Optimal Interpolation

Dong Jian

The goal of optimal interpolation (OI) is to produce an initial condition for a subsequent numerical forecast through a statistical combination of observations and first guess (background). Here it is applied to analyze 2m air temperature and sea level pressure with SYNOP observations.

The essence of OI can be summarized as the formulas below:

$$\mathbf{x^a} = \mathbf{x^b} + \mathbf{K}\left(\mathbf{y} - \mathbf{Hx^b}\right) \tag{1}$$

The analysis $\mathbf{x^a}$ is obtained by adding to the first guess $\mathbf{x^b}$ the product of the optimal weight matrix $\mathbf{K}$ and the innovation $\mathbf{y} - \mathbf{Hx^b}$ (the difference between the observation and the first guess).

$$\mathbf{K} = \mathbf{BH^T}(\mathbf{HBH^T} + \mathbf{R})^{-1} \tag{2}$$

The optimal weight matrix $\mathbf{K}$ is given by the background error covariance in the observation space $\mathbf{BH^T}$ multiplied by the inverse of the sum of the background and the observation error covariances.

$$\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H^T}\mathbf{R}^{-1}\mathbf{H} \tag{3}$$

The analysis precision, defined as the inverse of the analysis error covariance, is the sum of the background precision and the observation precision projected onto the model space.

The report discusses the influences of several key parameters of OI to investigate its capabilities. In the provided code developed by Xiang-Yu Huang etc., the correlation scale is set to be 150 km for both pressure and temperature with default resolution, observation distribution, and one box structure. The results are shown below (Fig.1). The following discussion focuses on analysis increment which involves the terms in $\mathbf{K}\left(\mathbf{y} - \mathbf{Hx^b}\right)$.

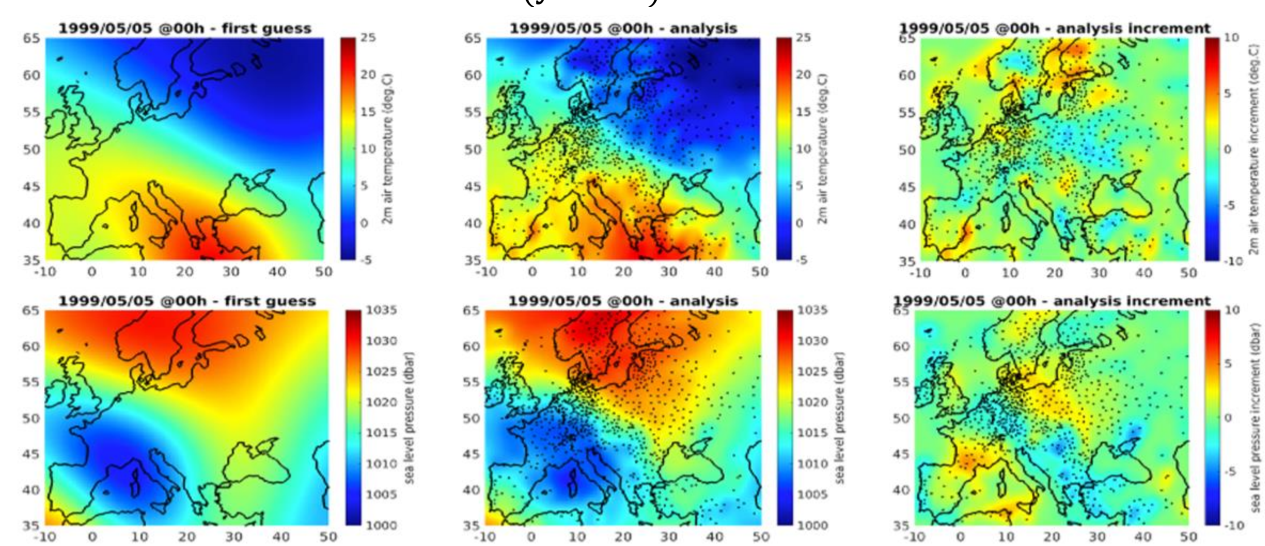

Figure 1 The first row is analysis of 2m air temperature in default experiment, first guess (left), analysis (middle) and analysis increment (right). The second row is the counterparts for sea level pressure. Black dots represent observations. Note $\boldsymbol{x^b}, \boldsymbol{x^a}, \boldsymbol{K}\left(\boldsymbol{y} - \boldsymbol{Hx^b}\right)$ *represent first guess, analysis and analysis increment respectively.*

1 Correlation Length L

The role of correlation length scale L comes into play in the background error covariance matrix **B**, which is often assumed homogeneous and isotropic, i.e., the correlation between two analysis points is a function of distance:

$$\mathbf{B} = \sigma_\mathrm{b}^2 \begin{pmatrix} 1 & \gamma_{12} & \cdots & \gamma_{1N} \\ \gamma_{21} & 1 & \cdots & \gamma_{2N} \\ \vdots & & \ddots & \vdots \\ \gamma_{N1} & \gamma_{N2} & \cdots & 1 \end{pmatrix}$$

γ_{ij} is a Gaussian function used to model the correlation between point *i* and point *j*:

$$\gamma_{ij} = \exp\left(-\left(r_{ij}/L\right)^2\right)$$

L is the correlation length scale that determines the distance over which the influence of the observation extends. r_{ij} is the distance between analysis points i and j, The resulting γ_{ij} is a value between 0 and 1, γ_{ii} will equal 1, since the distance from a point to itself r_{ii} is zero. As L gets larger, the γ_{ij} values will again approach 1, points separated by greater distances will contribute more to the overall weighting scheme **K** which can be understood by noting that the leftmost term in **K** is **B**.

The first exercise investigates the impact of the length scale by varying only the length scale in several experiments (see table.1).

Table 1 Experiments in exercise 1

Experiment	Correlation scale *L*
1	75 km
2	150 km (Default setting)
3	300 km
4	1500 km

For the instance of pressure plots (Fig.2), comparing with default setting (*L*=150 km), one notices smaller magnitude and less extend of analysis increment in experiment 1 (*L*=75 km), and greater magnitude and more extend of analysis increment in experiment 3 (*L*=300 km), especially over Poland and the border between Spain and France. However, over the Mediterranean Sea in experiment 4, the transition from data-dense region to data-sparse region is abrupt and unphysical, the left side of the Mediterranean Sea produces large positive increments while the right side produces large negative increments and the lower left bottom of the domain produces unexpected negative increment, due to too large length scale.

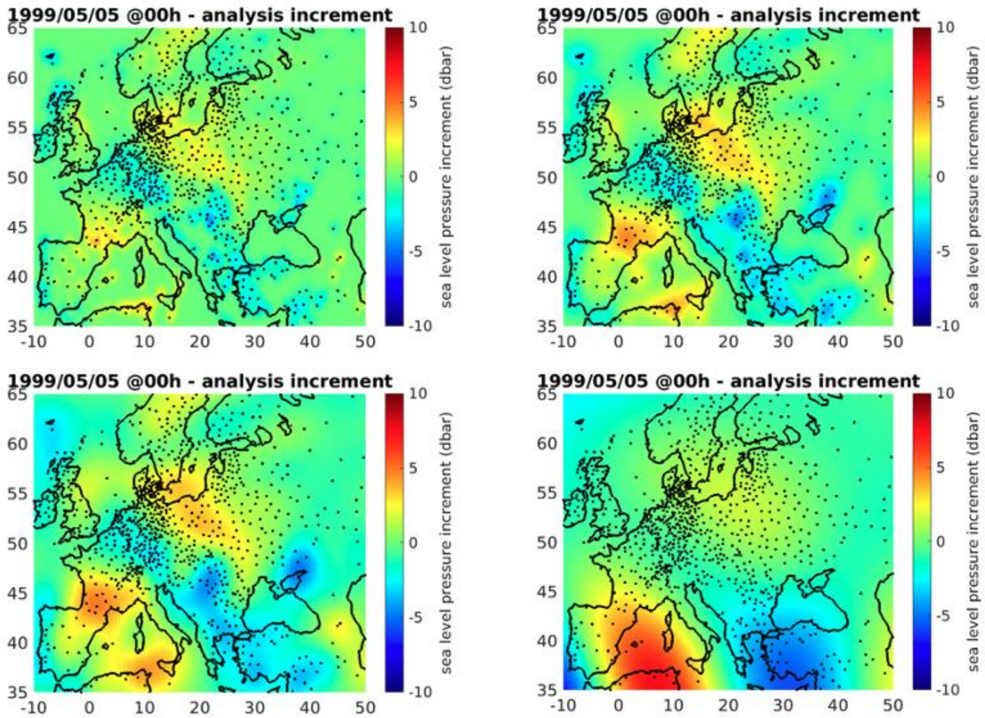

Figure 2 Analysis increment for sea level pressure in experiment1 (upper left, L=75 km), experiment 2 (upper right, L=150 km), experiment 3 (lower left, L=300 km), experiment 4 (lower right, L=1500 km)

If L is small, analysis increment contributes very little, analysis increment is comparatively localized. And it extends the influence range with larger L. If L is too large (experiment 4 L=1500 km), large magnitude of increment centered in the Mediterranean Sea (data-sparse areas, reddish and bluish patches) are not rational. The patterns would be more unphysical if L is further increased (Figure not shown).

Correlation scale L is the kernel of background error matrix $\mathbf{B}$ which influences information spreading. In data-sparse regions, the magnitude of the analysis increment is mostly determined by the covariance structures (for a single observation it is given by $\mathbf{BH^T}$). Hence the correlations in $\mathbf{B}$ will perform the spatial spreading of information from the observation points to a finite domain surrounding it.

L also influences information smoothing which can be noticed in data-dense regions, where analysis increments are comparatively smoother than the Mediterranean Sea.

In short, the larger the background error covariance, the larger the correction to the first guess. If background error variances are badly specified, which means either too short or too large length scale, it will lead to too short or too large or even unrealistic analysis increments.

2 Analysis domain resolution

Now, we use the averaged value of all observations as the background field and change domain resolution. Averaged observations are treated as the same observation, so state vector $\mathbf{x}^{\mathbf{b}}$ contains the same values and no need to consider the effects of **B** matrix if L is unchanged. Several experiments are conducted to investigate the effects of resolution for the same domain size and different length scales (see table.2).

Table 2 Experiments in exercise 2

Experiment	L (km)	Resolution
5	150	default resolution
6	300	default resolution
7	150	½ default resolution
8	300	½ default resolution
12	150	1/5 default resolution

Domain resolution refers to model discretization. The higher the resolution, the smaller the cell size, thus, the greater the details. Decreasing spatial resolution, those observations which are very close would be binned into one cell and they are treated as one single observation. So, the analysis increment is the same in each cell.

Comparing analysis increment with different resolutions (experiment 5,7&12, see Fig.3), the difference is obvious, coarser resolution has larger pixel cells, thus larger representativeness errors, the presence in the observations of subgrid-scale variability not represented in the grid-average values of the model and analysis.

One can say that a much coarser resolution will ignore some fine processes within grids. Because the spatial resolution of the data is reduced and that small-scale structures visible in the original data are not present in the binned observations. This involves the linearized observation operator **H** which interpolates from the model discretization to the observation points.

On the edges of grids, analysis increments are not smoothing and sort of zigzagging due to decreasing resolution, especially in the transition zone between high temperature and low temperature regions.

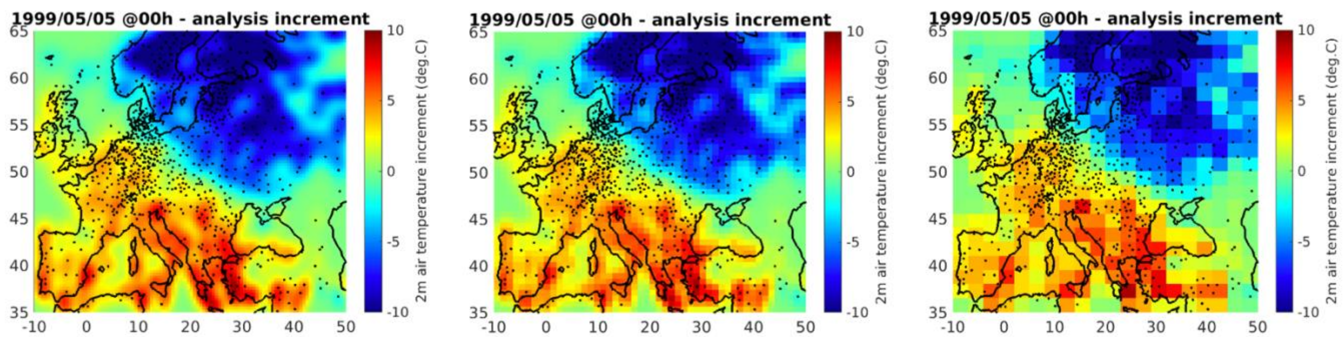

Figure 3 Analysis increment for 2m air temperature in experiment 5 (left, default resolution), experiment 7 (middle, half resolution) and experiment 11 (right, 1/5 default resolution)

Comparing analysis increments with different resolution and different length scales (experiment 5,6,7&8, see Fig.4), decreasing resolution means analysis grids length is increased, those observations within the same grids are regarded as one and thus contribute same to the observation error covariance matrix **R**. The analysis increments are sort of confined and averaged in each grid, causing unsmooth transition between grids.

Comparing experiments 7 and 8 (see Fig.4 second row), when combing the effects of increasing length scale produces a smoother analysis increment, which sort of mitigates the effects of decreasing resolution, yet analysis increment is a bit larger.

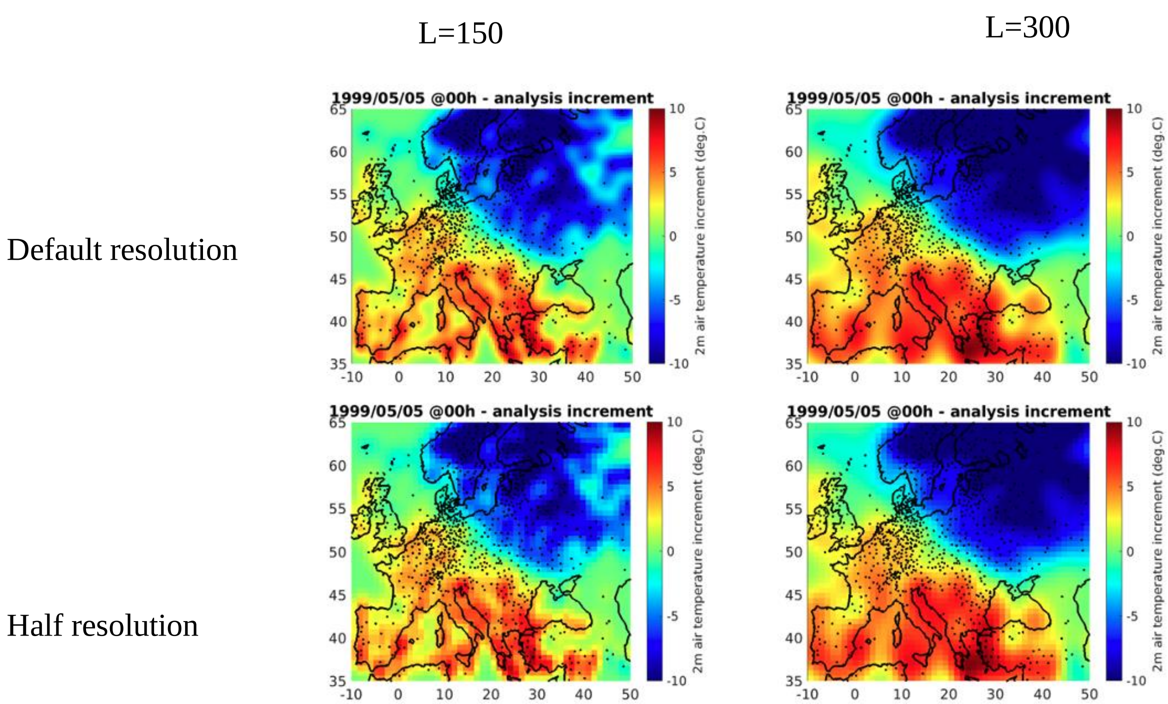

Figure 4 Analysis increment for 2m air temperature increment in experiment5 (upper left, default setting), experiment 6 (upper right, high resolution, L=300 km), experiment 7 (lower left, low resolution, L=150 km), experiment 8 (lower right, low resolution, L=300 km)

In general, a higher resolution is always better than low resolution if not considering computational capability. Lowing resolution will increase representativeness errors which indicates the observation is not representative of the regionally averaged measurement required by the model grid It is not negligible when analysis phenomena cannot be well represented in model space.

When one must use a coarser resolution, the correlation length scale should be increased appropriately, but not too large.

3 Number of observations

In most cases, the analysis problem is under-determined because data is parse. Next, the number of observations is further reduced in several experiments (table 3) to discuss the effect of having fewer observations.

Table 3 experiments in exercise 3

Experiment	Observations
1	1062, default setting
9	531, 50 % observations
10	Remove observations in France
11	Remove observations in France but *L* = 30 km

In temperature plots (Fig.5 first row), comparing default experiment, analysis increment in experiment 9 is reduced regionally, this can be seen over Finland, west coast of Norway, east of Spain. When observations in France are removed, analysis values in France remain the same value of first guess, analysis increment is zero in France, even though there might be analysis increment near the edge of France.

In pressure plots (Fig.5 second row), reduction of analysis increments can be seen in central Europe (yellow patches), the border between Spain and France (reddish patches), and some regions in east Europe (Blue patches). Analysis increment is void in France where no observations exist.

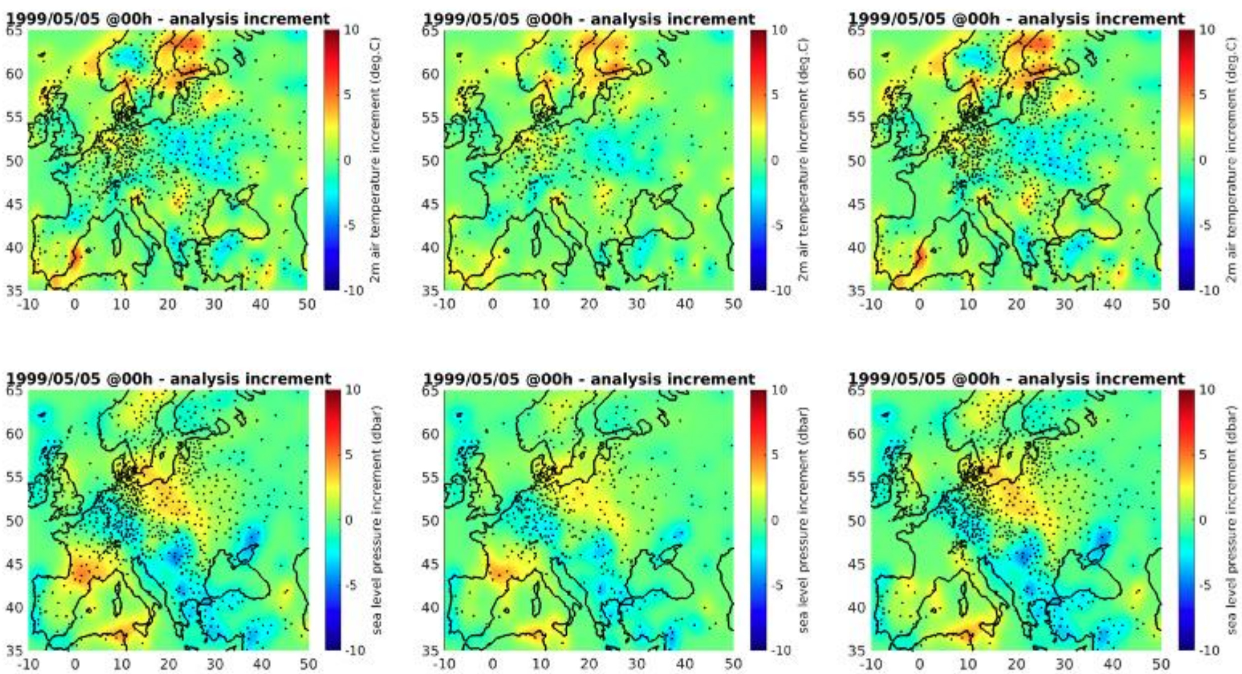

Figure 5 Analysis increment for 2m air temperature increment in experiment 1 (upper left, default setting), experiment 9 (upper middle, 50% observation),experiment 10 (upper right, removed observations in France) and sea level pressure increment in experiment 1 (bottom left, default setting), experiment 9 (bottom middle, 50% observation), experiment 10 (bottom right, removed observations in France)

Refer to the formula of analysis increment, The key to data analysis is the use of the discrepancies between observations and state vector given by $\mathbf{y} - \mathbf{H}\mathbf{x}^{\mathbf{b}}$. When the number of observations is reduced, the dimension of observation vector and observation departures are reduced, causing less analysis increments or say analysis tends to be suboptimal (too close to the background).

Comparing experiment 10 with 11 which analysis scale is increased (Fig.6), the missed information in France would be filled partially (we see a yellower color in the center of France in experiment 11) even though not as accurate as of that with full observations and it affects other regions as well. When the number of observations is quite a few, the correlation scale is increased, allowing to spread of the information more uniformly for the data-missing region.

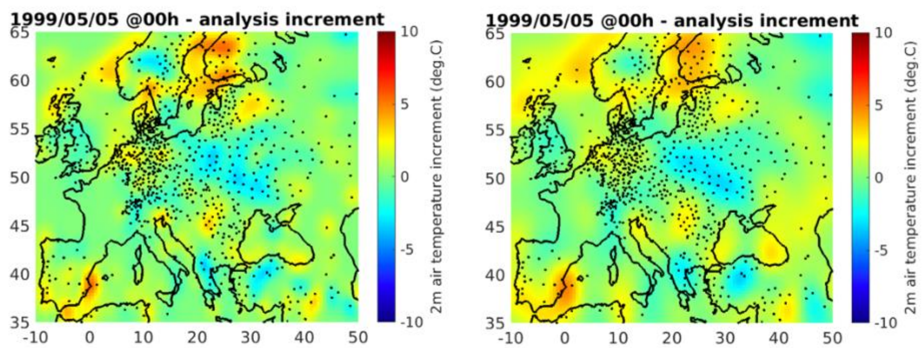

Figure 6 Analysis increment for temperature in experiment 10 (left, removed observations in France) and experiment 11 (right, removed observations in France but correlation scale =30 km)

In short, having more observations would result a better analysis increment. When data in a particular region is vacant, one may consider increasing correlation scale appropriately.

4 Box structure

Consider again the analysis equation (1), in practice mainly due to the large size of $\mathbf{HBH^T} + \mathbf{R}$, which is often difficult to invert. To produce a computationally feasible algorism, a box method has been used which divides the analysis domain into several boxes. The boxes are often made to overlap to have smooth transitions at the box borders so that most of the observations selected in two neighbouring analysis boxes are identical. In the end, the final analysis is obtained by putting all the box analyses together (Fig.7).

Two experiments are compared to investigate the effect of box structure (table 4). For the 9 boxes experiment, the same type of OI analysis is performed in each selected box, but with a reduced number of observations (Fig.8). Now the dimension of $\mathbf{HBH^T} + \mathbf{R}$ is equal to the number of observations in each box which is reduced and makes it feasible to invert it.

Table 4 experiments in exercise 4

Experiment	Box structure
12	1 box
13	9 boxes

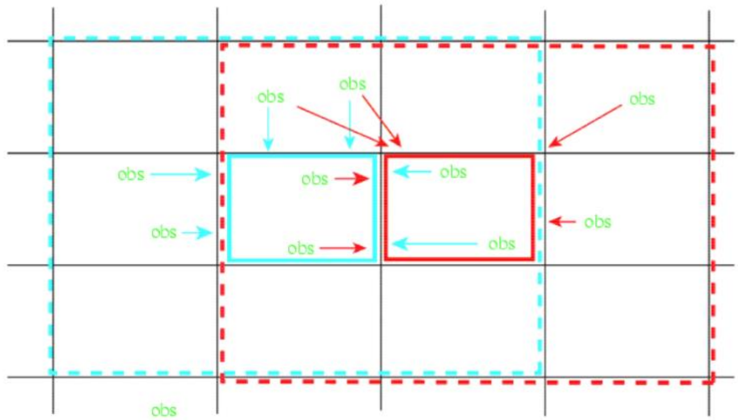

Figure 7 Sketch of box method, all the points in an analysis box (full rectangle), all observations located in a bigger selection box (dashed rectangle), so that most of the observations selected in two neighboring analysis boxes are identical. (credit: F. Bouttier and P. Courtier, ECMWF)

```
=======================================================
    nobs in all:     1062
    nobs in each box:    662  670  341
                         741  764  386
                         505  551  271
=======================================================
```

Figure 8 Snapshot of code showing number of observation in each box

It takes 1m19.113s to run the analysis for 9 boxes experiment, while 1m3.458s for the 1 box model, which means the box method is slightly more expensive. Although the box method spends a few seconds more, the box method is still computationally efficient because it makes the inversion of large dimension matrix feasible.

Analysis increment is shown below (Fig.9), box method produces greater analysis increment and extends more influence range, as can be seen over the north-east and south part of the domain in temperature figures. In pressure plots, analysis increment extends to the north-east of the domain for high-pressure system and extends to the Mediterranean Sea for low-pressure system.

In the 9-box settings, the dimension of the total analysis grid is set 21 by 21, 9 selected boxes with 11 by 11 dimension grid are filled in the total domain (similar with Fig.7), thus most of observations in two neighboring boxes are the same, which soothes the transitions between box borders, this can be noticed in the region between positive increment and negative increment.

Comparing the south-east corner box (fewer observations) with the north-west corner box (more observations), the box without enough observation is quite different with its counterpart without using box method. So, if fewer observations in the selected box, it may produce some uncertainties.

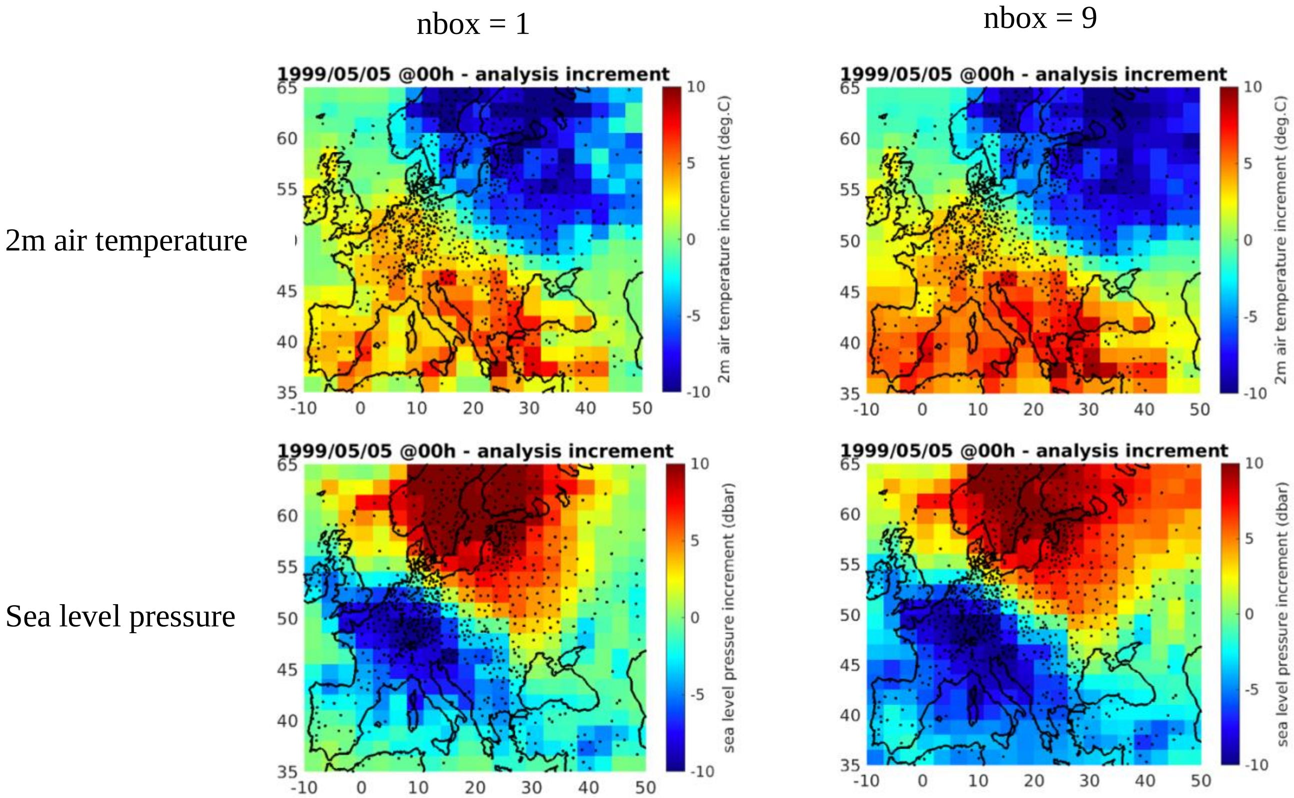

Figure 9 Analysis increment for 2m air temperature increment in experiment 12 (upper left, 9 boxes), experiment 13 (upper right, 1 box), and sea level pressure increment in experiment 12 (bottom left, 9 boxes), experiment 13 (bottom right, 1 box)

5 Conclusion

Four exercises are discussed to investigate the effects of correlation scale, domain resolution, observation distribution, and box method respectively. All these factors contribute to the quality of analysis adjustment.

Correlation length scale determines influence range of observations, resolution decides representativeness error and thus observation errors, number and distribution of observations influence the interpolation and departure between observation and state vector, box methods make computation feasible yet induce uncertainty in data-sparse box.

Essentially these parameters/observations have their respective influence on background error correlation $\mathbf{B}$, observation error correlation $\mathbf{R}$, and departure $\mathbf{y} - \mathbf{H}\mathbf{x}^{\mathbf{b}}$, influencing analysis increment. When the right parameters are chosen, appropriate assumptions are made, data quality is satisfied, optimal interpolation can produce a good and justified analysis result with its simplicity and relatively small cost. Otherwise, spurious noise may be produced in the analysis field.