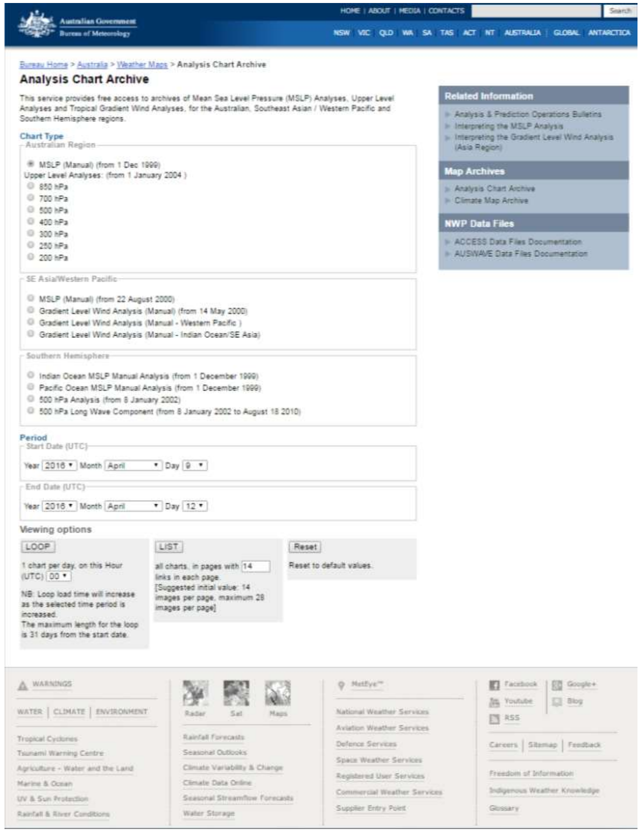


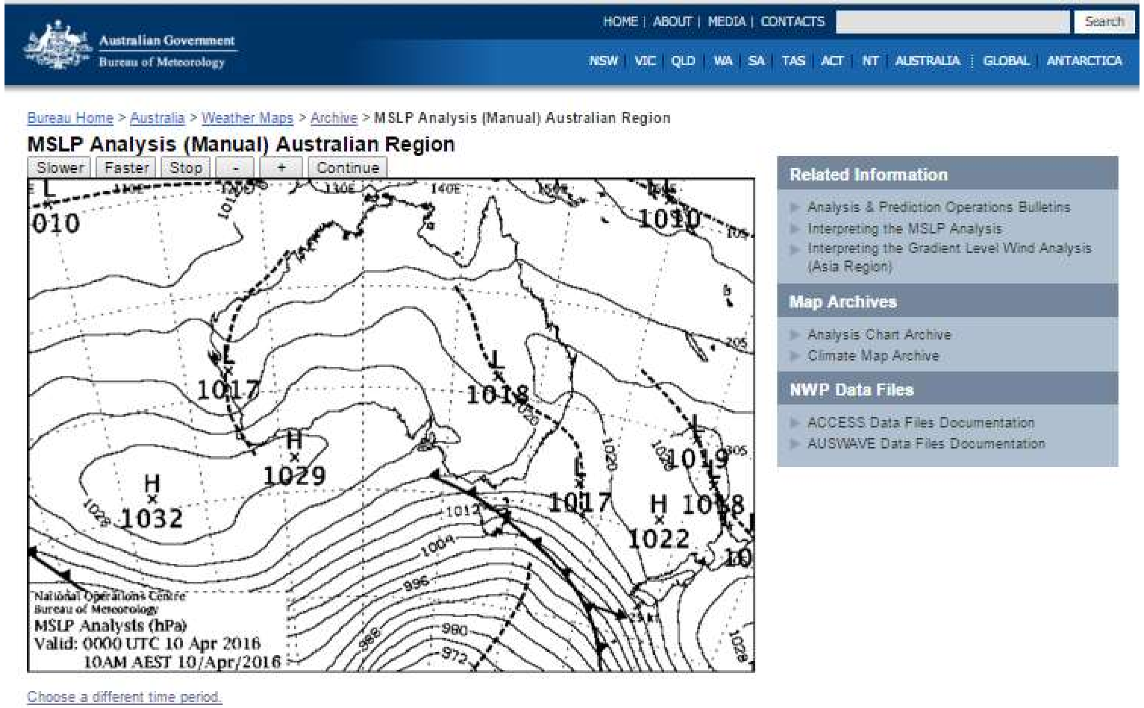




http://giovanni.sci.gsfc.nasa.gov/giovanni/

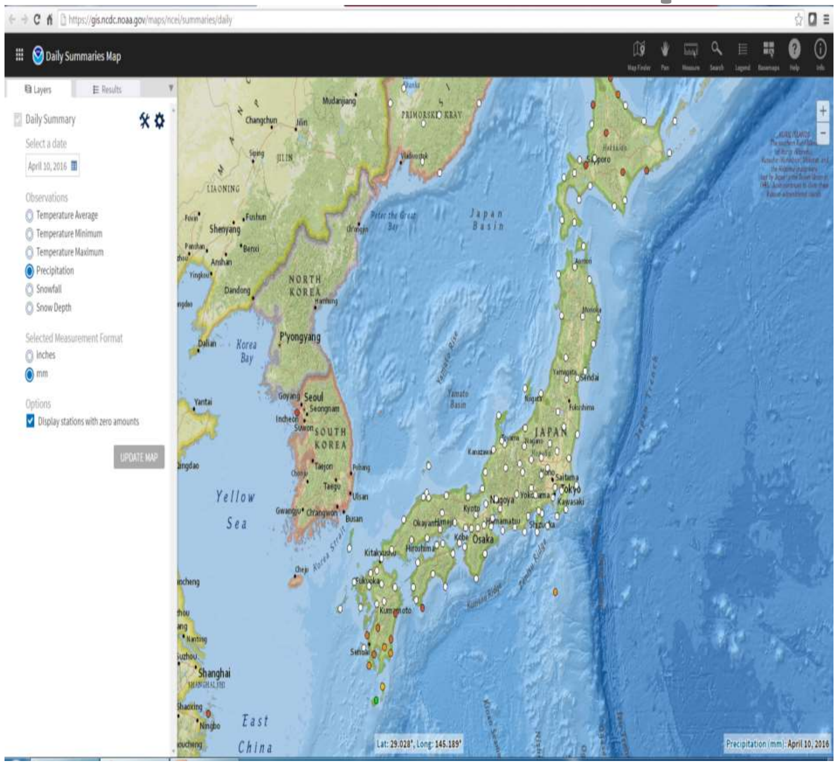
Compare with satellite images

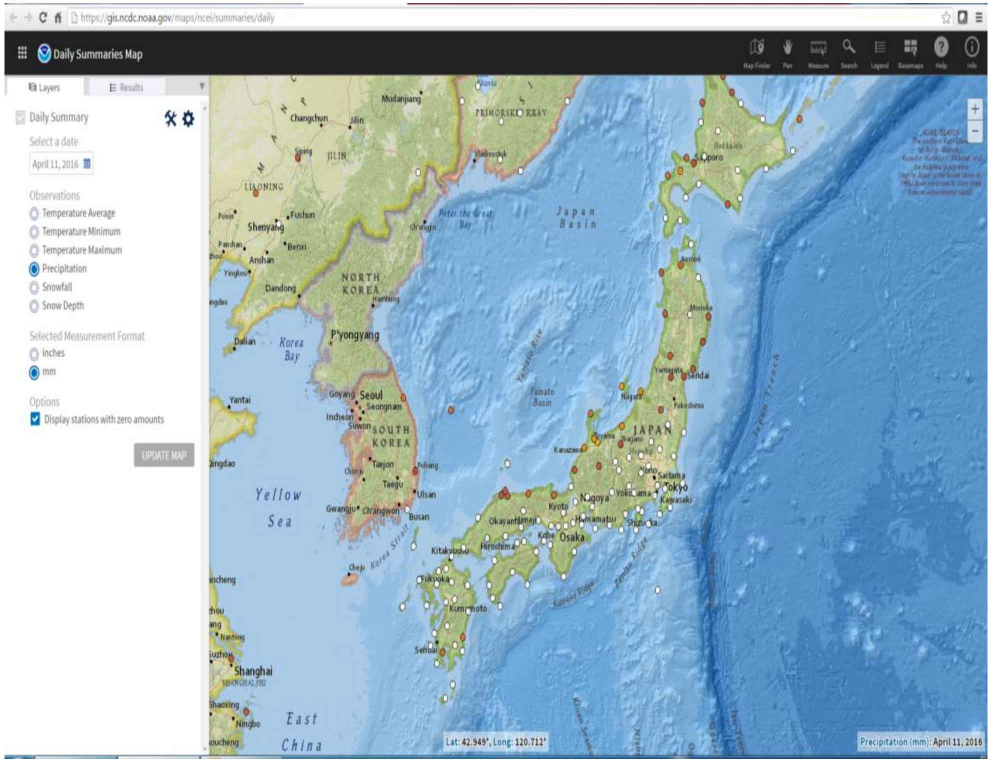




http://www.bom.gov.au/australia/charts/archive/

Compare with surface stations





https://gis.ncdc.noaa.gov/maps/ncei/summaries/daily

- Download station measurements
- *Tip – show stations with zero precipitation*

Smoke - biomass burning

Biomass burning (biochemical process)

gases

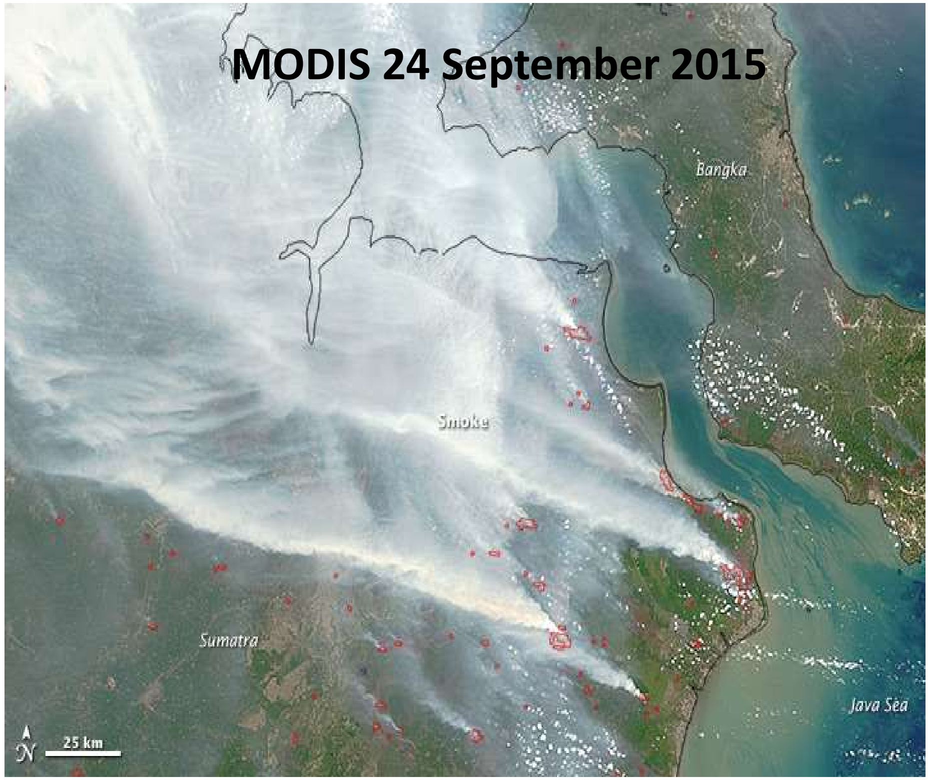
aerosols

Smoke risk for inhabited areas

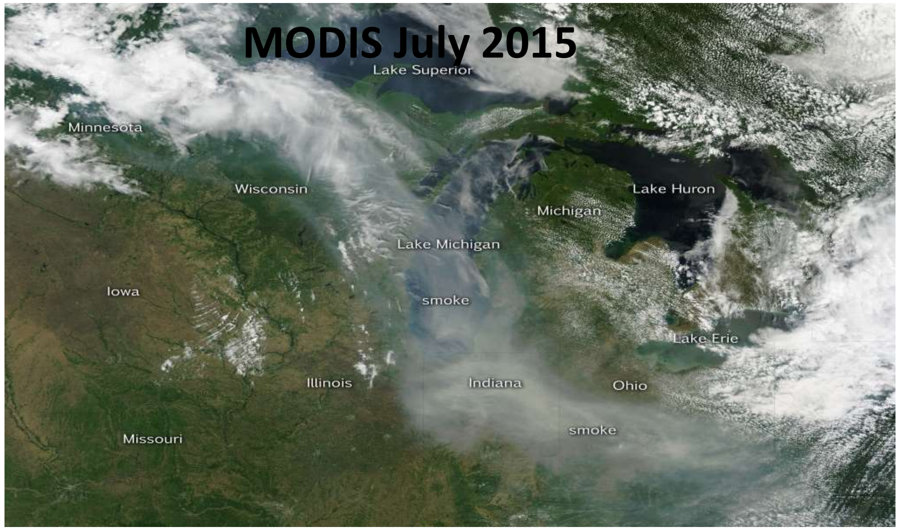
Climate change (tropical fires release 30% of global CO_2)

Satellite detection of fire hot-spots and biomass smoke

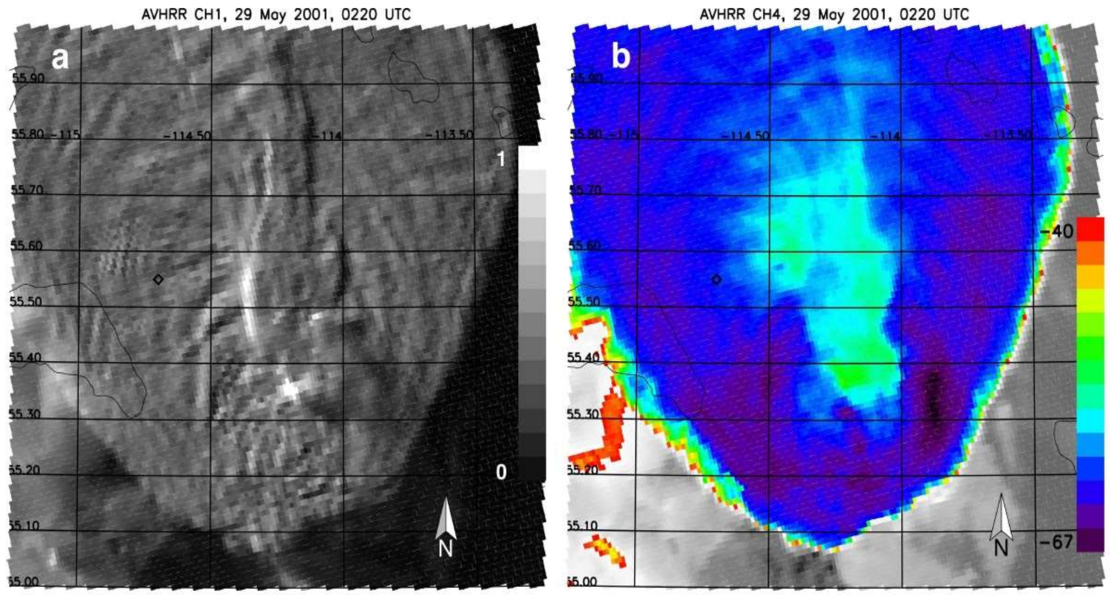
MODIS 24 September 2015



MODIS July 2015



Smoke - biomass burning



(a) 0.65 μm reflectance (b) 10.8 μm brightness temperature

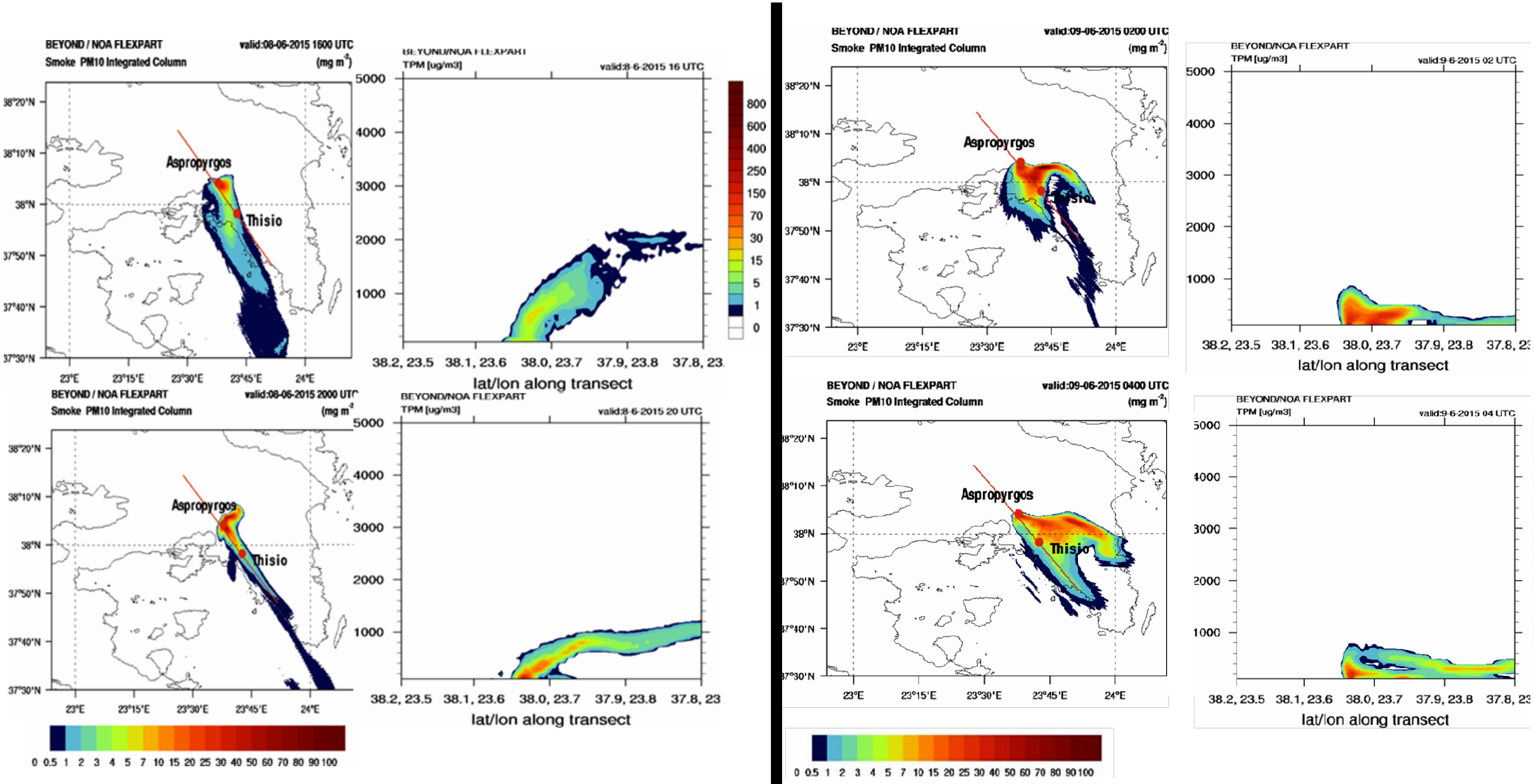
AVHRR reflectance (a) and brightness temperature (b) images of pyroconvection (overshooting – gravity waves)

Plume rise

- PBL
- Pyroconvection
- Injection of smoke in upper troposphere / lower stratosphere
- Generation of gravity waves enhances mixing at the top of the pyroCb
- Residence time of smoke in the atmosphere increases dramatically

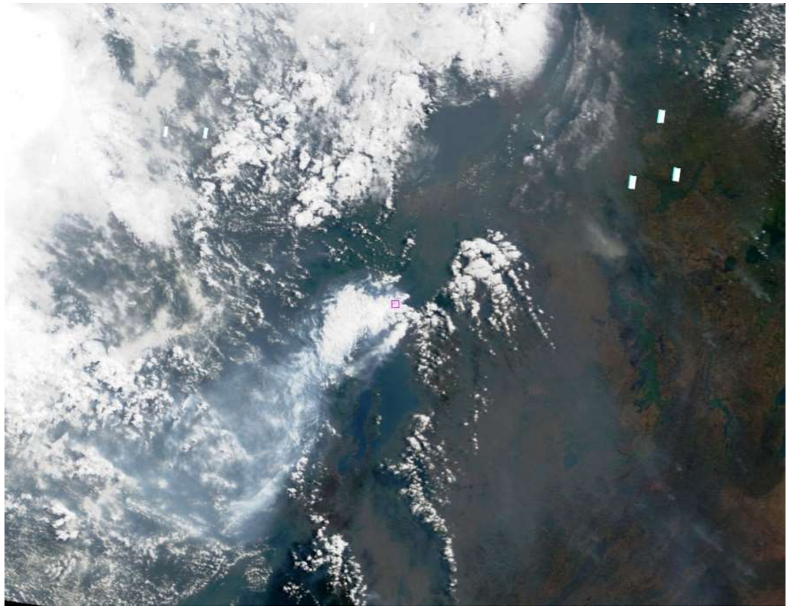
Rosenfeld et al., 2007

Smoke - biomass burning

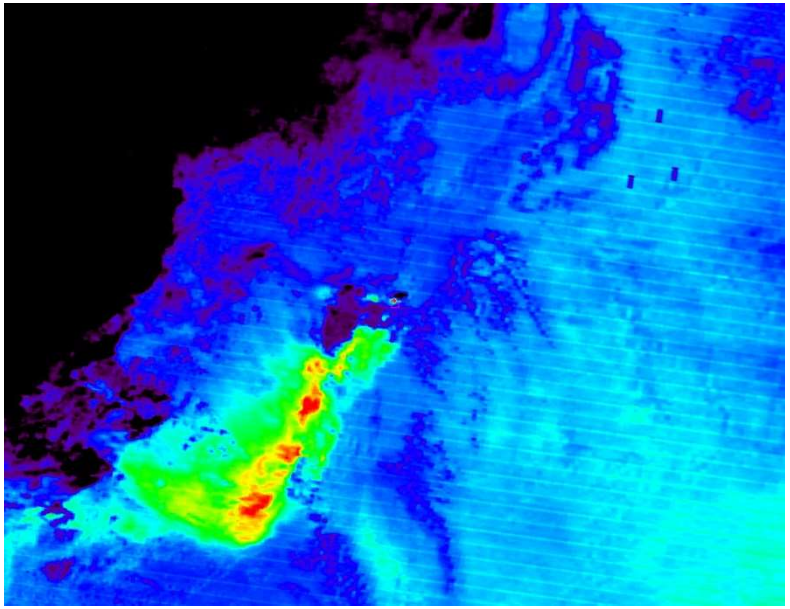


High PBL top - Deep mixing layer **Low PBL top - Temperature Inversion**

Volcanic emissions



MODIS visible image



MODIS infrared image

Volcanic emissions

Volcanic Eruptions → gases, aerosols

Volcanic Eruptions → Health risk for inhabited areas; Climate change considerations

Alaska, 1990

Emissions

- **H_2O** Water vapor (climate)
- **CO_2** Carbon Dioxide (health – climate)
- **SO_2** Sulfur Dioxide (health – climate effect (sulphates, ozone), satellite proxy)
- **H_2S** - Hydrogen Sulfide (toxic)
- **HF**, **HCl**, **HBr** - Hydrogen Halides (toxic)

Human nose is the most sensitive instrument to H_2S (0.000001% H_2S) - rotten egg smell.

Volcanic emissions

The AURA Ozone Monitoring Instrument (OMI) (daytime detections of SO_2)

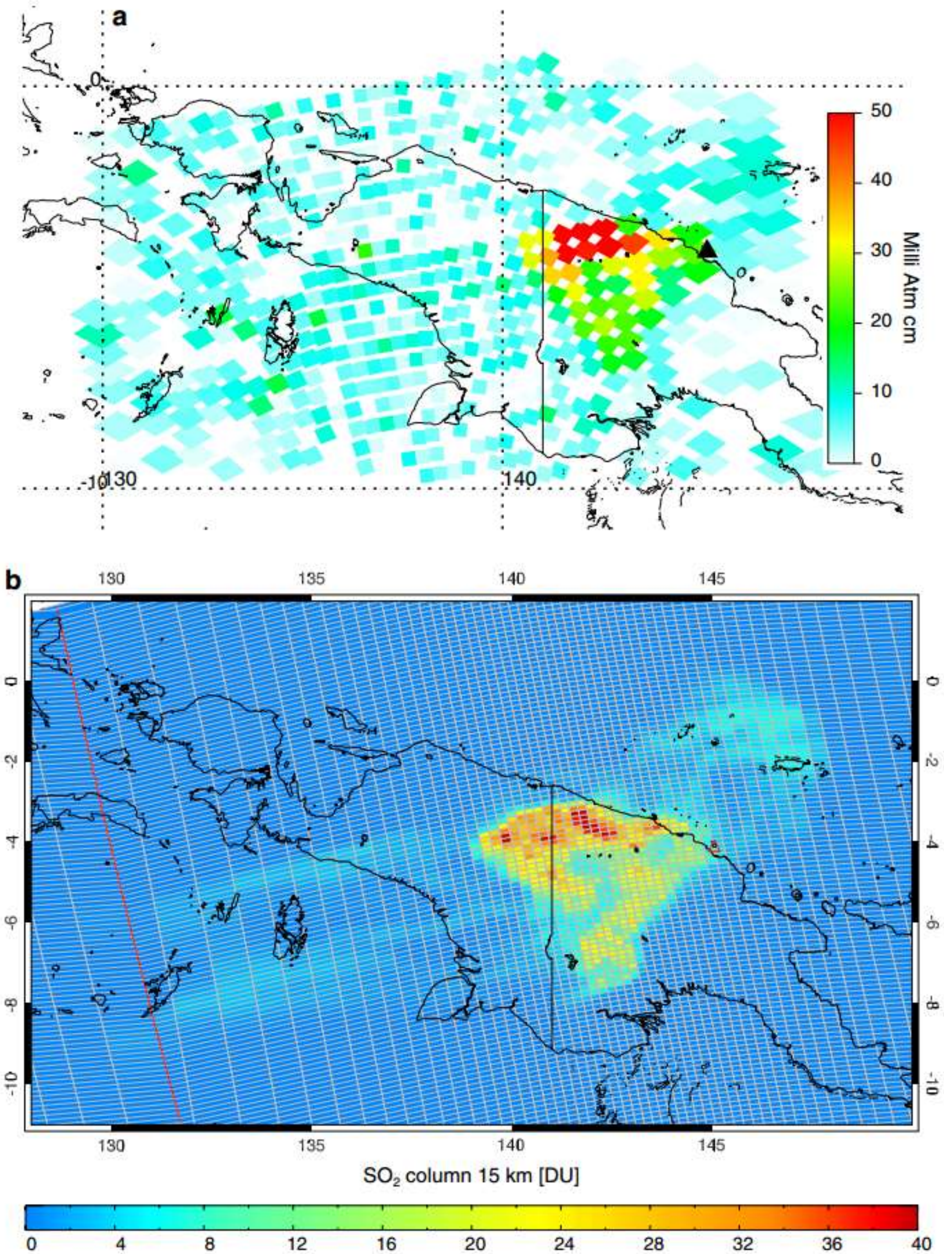


Fig. 1 Comparison of TOMS and OMI SO_2 retrievals for the Manam (Papua New Guinea) eruption of January 27, 2005 at 14:00 UT (00:00 LT on January 28). Both images show actual satellite footprints, which increase in size toward the edge of the orbit swath. (**a**) EP-TOMS overpass (orbit 45707) at 01:39–01:42 UT (11:39–11:42 LT) on January 28, 2005. *Color scale* shows retrieved SO_2 vertical column amount in milli atm cm (equivalent to Dobson Units). A *black triangle* indicates location of Manam; (**b**) OMI overpass (orbit 2867) at 04:13–04:15 UT (14:13–14:15 LT) on January 28. A *red triangle* indicates location of Manam; the *red line* to the left of the image is the edge of the next OMI orbit. Note the high background noise in the TOMS retrieval (~10 DU), which inhibits detection of the diffuse portions of the SO_2 cloud that can be seen northeast and west of the main cloud mass in the OMI image

Volcanic emissions

SO_2 satellite retrievals

Global Ozone Monitoring Experiment (GOME- 2)

- UV/visible spectrometer covering the 240–790 nm wavelength interval with a spectral resolution of 0.2–0.5 nm
- On board the Meteorological Operational satellite-A (MetOpA)
- Ground pixel size 80 km × 40 km.

The hyperspectral Infrared Atmospheric Sounding Interferometer (IASI)

- Spectral coverage from 645 to 2760 cm^{-1} , resolution 0.5 cm^{-1}
- Onboard MetOp-A
- Resolution 12 km at nadir

OMI

- Nadir-viewing imaging spectrograph
- Measures atmosphere-backscattered sunlight in the ultraviolet-visible range from 270 to 500 nm with a spectral resolution of about 0.5 nm
- Resolution 13 km × 24 km at nadir

Moderate Resolution Imaging Spectroradiometer (MODIS)

- Multispectral instrument on board the Terra and Aqua polar satellites
- 36 spectral bands from visible to thermal infrared
- Spatial resolution varies between 250, 500 and 1000 m.

Volcanic emissions

DU=Dobson Unit=0.01 mm thickness at STP

SO_2 satellite retrievals + modelling emissions

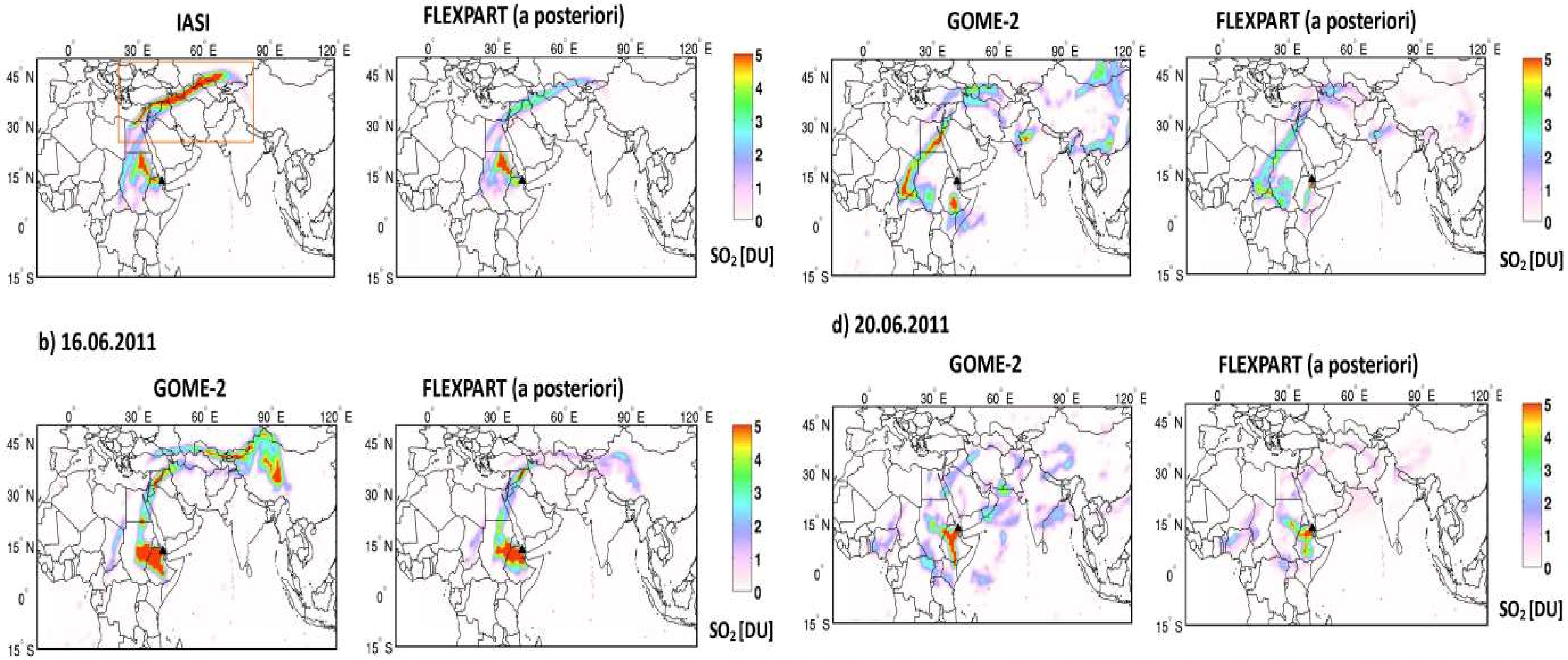


Fig. 10. SO_2 columns measured by IASI (morning overpass) and GOME-2 and simulated by FLEXPART using the a posteriori emissions from the inversion for 15 June 2011 **(a)**, 16 June 2011 **(b)**, 18 June 2011 **(c)** and 20 June 2011 **(d)**. The measured columns are averaged over the $1° \times 1°$ FLEXPART output grid. The simulated columns are (1) calculated from the SO_2 profiles weighted by the corresponding altitude-dependent measurement sensitivity functions (averaging kernels) and (2) interpolated at the time of observations. The Nabro volcano is marked by a black triangle. The orange box shows the SO_2 plume released in the first 15 h of the eruption.

Theys et al., 2012, ACP

Data assimilation

Why use remote sensing data in numerical models ?

1. Numerical models solve initial and boundary value problems (differential equations)
2. These conditions must be provided by observation (weather stations, balloons, etc.)
3. Some air-quality models (e.g. dust models) rely on their own forecasts for initial and boundary conditions (warm start)
4. Even at the idealized case of a perfect model run, this methodology would imply error propagation from numerical diffusion itself
5. For natural hazards such as **biomass smoke** or **volcanic ash** there is **no other way** to get initial conditions

Data assimilation

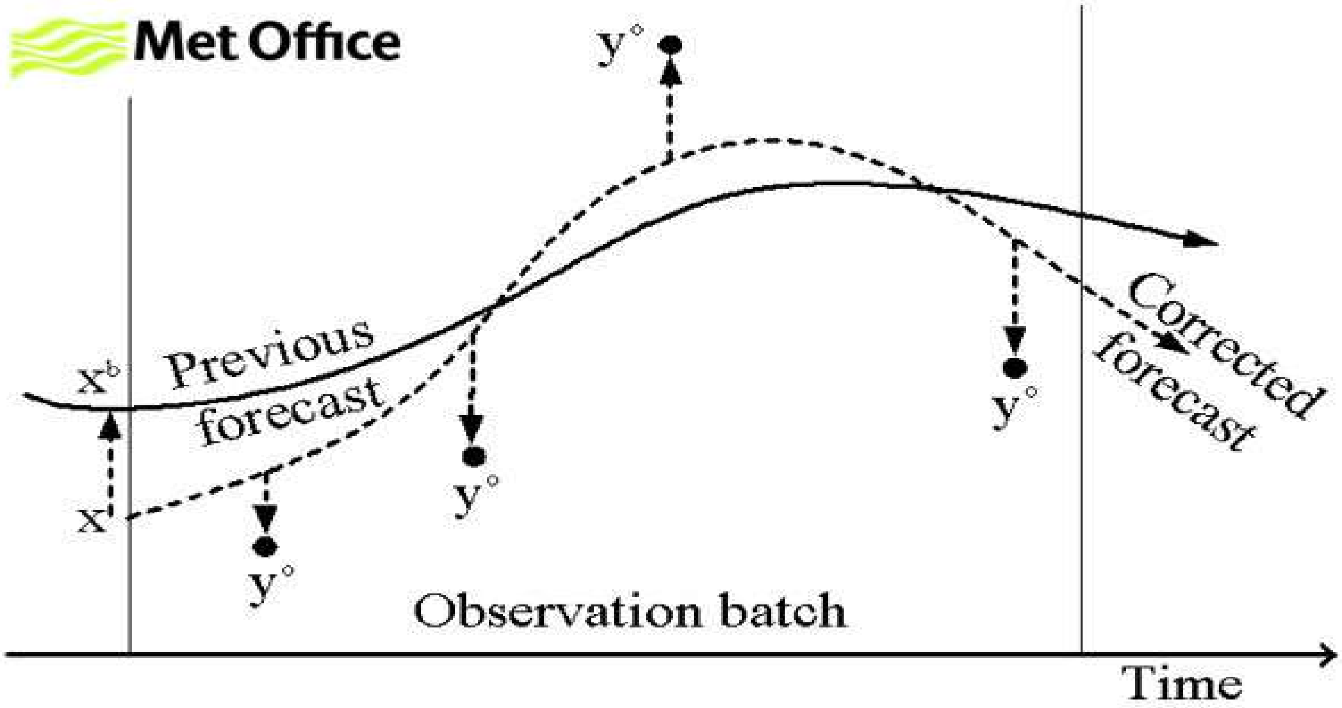
Data assimilation is an analysis technique in which the observed information is accumulated into the model state by taking advantage of consistency constraints with laws of time evolution and physical properties.

- **sequential assimilation** considers only observations made in the past (real-time forecasting systems)
- **non-sequential**, or retrospective assimilation, where observation from the future can be used, for instance in a reanalysis exercise

- **intermittent method**, observations are processed in small batches (technically convenient)
- **continuous method**, observation batches over longer periods are considered, and the correction to the analyzed state is smooth in time,which is physically more realistic

Meteorological Training Course Lecture Series, ECMWF, 2002

Data assimilation



The need for a statistical approach

- If we have a preliminary estimate of the analysis with a good quality, we do not want to replace it by values provided from poor quality observations.
- When going away from an observation, it is not clear how to relax the analysis toward the arbitrary state
- An analysis should respect some basic known properties of the true system, like smoothness of the fields, or relationship between the variables

Data assimilation

- The data that can go into the analysis system comprises the **observations**, the **first guess** and the **known physical properties of the system**.
- All pieces of data are important sources of information.
- There are errors in the model and in the observations, so we can never be sure which one to trust.
- However we can look for a strategy that minimizes on average the difference between the analysis and the truth.
- To design an algorithm that does this automatically, it is necessary to represent mathematically the **uncertainty** of the data.
- This uncertainty can be measured by calibrating (or by assuming) their error statistics using probabilistic concepts.
- Then the analysis algorithm can be designed on a formal requirement that in the average the analysis errors must be minimal.
- This will allow us to write the analysis as an optimization problem.

Data assimilation

- **State vector x** = a column matrix that represents the atmospheric state of the model
- **True state $\mathbf{x_t}$**= the best possible representation of reality
- **First guess (background) state $\mathbf{x_b}$**= The a priori or background estimate of the true state before the analysis is carried out
- **Analysis $\mathbf{x_a}$** = This is what we are looking for, $x_a=x_b+\delta x$
- **Space Operator H** = Interpolation from model space to observation space
- **Vector of errors $\boldsymbol{\varepsilon_b}$**= before doing an analysis, there is one and only one vector of errors that separates $\mathbf{x_b}$ from the true state, $\varepsilon_b=x_b-x_t$

The analysis problem is to find a ***correction δx*** *such that x_a is as close as possible to x_t*

Why is it not possible to precisely represent reality ?
Representativeness errors due to model discretization

Data assimilation

We don't want to know the errors but we need to know their statistics!

- Given a background field just before doing an analysis, there is one and only one **vector of errors (ε_b)** that separates it from the true state: $\varepsilon_b = x_b - x_t$
- If we were able to repeat each analysis experiment a large number of times, under exactly the same conditions, but with different realizations of errors generated by unknown causes, $\boldsymbol{\varepsilon_b}$ would be different each time.
- We calculate statistics such as averages, variances and histograms of frequencies of error and expect the statistics to converge to values which depend only on the physical processes responsible for the errors.
- The best information about the distribution of error is given by the *probability density function* PDF
- From this function one can derive all statistics, including the average (or expectation) and the variances

Data assimilation

Error variables

- **background errors**: $\varepsilon_b = x_b - x_t$, average $\overline{\varepsilon_b}$, **covariance** $\boldsymbol{B}$

The difference between the background state vector and its true value. They do not include discretization errors.

- **observation (radiance) errors**: $\varepsilon_o = y - H(x_t)$, average $\overline{\varepsilon_o}$, **covariance R**

They contain errors in the observation process (instrumental errors, because the reported value is not a perfect image of reality), errors in the design of the operator , and representativeness errors

- **analysis errors**: $\varepsilon_a = x_a - x_t$, of average $\overline{\varepsilon_a}$

They are the estimation errors of the analysis state, which is what we want to minimize

The averages of errors are called biases and they are the sign of a systematic problem in the assimilating system: a model drift, or a bias in the observations, or a systematic error in the way they are used.

Data assimilation

Best Linear Unbiased Estimator (BLUE)
How the least-squares estimation can be simplified to yield the most common algorithms used nowadays in meteorology and oceanography.

The BLUE analysis is equivalently obtained as a solution to the *variational optimization problem*:

$X_a = \mathit{ArgMin}(J)$

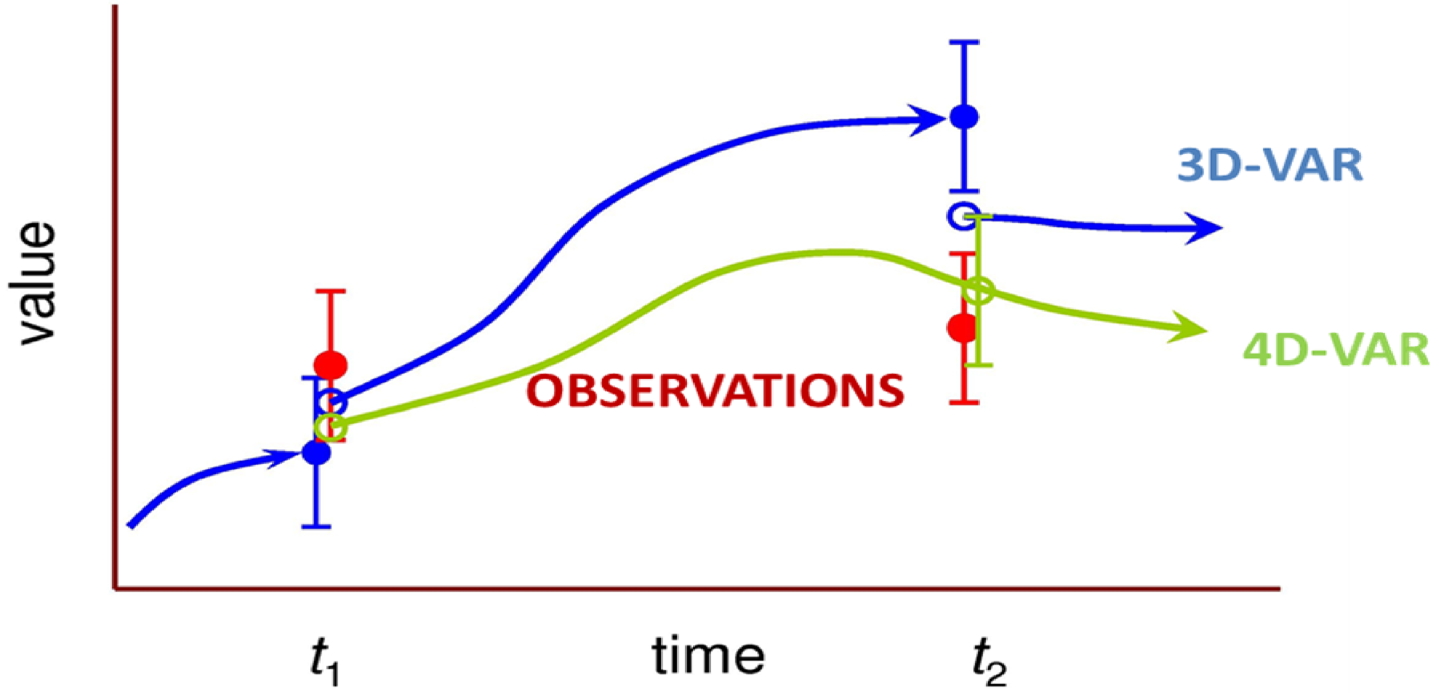
$J(x) = (x - x_b)^T B^{-1} (x - x_b) + (y - H[x])^T R^{-1} (y - H[x])$, **3D-var**

$J(x) = (x - x_b)^T B^{-1} (x - x_b) + (y - H[x])^T R^{-1} (y - H[x]) +$
$(H_2[x] (M(x)) - y_2)^T R_2^{-1} (H_2[x] (M(x_a) - y_2)$, **4D-var**

J = cost function, M is model forecast (t_1 -> t_2)

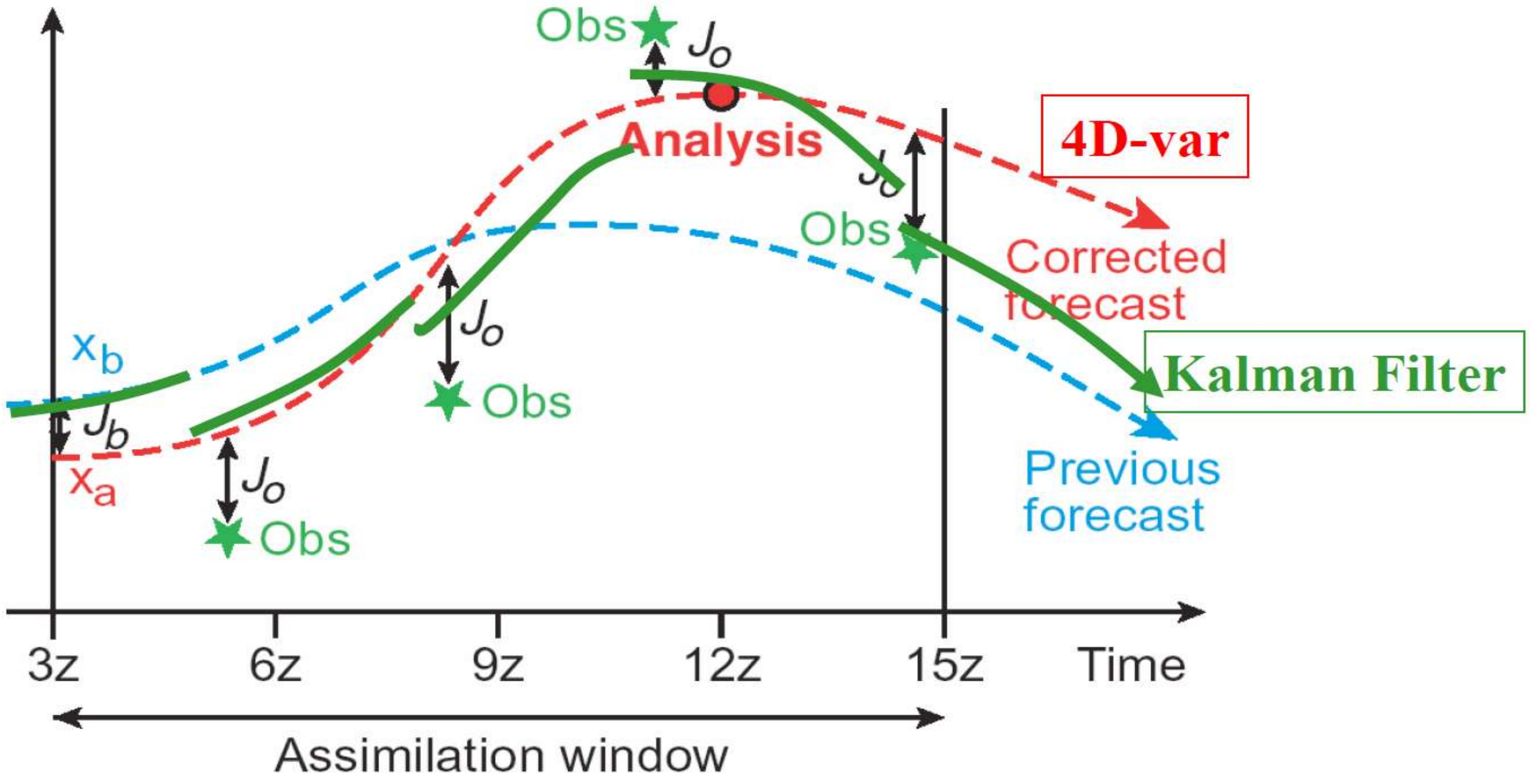
3D-var 4D-var assimilation techniques are based on the minimization of J

Data assimilation

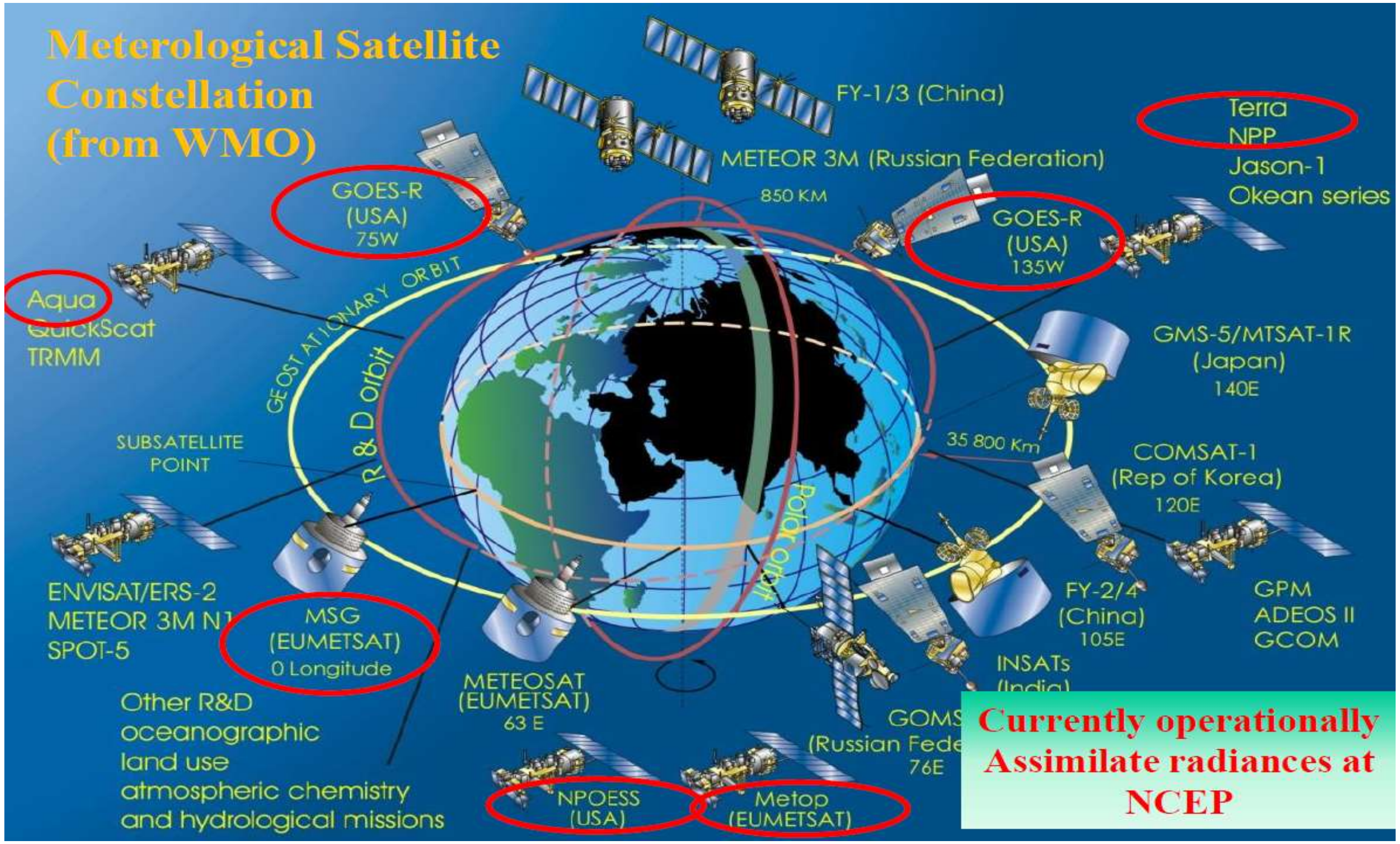


- 3D-VAR assumes all observations are at analysis time
- 4D-VAR incorporates also the time dimension

Data assimilation



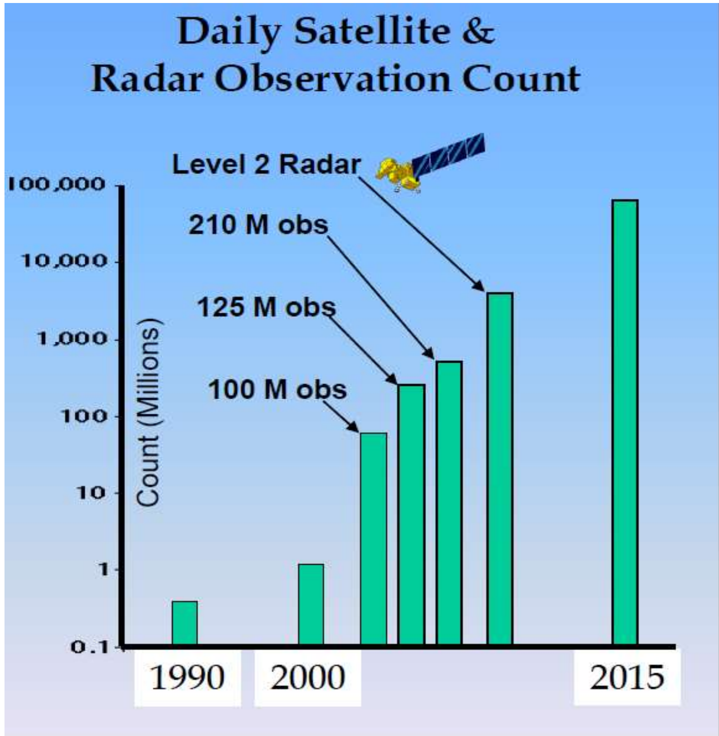
Data assimilation



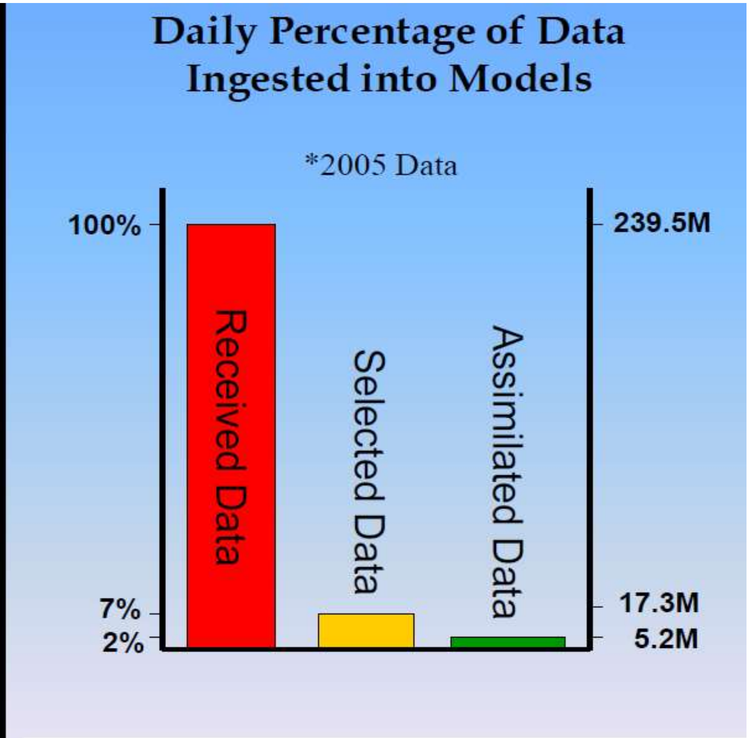
Satellite data must be treated carefully

- Important to be aware of instrument characteristics before attempting to use data.
- No current component of observing system is used “perfectly” or “as well as possible”.
- Computational expense plays important role in design of system.

Data assimilation



Five Order of Magnitude Increases in Satellite Data Over Fifteen Years (2000-2015)



Received = All observations received operationally from providers
Selected = Observations selected as suitable for use
Assimilated = Observations actually used by models

- Satellite instruments do not directly measure the atmospheric state
- Instead they measure radiation emitted by and/or transmitted by the atmosphere that is representative of the atmospheric state
- **But NWP need atmospheric variables**

Data assimilation

Quality Control

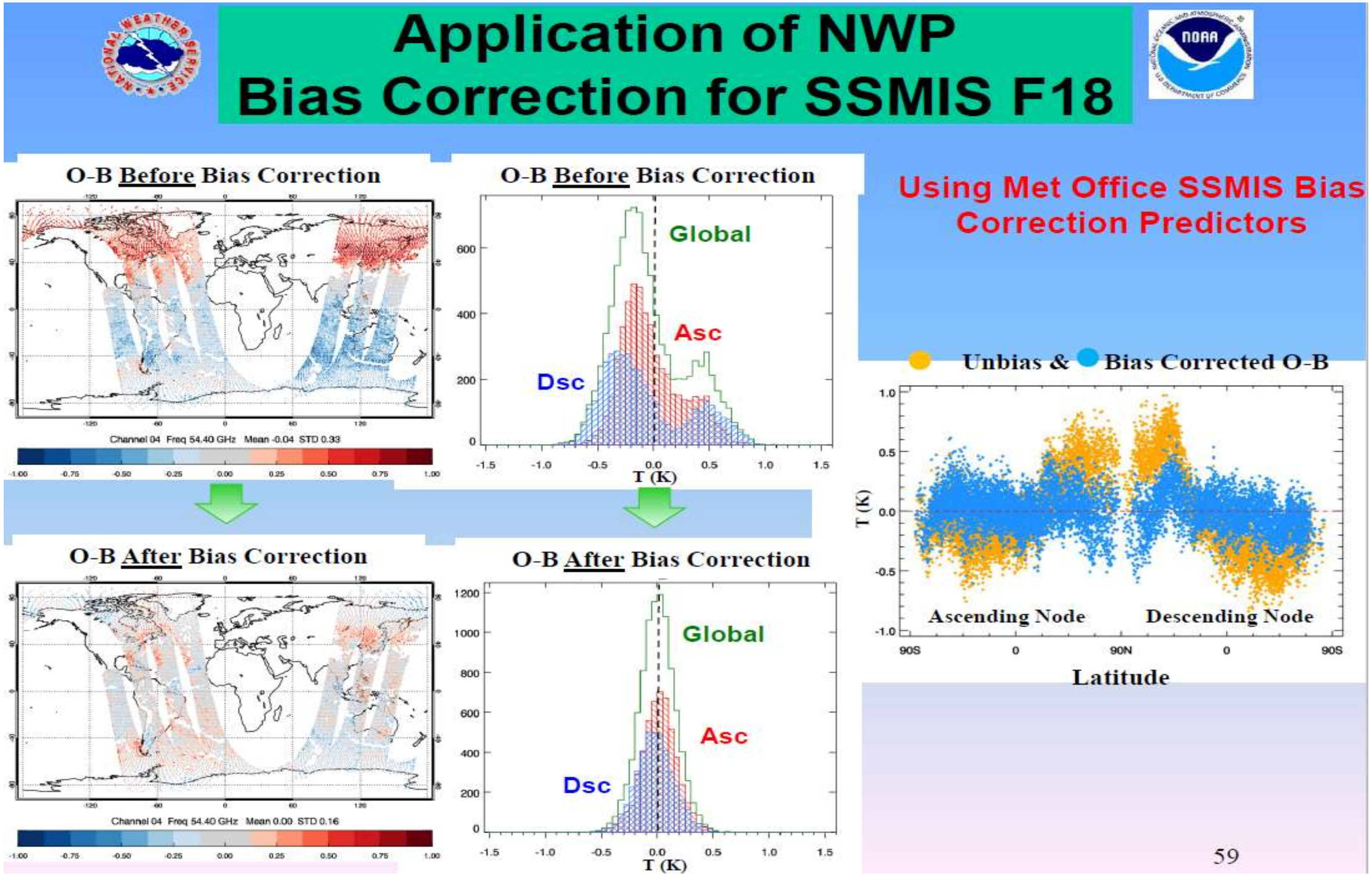
- The quality control step may be the most important aspect of satellite data assimilation.
- Most problems with satellite data come from 4 sources:
  1. Instrument problems.
  2. Clouds and precipitation simulation errors.
  3. Surface emissivity simulation errors.
  4. Processing errors (e.g., wrong height assignment, incorrect tracking, etc).
- IR cannot see through most clouds.
- Microwave impacted by clouds and precipitation but signal is smaller from thinner clouds.
- Surface emissivity and temperature characteristics not well known for land/snow/ice.
- Also makes detection of clouds/precip. more difficult over these surfaces.
- Error distribution may be asymmetric due to clouds and processing errors.

Data assimilation

Bias Correction

- The differences between simulated and observed observations can show significant biases.
- The source of the bias can come from:
  1. Inadequacies in the characterization of the instruments.
  2. Deficiencies in the forward models.
  3. Errors in processing data.
  4. Biases in the background.
- Except when the bias is due to the background, we would like to remove these biases.
- Currently bias correction only applied to a few data sets:
  1. Radiances.
  2. Radiosonde data (radiation correction and moisture).
  3. Aircraft data.
- For radiances, biases can be much larger than signal. Essential to bias correct the data.

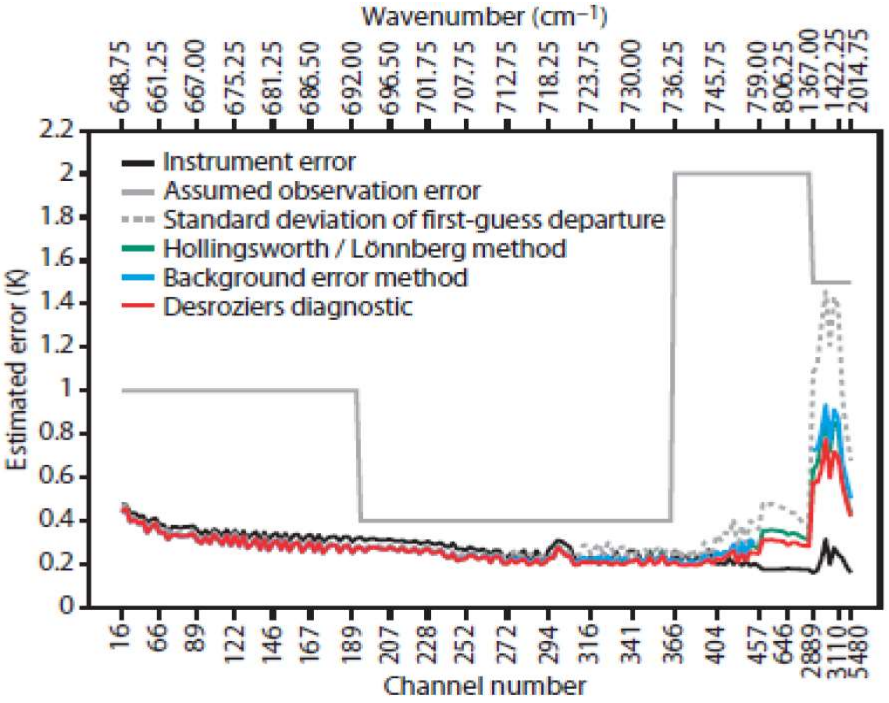
Data assimilation



Data assimilation

Observational Errors

- Observation errors specified based on instrument errors and statistics
- Generally for satellite data, variances are specified a bit large since the correlated errors (from RT and instrument errors) are not well known.
- Observation errors are also generally specified as being uncorrelated spectrally, but efforts are being made to determine the off-diagonal components of the observation error covariance matrix.



IASI Observation Errors in ECMWF System

Data assimilation

Thinning

–Reducing spatial or spectral resolution by selecting a reduced set of locations or channels.

–Can include "intelligent thinning" to use better observation.

Superobbing

–Reducing spatial or spectral resolution by combining locations or channels.

–Can reduce noise.

–Includes reconstructed radiances.

–Can include higher moments contained in data.

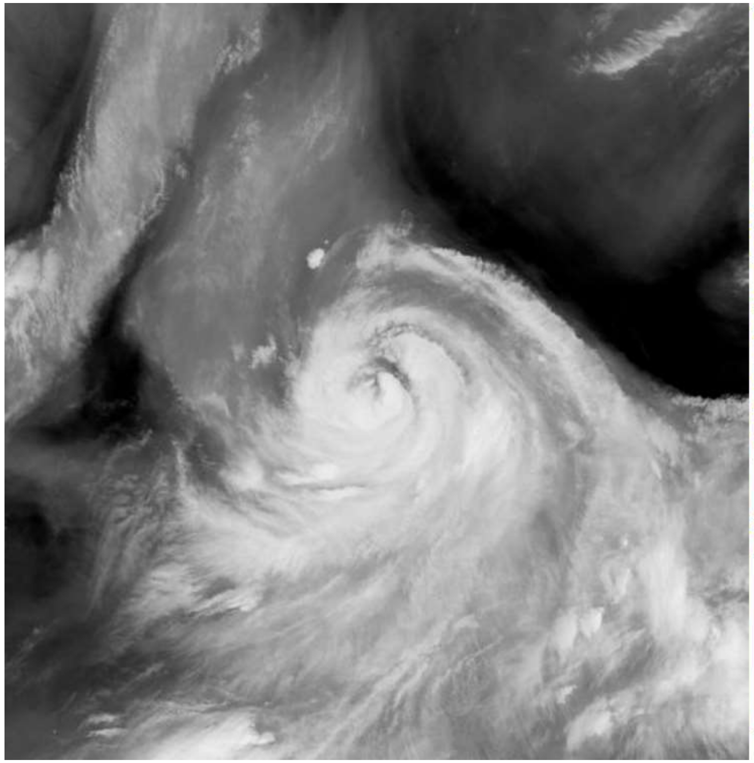
Both can be used to address 3 problems:

–Redundancy in data.

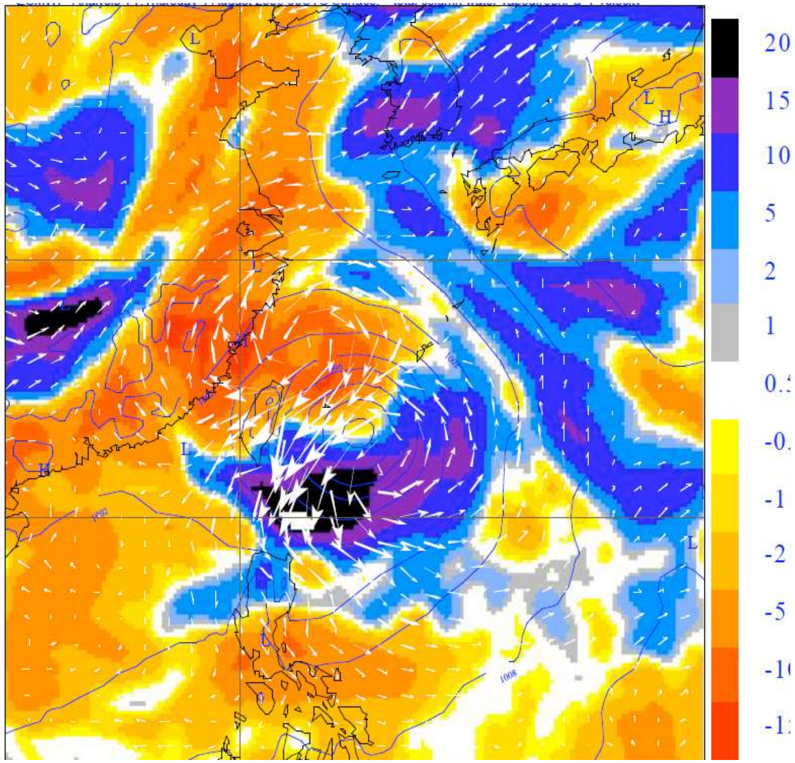
–Reduce correlated error.

–Reduce computational expense.

Data assimilation

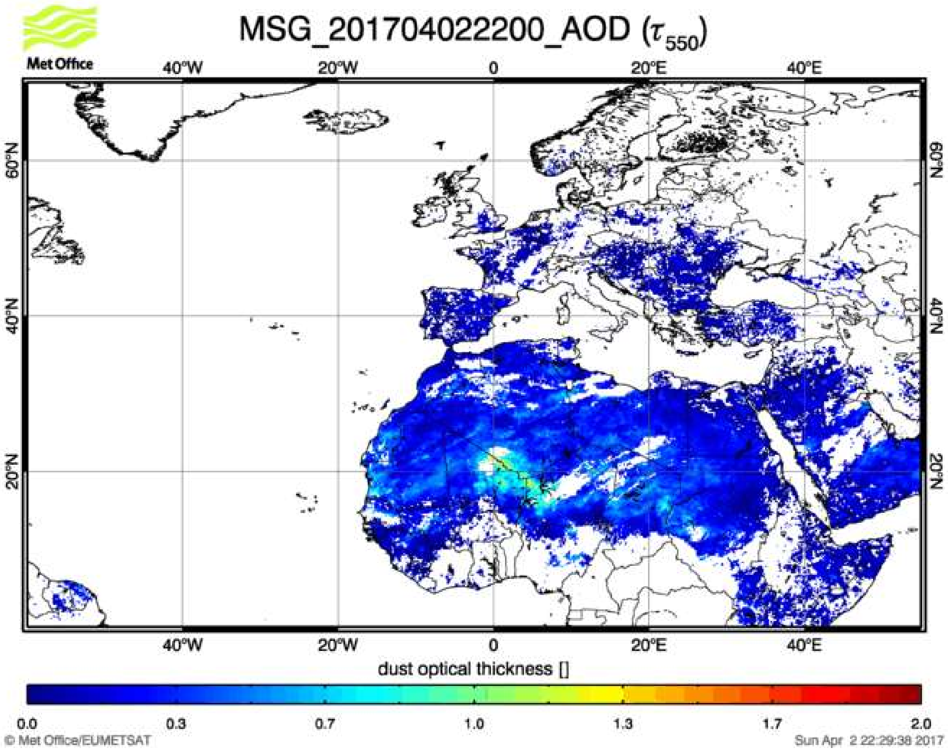


MTSAT Infrared image of typhoon MATSA approaching Taiwanese and Chinese coast on August 4, 2005, 00 UTC.



4DVar moisture increments with rain assimilation (colors in %), 900 hPa wind increments (white arrows), surface pressure (isolines), ECMWF model

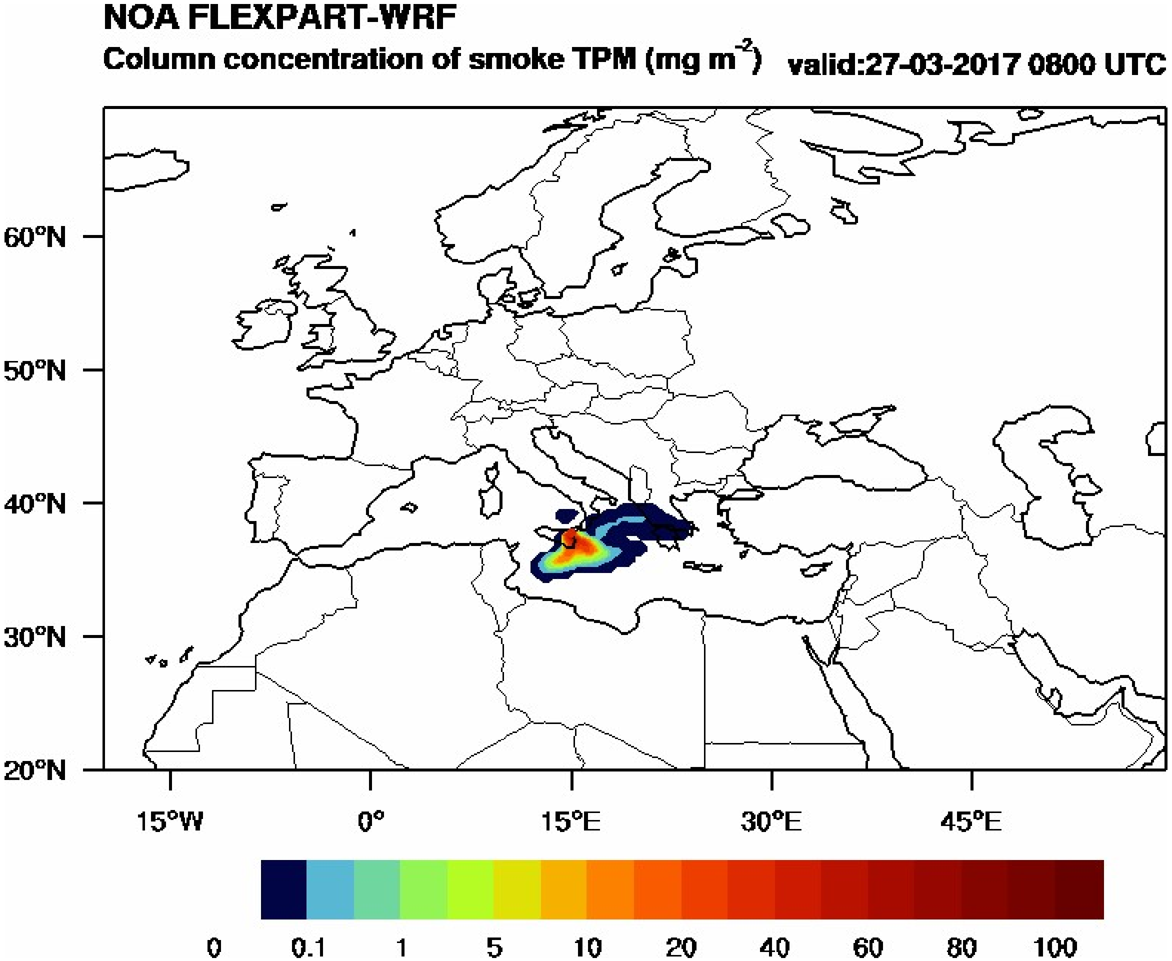
Data assimilation



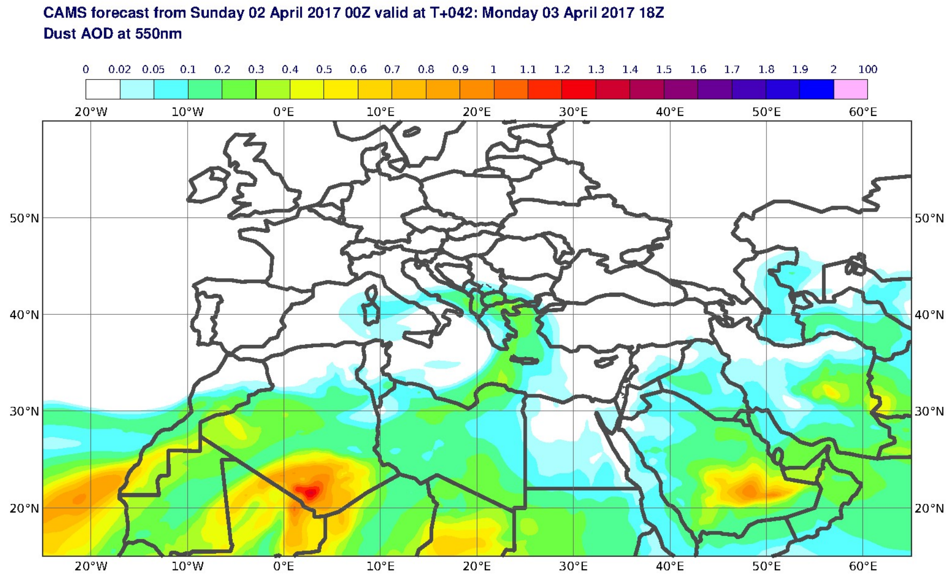
Dust Optical Depth from the UK Met Office SEVIRI retrieval algorithm

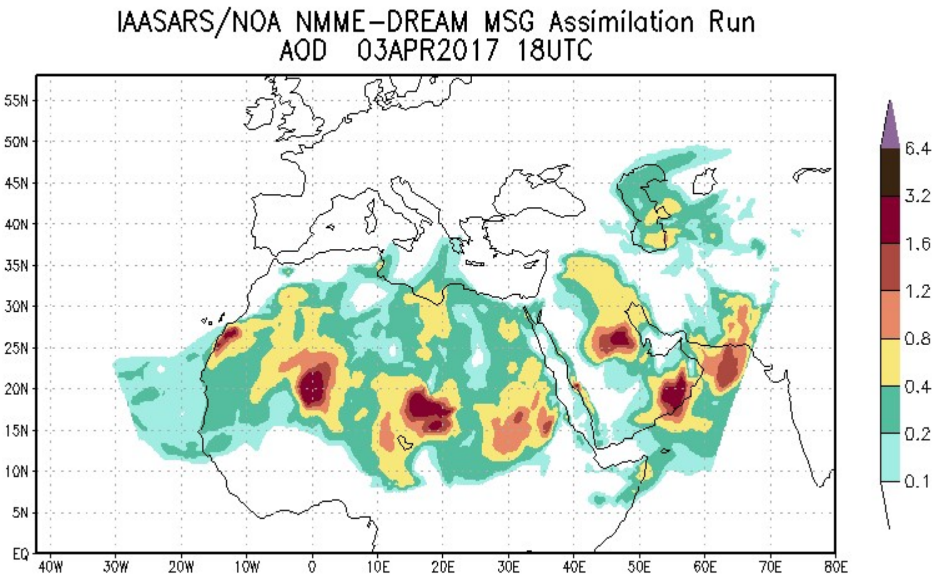
The U.K. Met Office MSG dust product shows an estimation of the dust optical thickness retrieved from empirical relationship between SEVIRI infrared (10.8 µm) radiance and aerosol optical depth at 550nm.
It is generated by transforming original retrievals to regularly-spaced grids (0.18 degree) using simple average method.

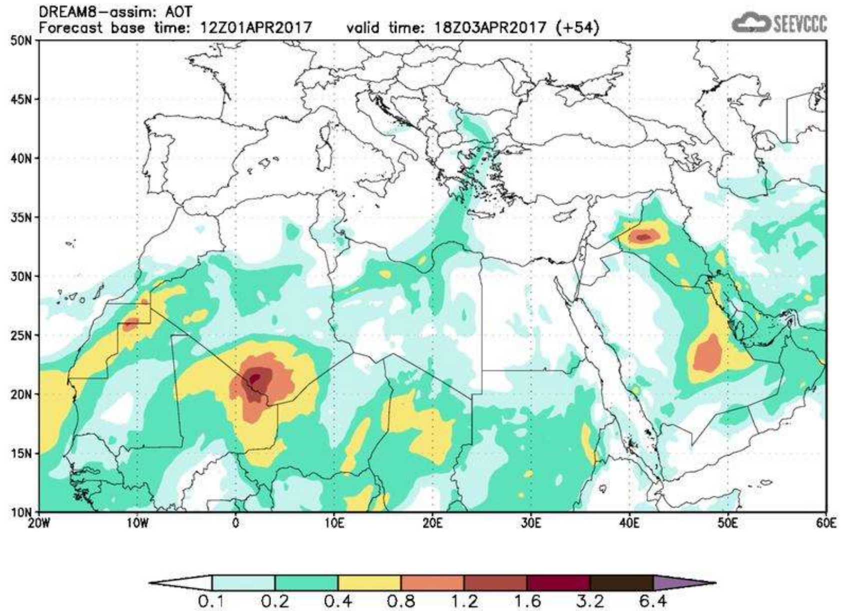
Brindley, H. E., and J. E. Russell (2009), JGR


NOA FLEXPART-WRF
Column concentration of smoke TPM (mg m^{-2}) valid:27-03-2017 0800 UTC
60°N
50°N
40°N
30°N
20°N
15°W
0°
15°E
30°E
45°E
0
0.1
1
5
10
20
40
60
80
100

Satellite data assimilation

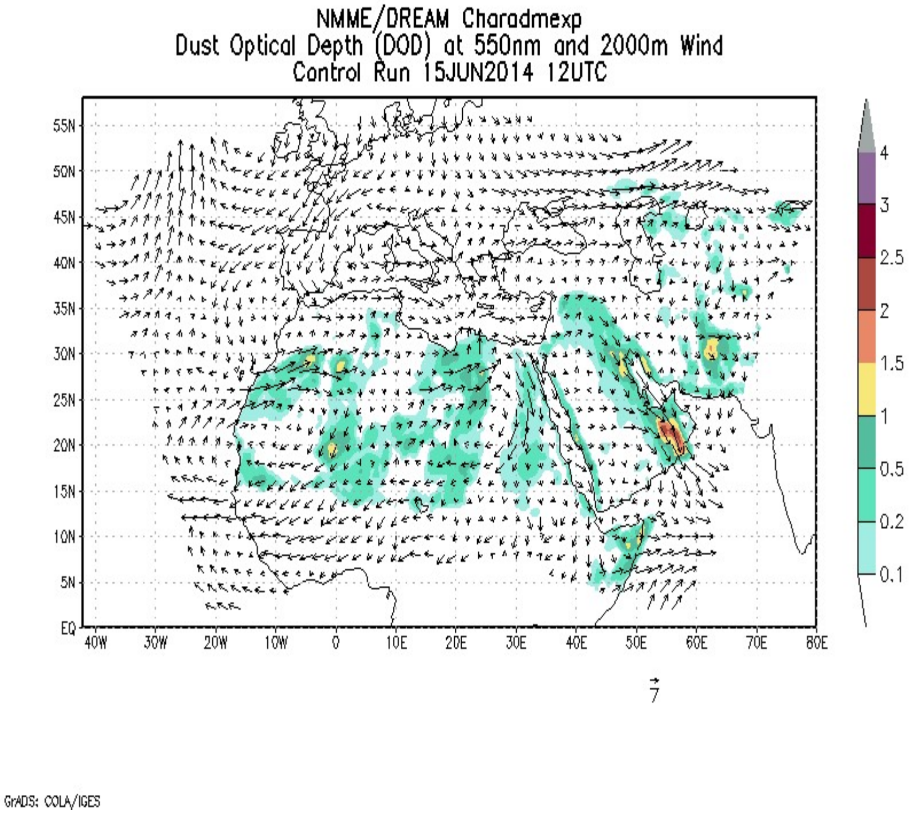


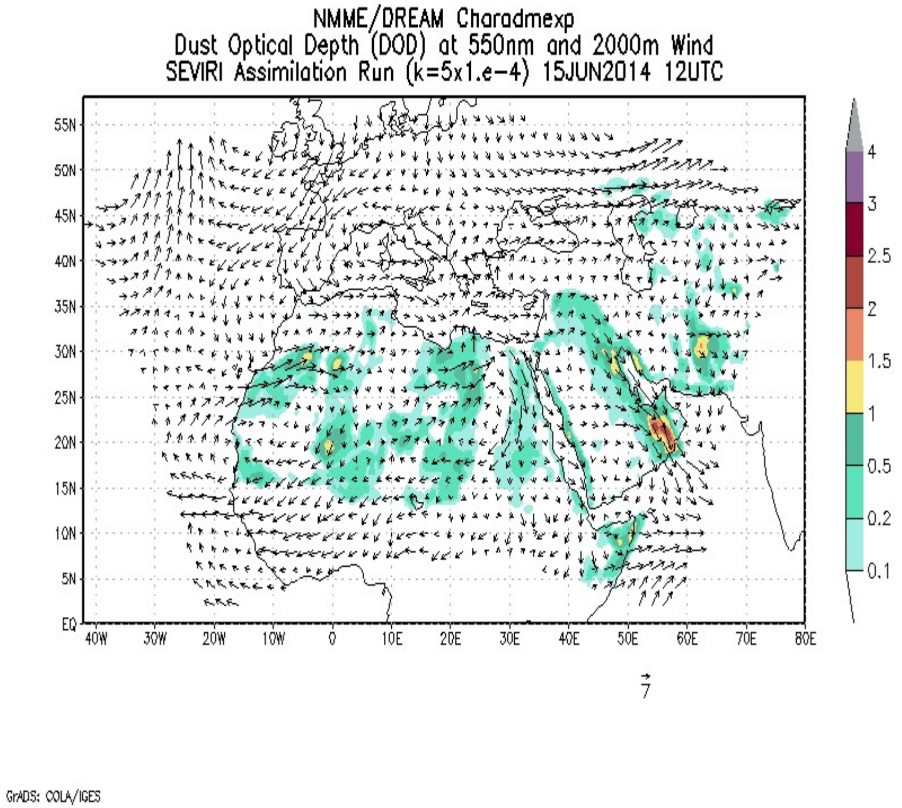




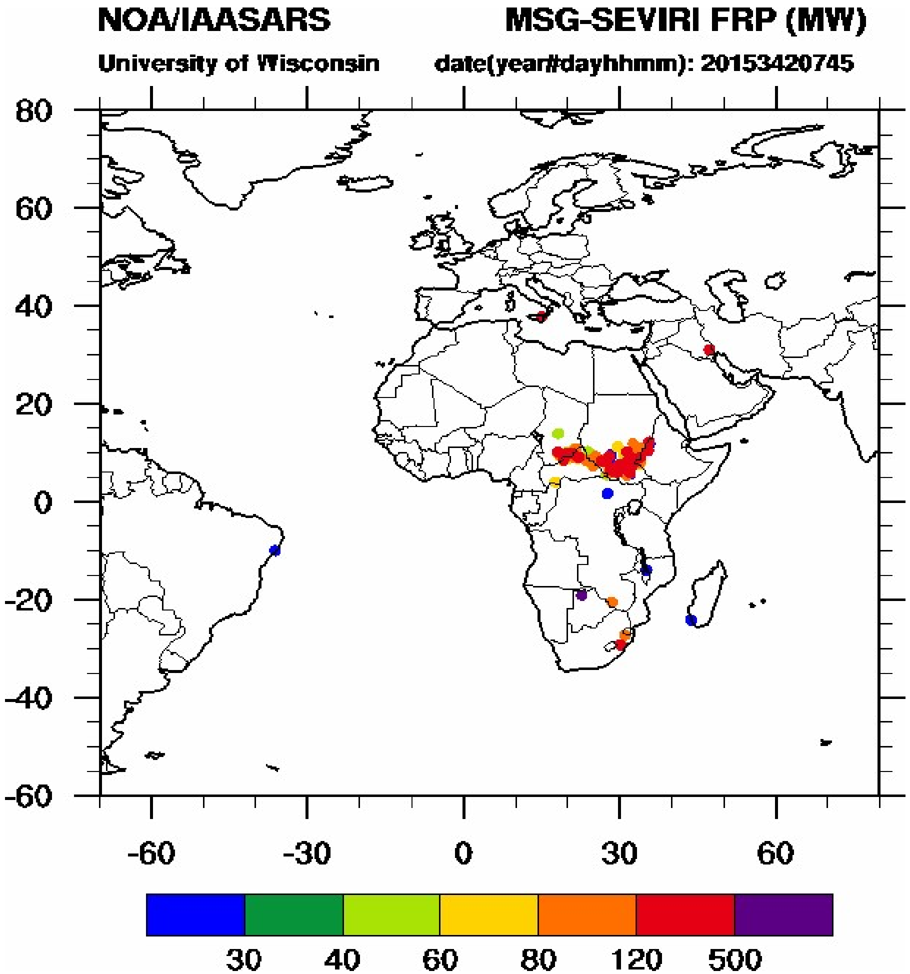
Data assimilation

Assimilation of dust retrievals from a geostationary sensor (MSG-SEVIRI) in atmospheric dust models (NMM-DREAM)



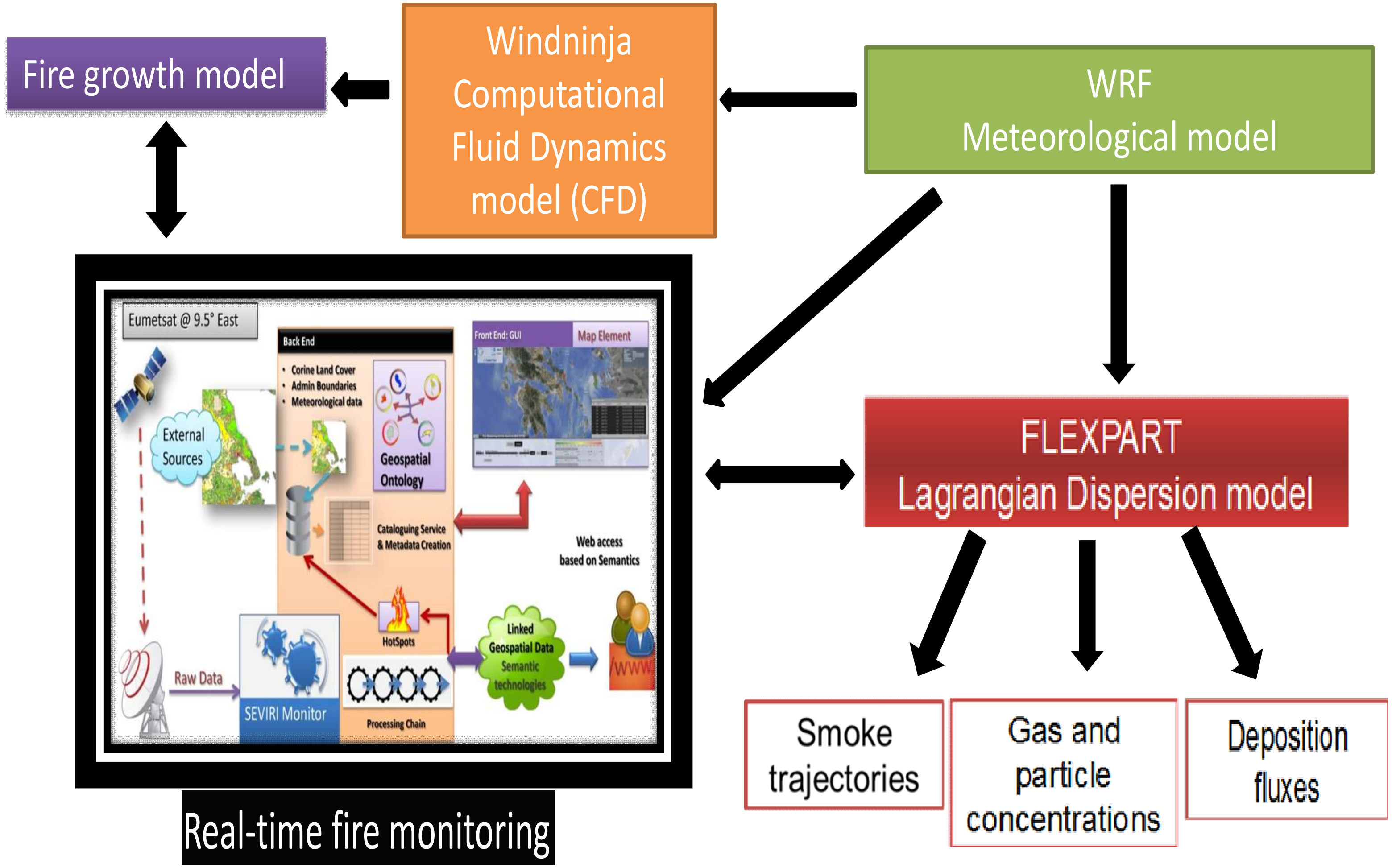


Data assimilation



- Fire Radiative Power (FRP) is a measure of fire intensity
- Assimilation of FRP in smoke dispersion models is used for the calculation of (i) smoke injection heights and (ii) smoke emission rates.

Data assimilation



Data assimilation