REGIONAL SST PRODUCT DEVELOPMENT AND ASSIMILATION WITHIN THE MEDITERRANEAN FORECASTING SYSTEM

B. Buongiorno Nardelli[1], C.Fratianni[2], H.Roquet[3], R. Santoleri[1]
[1] Istituto di Scienze dell'Atmosfera e del Clima-CNR-Roma
[2] Istituto Nazionale di Geofisica e Vulcanologia (INGV)
[3] CMS-MeteoFrance

ABSTRACT

A regional Sea Surface Temperature product covering the Mediterranean Sea and the western Atlantic Ocean at 1/16° spatial resolution has been developed and is operationally produced and distributed in near-real time in the framework of the MFSTEP (Mediterranean Forecasting System Toward an Environmental Prediction) project. The MFS processing chain is described here. It includes several steps, from the cloud detection and removal, to the images compositing and merging of CNR-ISAC and CMS datasets. Night time maps of SST for each day are interpolated on the model domain through statistical methods and then assimilated in the MFS General Circulation Model. The data assimilation techniques and their impact on the modelling tasks are also briefly discussed.

1. INTRODUCTION

The Mediterranean Forecasting System Towards Environmental Prediction project (MFSTEP) aims at the further development of the Mediterranean Forecasting System Pilot Project (MFSPP) operational forecasting system for the Mediterranean Sea. It is based upon three main components: a) a Near Real Time (NRT) Observing system; b) a numerical forecasting systems at basin scale and for regional areas; c) a forecast dissemination/exploitation system.

The problems to be solved within MFSTEP are related to three major categories:
1) Technology development, connected to the new instrumentation and processing chains for NRT monitoring and the provision of NRT protocols for data dissemination, comprehensive of telecommunication technology and quality control procedures;

2) Scientific development, connected to the understanding of the sampling scheme for different measuring platforms, the design and implementation of data assimilation schemes for different spatial scales, the ecosystem modelling validation/calibration experiments at the basin and the coastal areas scale and the development of data assimilation techniques for biochemical data;

3) Exploitation development, consisting of software interfaces between forecast products and oil spill modelling, general contaminant dispersion models, re-locatable emergency systems, search and rescue models, and fish stock observing systems, together with studies of economic impact of the system.

In the framework of MFSTEP, an improved near real time acquisition and processing system of SST remote sensing data has been set up and tested for the Mediterranean sea. The system is based on existing facilities and could serve as a prototype also for future operational systems. Near real time (NRT) remote sensing data provide unique data sets for the setting up of the operational Mediterranean Forecasting System, as they are assimilated and used for the initialization of the basin scale forecasts. The assimilation cycle is weekly and the SST data assimilated are daily SST products optimally interpolated on the 1/16°x1/16° model grid.

In this paper the MFSTEP SST processing is described. In the following, the data used and the main modules of the chain will be introduced, together with a detailed description of the optimal interpolation algorithm. Finally, some preliminary results of the assimilation of daily SST data into an operational model are discussed.

2. THE IMPROVED MFS PROCESSING CHAIN

The MFS SST product is based on night-time AVHRR images acquired and processed by CMS and CNR-ISAC. The processing chain developed during MFSPP has been revised in order to allow a higher frequency assimilation of the sea surface temperature data in the MFSTEP model. Originally, MFSTEP technical annex indicated that NRT daily composite maps of SST should be computed leaving data voids in the cloud contaminated regions. In contrast, daily interpolated SST maps on the MFSPP grid (1/8°x1/8°) were expected to be available by the end of the project. However, the modelling component of MFSTEP successively indicated the need for completely filled SST fields also at the higher resolution 1/16°x1/16° grid. This in turn required a complete revision of the processing chain. In particular:

1) A new procedure to generate the data products was defined. It required moving the responsibility for MFSTEP SST data production, and distribution, from CMS to CNR-ISAC.
2) A new data exchange/merging procedure between CMS and CNR-ISAC was developed In particular, mapped AVHRR SST acquired and processed at CMS are now uploaded from CMS to CNR-ISAC, and merged to the AVHRR SST acquired and processed at CNR-ISAC.
3) A new Optimal Interpolation (OI) scheme was developed at CNR-ISAC.

3. MERGING OF CMS AND CNR-ISAC DATA

This module deals with the merging of declouded data produced at CMS and CNR-ISAC. Due to the different satellite visibility circles and the need for coverage extending from 18°W to the eastern limit of the Mediterranean Sea a composite of the data acquired at the two stations was mandatory to meet the project's geographic coverage requirements.

Different SST algorithms are used at CMS and CNR-ISAC (CMS algorithm and NASA Pathfinder algorithm (PF), respectively), as well as different cloud detection algorithms. It is thus necessary to eliminate the bias between data produced at the two stations, in order to avoid artificial features and lower quality estimates in the composite SST.

At present, the best algorithm for AVHRR SST estimate in the Mediterranean is considered the delayed time PF algorithm (Kearns et al. 2000, D'Ortenzio et al., 2000). PF-SST is obtained through a nonlinear SST algorithm that makes use of coefficients based upon comparison with PFMDS (Pathfinder Matchup Data Set). PF-SST coefficients have a monthly temporal resolution (based upon 5-month weighted average) and are calculated by the RSMAS Remote Sensing Group at the University of Miami. On the opposite with respect to fixed coefficients operational algorithms, the monthly resolution of PF coefficients affords sensitivity to the seasonally varying atmospheric concentration of water vapour, the major constituent that enters the atmospheric correction procedure. Moreover, these coefficients take into account major input of stratospheric dust that can be considered of transient nature (for a full description of the Pathfinder SST algorithm the reader is referred to Kilpatrick et al. 2001). A comparison of MFSPP operational SST fields with delayed time PF-SST data showed a bias error characterized by a clear seasonal behaviour both using CMS algorithm and previous CNR-ISAC operational algorithm (Buongiorno Nardelli et al., 2003).

As PF coefficients cannot obviously be available in NRT, tests have been made to check the possibility of using previous year's coefficients to remove part of this seasonal error. As a consequence of the positive results of these tests, few adaptations were made to allow a real-time usage of PFSST algorithm at CNR-ISAC. CMS data are consequently adjusted to CNR-ISAC data by adding the average difference, computed from all the cloud-free pixels present in both composite images.

4. QUALITY CONTROL AND FINAL CLOUD REMOVAL OF MERGED PRODUCT

This module deals with a visual check aimed at detecting any residual cloud contamination that has not been flagged by automated cloud detection procedures. This step is necessary as residual cloudy pixels could cause the optimal interpolation to spatially propagate the error (smearing) to a larger area, especially if the cloud-contaminated area is near an area that is cloud flagged (devoid of data). In other words, from the point of view of generating optimally interpolated fields, the final product quality is more adversely affected (the error is higher) if few bad data points are retained than if few good data points are discarded.

For each weekly set of images the operator displays on the screen the entire images and subjectively decides if there are residual points that are cloud contaminated. If so, the operator removes these points by graphically encircling them with a cutting tool and overwrites the input file.

5. OPTIMAL INTERPOLATION MODULE

The classical optimal interpolation method (Gandin, 1965, Bretherton et al., 1976) has been adapted to interpolate on MFSTEP grid AVHRR SST observations. Optimal interpolation method requires an a priori statistical knowledge on the covariance function of the signal to be mapped and of the data noise, and has been applied here following previous results by Santoleri et al. (1991) and Leonardi et al. (1994). The strategy adopted is detailed in the following.

5.1 Description of the interpolation scheme

A sub-optimal interpolation scheme has been implemented in the AVHRR operational processing chain to produce SST estimates on MFSTEP grid. The scheme is sub-optimal in the sense that the data used to interpolate at a certain time-space location are selected within a limited subdomain, close to the interpolation point (within a space-time influential radius), while the optimal scheme would require the use of all available data. This step is necessary as, otherwise, given the high number of measurements available, huge matrix operations (inversions...) would be involved, leading to long (if not unfeasible for operational purposes) computing time.

As the Mediterranean sea is characterized by the presence of many islands and peninsulas, the scheme drives a ‘multi-basin’ analysis to avoid data propagation across land from one sub-basin to the other. In practice, the optimal interpolation is run several times, one for each sub-basin, applying different masks. Six sub-basin have been identified: eastern Atlantic, Western Mediterranean basin, Tyrrenian sea, Adriatic sea, Levantine basin, Aegean sea. In order to avoid artefact effects at the border of the different regions, two different masks are used for each sub-basin, one identifying the interpolation points and one for the selection of the observations. This last includes buffer zones at the borders.

The single analysis scheme includes different modules and is detailed in the following:

- IMAGES SELECTION: only images within a temporal window of -10/+6 days with respect to the interpolation day (J) are included in the analysis. Consequently, the first day of the week (J_1) is built selecting the data between J_1-10 and J_1+6, while the last day (J_7) can use only data between J_7-10 and J_7.

- DATA PREPARATION: all selected images are checked for residual cloudy pixels. Cloud margins are firstly eroded, flagging all values within a distance of *m* pixels to a pixel already flagged as cloudy. The second step consists in rejecting the SST values lower than a minimum threshold SST value (that can be changed seasonally). The third step is the comparison to the closer (in time) analysis available, that is used as a reference only if the analysis error (as defined in the previous section) is lower than a fixed value. Data that differ from the reference field for more than a defined threshold (usually 2σ, where σ represents the average standard deviation between consecutive night images =.7 °C) are not included in the analysis. The mean bias between selected images estimated from co-located valid pixels is removed.

- INFLUENTIAL DATA SELECTION: for each grid point, observations are selected within a space-time influential radius. Actually, even if more data are available inside this influential bubble, only *n* observations can be effectively used because of computational time limitations. As a consequence, a strategy to remove most cross-correlated/least correlated (to the interpolation point) data is needed. If not, isolated observations that are only slightly less correlated to the interpolation point than more covered areas could be excluded from the analysis, while selected data would be too much cross-correlated, leading possibly to numerical problems in the inversion of the covariance matrix and surely to biased estimates with respect to the original data coverage. Different methodologies have been tried, that gave similar results. The method chosen here is based on sorting the data as a function of their correlation to the interpolation point. The most correlated observation is selected first, while all successive data are selected only if they are found along a new direction in the space-time (until *n* observations are found). This allows a more balanced coverage within the influential bubble, even selecting a small number of observations.

- INTERPOLATION: The covariance between the interpolation point and the selected observations and the covariance matrix of the observations are computed using the correlation function defined within Medspiration project. The covariance matrix is then inverted and the SST estimate is computed.

- OUTPUT: interpolated SST fields are written in netCDF format and transferred via FTP to the MFS web page. An example of an interpolated map is presented in Figure 1.

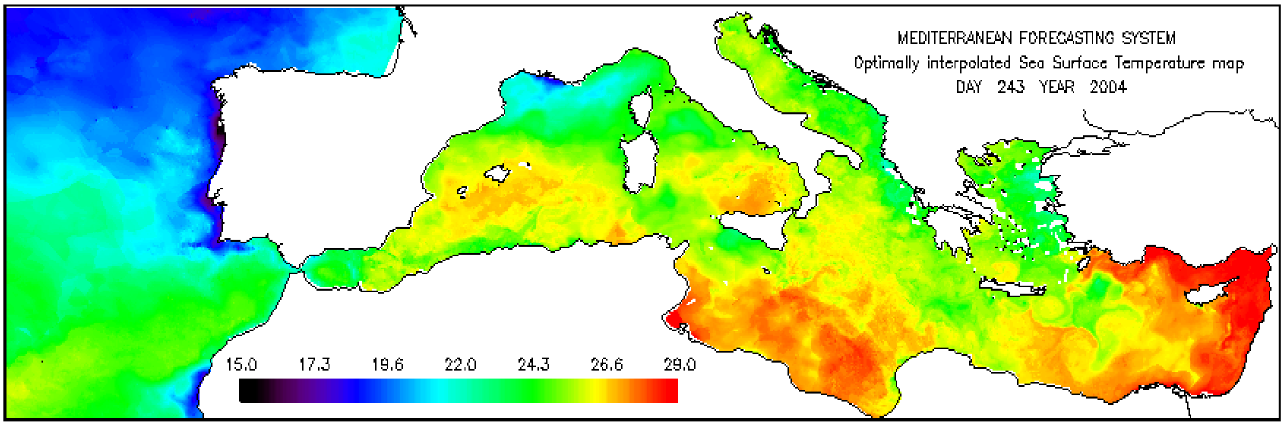

Figure 1. Example of an SST Optimally Interpolated map at 1/16° resolution.

5.2 Optimal interpolation configuration

Given the seasonal variability in the SST field and mainly in the cloud cover percentage and type over the Mediterranean area, the operating data interpolation scheme will inevitably need some changes being made to its configuration. These changes will affect mainly the data selection and the threshold values for data acceptance inside the interpolation procedure. The system design has consequently been chosen to be flexible, allowing for changes in the processing to be made simply by changes in a configuration file. The parameters defined in the configuration file are listed and described in Table 1.

FLAG_INT	**FLAG_INT=1** does not interpolate if a valid observed SST value is present at the interpolation time. FLAG_INT=0 interpolates anyway.
LIMIT	maximum number of selected data for each interpolation point **=50**
DIST	starting influential radius for data selection. **300 km**
RMAXDIST	maximum influential radius for data selection (the influential radius is incremented if data selected within DIST are less then LIMIT). **600km**
NPIX	number of values selected in time for each pixel. =**3**
THRESH	maximum difference admitted between the SST in input and the reference SST (used for residual cloudy pixels removal). The reference SST is either the corresponding day optimal field (for days before the interpolation day J) or the J-1

	analysis for the interpolation day (J) and successive days (J+1, J+2,...) **=1.2 °C**
RMAXERROPT	maximum % error admitted on the reference SST field to activate the residual cloudy pixels removal. **=40**
IBX	dimension of the moving window used for cloud erosion =**3**
MINSSTVALID	minimum SST value considered valid =**4**

Table 1. Configurable parameters in MFS OI scheme.

5.3 Covariance model

The correlation function has been estimated directly from 6 years of Pathfinder SST (1997-2002). If no signal is filtered from the time series, the seasonal cycle clearly dominates the variability. As seasonality is characterized by very long spatial and temporal scales, the interpolated field would be strongly smoothed, and higher resolution features would be filtered out. In order to reveal shorter scales, we removed the spatial average estimated from a first guess interpolated field from each image before estimating the covariance. This means that, before interpolating the data, we need to remove mean biases between all the images selected. In the analysis, biases are estimated from co-located valid pixels analogously to the *composite* procedure. The functional form that best fits the computed covariance is:

$$C(r,\Delta t) = e^{-\frac{\Delta t}{\tau}} e^{-\frac{r}{L}} \qquad (1)$$

where r is the relative distance, Δt is the time lag, L (=180 km) is the spatial decorrelation length and τ (7 days) is the temporal decorrelation scale.

6. ASSIMILATION OF DAILY SST DATA VS WEEKLY SST DURING MFS

In the MFSPP operational model (MOM with a horizontal resolution of 1/8 x 1/8 degrees) the SST is used to correct the air-sea fluxes at the surface of the model. The flux corrected heat fluxes are:

$$Q_{corr} = Q - \left.\frac{\partial Q}{\partial T}\right|_{T=T^*} (T - T^*) \qquad (2)$$

where Q is the net heat flux at the air-sea interface, T^* is the observed SST, T is the SST produced by the model. The relaxation coefficient $\left.\frac{\partial Q}{\partial T}\right|_{T=T^*}$ is taken to be constant and equal to 69 (W/m^2 ^{0}C). Due to the possibility of residual cloud pixels in the data, it was decided to use the flux correction only if the difference between observed and model SST is lower then 5° C. The value of the relaxation constant of 69 W/m^2 ^{0}C corresponds to a relaxation time of 7 days for the model surface layer thickness which is 10 meters (10m/7dy). We also used a relaxation time corresponding to 1 day with the same surface layer thickness (10m/1dy). A comparison has also been performed between XBT 5 meters temperatures and SST at the same points and background (i.e. model) for the run with daily SST and a relaxation coefficient of 10m/1dy (Fig. 2) and the run with weekly SST using a constant relaxation of 10m/7dy (Figure 3).

The daily SST run is clearly better than the run with weekly SST. Unfortunately, after a month of running with SST daily and the relaxation time of 1day, the heat flux field showed the development of small scale noise that could produce the model instability. This is partly due to the fact that the mixing scheme (constant vertical mixing value) is not sophisticated enough to be tightly coupled to daily SST.

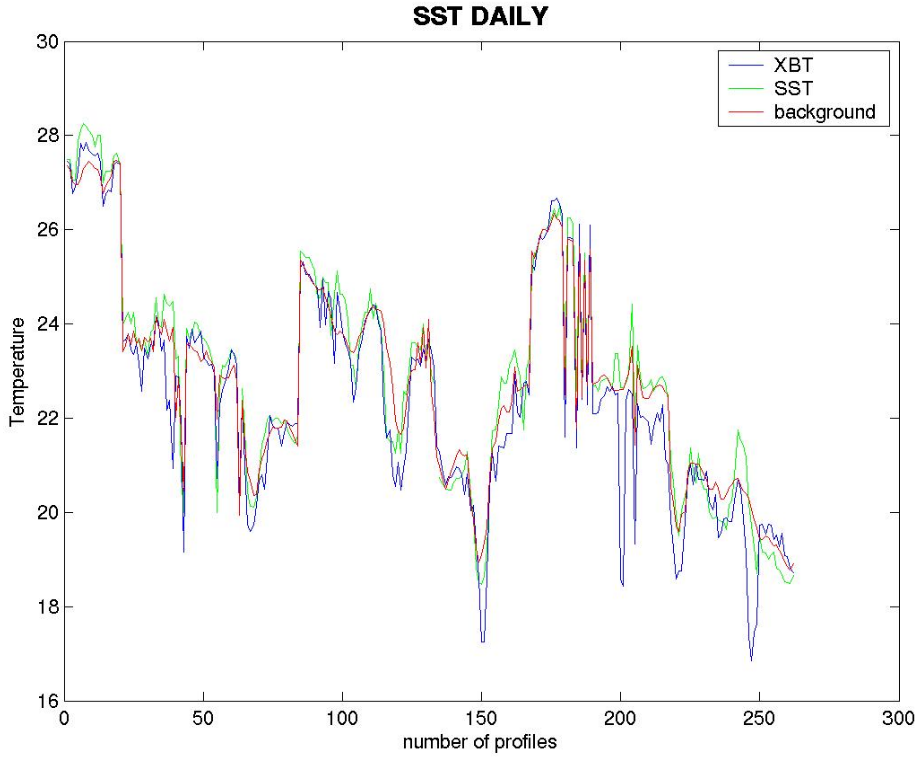

Figure 2. Comparison between XBT5m, SST daily analyses and SST background with 1d relaxation time.

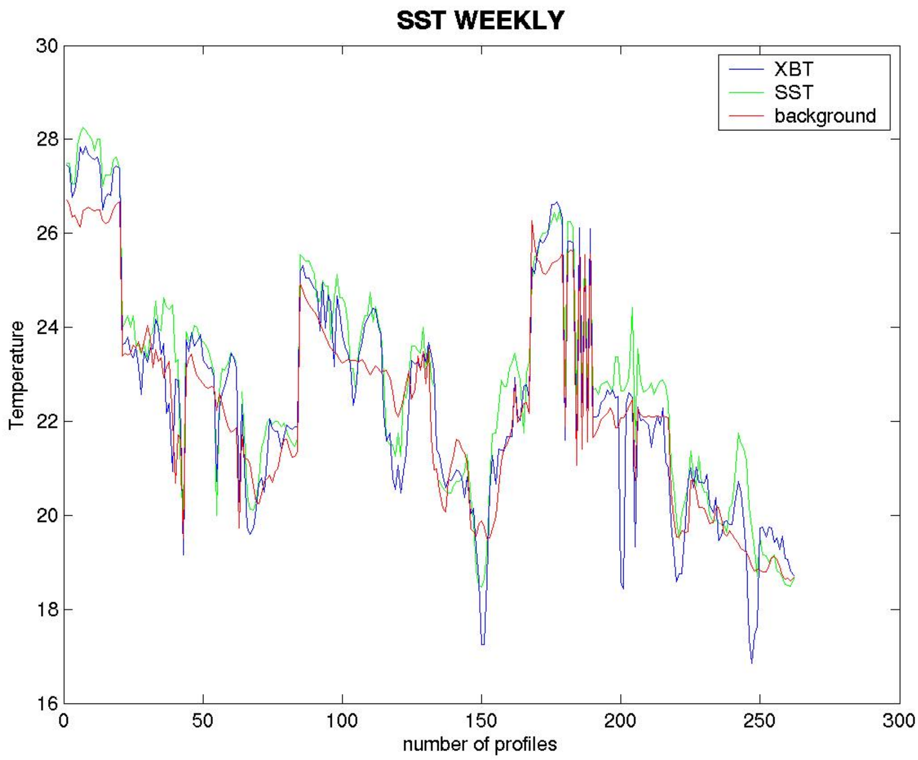

Figure 3. Comparison between XBT5m, SST weekly analysis and SST background with 7d relaxation time.

Thus we decided to use a constant relaxation of 10m/7dy also using the SST daily. The result of this run in terms of XBTs, SST and background is shown in Figure 4.

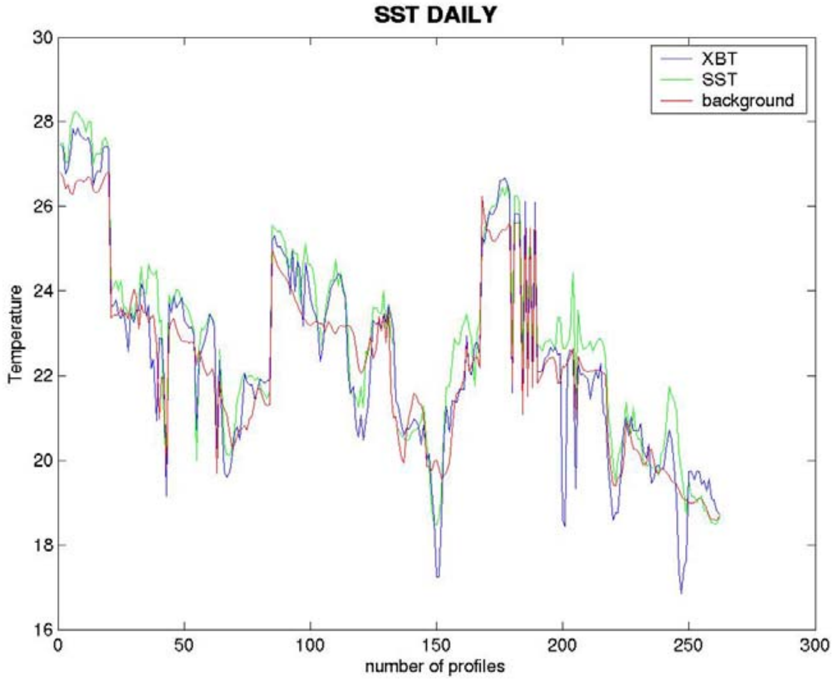

Figure 4. Comparison between XBT5m, SST daily analysis and SST background with 7d relaxation time.

Thus the benefit of having daily SST is not evident anymore, even if some minor improvements are evident.

7. Summary

The MFS SST processing chain has been described. The new SST product consists of a regional product covering the Mediterranean Sea and the western Atlantic Ocean at 1/16° spatial resolution. SST data are operationally produced and distributed in near-real time. Present SST data assimilation techniques used in MFS have been briefly described and their impact on the forecast performance has also been discussed.

REFERENCES

Buongiorno Nardelli B., G. Larnicol, E. D'Acunzo, R. Santoleri, S. Marullo e P.Y. Le Traon, 2003. Near Real Time SLA and SST products during 2-years of MFS pilot project: processing, analysis of the variability and of the coupled patterns, Annales Geophysicae, **21**, 103-121.

D'Ortenzio, F., S. Marullo, R. Santoleri, Validation of AVHRR Pathfinder SST's over the Mediterranean Sea, 2000: Geophy. Res. Let., **2**, 241-244.

Kearns, E.J.,J.A. Hanafin, R. H. Evans, P.J. Minnett, and O.B. Brown, 2000: An independent assessment of Pathfinder AVHRR sea surface temperature accuracy using the Marine-Atmosphere Emitted Radiation Interferometer, Bull. Meteorol. Soc., **81**, 1525-1536.

Kilpatrick, K.A., G.P. Podestà, and R. Evans, 2001: Overview of the NOAA/NASA advanced very high resolution radiometer Pathfinder algorithm for sea surface temperature and associated matchup database, J. Geophys. Res., **106**, 9179-9197.

Leonardi, R.M., Santoleri, R., Perilli, A.O., 1994. Shark research study on objective analysis, Final Report, European Space Agency, Contract Rep. No. 10180r93rYT-I- SC .

Santoleri R., S. Marullo, E. Böhm, 1991: "An objective analysis scheme for AVHRR imagery" Int. J. Remote Sensing, vol.**12**, no.4, 681-693.