WORK ON 3-I AT KNMI

G.J. Prangsma
KNMI
De Bilt, The Netherlands

1. INTRODUCTION

At KNMI a development program is in progress aiming at automatic production of short-range (24-30 hours) forecasts. Based on time constraints for the availability of the computed products to the operational forecaster, the goal has been set for a three-hourly update-cycle.

The fine-mesh, limited-area model (FMLAM) used as the central information carrier and forecast model is the limited-area version of the ECMWF grid point model. As a result of the design-studies it was deemed essential to use TOVS data for the three-hourly update-cycle to be realistic. Hence, a choice had to be made concerning a TOVS processing scheme compatible with the overall availability criteria.

Some of the considerations determining this choice were found to be less trivial than they seemingly were at first sight. Thus the basic characteristics as found in available reports and literature have been compiled and are reproduced in section 2 of this paper.

In the same spirit the, again seemingly trivial, properties of TOVS retrievals, relevant to impact studies of these retrievals in NWP models are outlined in section 3.

The current status of the TOVS implementation at KNMI and the accepted projects are described in section 4. Prospects and further outlook are discussed in section 5.

2. CHARACTERISTICS OF RETRIEVAL SCHEMES

All methods for atmospheric profile retrieval from satellite-based radiation measurements are based on the inversion of the integral equation describing the radiative transfer of the atmosphere for a number of wavelengths (channels) in both the microwave and the infrared part of the electromagnetic spectrum.

I The form of this integral equation implies that the radiation depends on the temperature profile in a highly non-linear fashion, whereas the sensitivity to variations in water vapour content, ozone concentration and other - minor - atmospheric constituents is near-linear and their effects are 1 or more orders of magnitude smaller than the changes due to temperature variations. Furthermore some channels are almost insensitive to water vapour, ozone and other trace gases.

This fact of nature delineates the first difference between retrieval methods:

a. those that try to assess all the atmospheric parameters (temperature and humidity profile, ozone concentration and possibly others) in one single process;
b. those methods which aim as a first step at the optimal retrieval of the temperature profile using only those channels that are nearly insensitive to water vapour etc. and as a second step use temperature-compensated radiance data of the sensitive channels to assess the other parameters.

II The next basic problem which retrieval schemes have to address is the problem of clouds. It is well-known that for otherwise identical atmospheric profiles - especially in the infra-red part of the electromagnetic spectrum - radiances are essentially different in clear and cloudy areas.

This process of so-called cloud-clearing, in which the observed radiances are corrected for the effects of couds, has been solved in a variety of ways. The main differences can be described by the two major schemes known at the time:

c. the so-called N-star method, which is based on the following - implicit - assumptions: clouds occur in one single layer only and: cloud top temperatures are locally uniform. The N-star method then tries to establish clear-sky radiances for a given observational point by substituting with averaged radiances from neighbouring points. This method fails if the surroundings are cloudy as well and will give unreliable results in meteorologically active areas with - relatively- large gradients;
d. the so-called Ψ-method ("pseudo-channel" method) employs the fact that some - but not all - microwave channels used in atmospheric sounding, are rather insensitive to the presence of clouds. By comparing the differences in radiance between microwave and infrared channels with similar weighting functions* "clear-sky" radiances are obtained in the following process: based on the microwave radiances and some infrared channels, which sense basically the upper troposphere and the stratosphere, a first initialization profile is obtained. The differences between the observed radiances in these insensitive channels and their counterparts corresponding to the obtained initialization profile together with the computed radiances from this initial profile for the cloud-sensitive channels yield the required "cloud-cleared" radiances.

III A third difference between current retrieval methods is the way in which differences caused by the viewing angle (the angle between the ray-path through the atmosphere and the nadir view of the satellite) are treated:

---

* Weighting functions in radiative transfer computations measure the relative contributions of atmospheric layers to the overall radiance observed at satellite altitudes.

e. One group of methods - diminishing in number - applies the retrieval algorithm on radiances which have first been corrected for this so-called limb-effect, using correction-coefficients obtained through curve-fitting to a theoretical radiative transfer model.
In so doing the residual errors in this curve-fitting process are introduced in an unknown way into the input data for the retrieval scheme, on top of the model errors.

f. A second group takes the viewing angle into account during the retrieval step, so that only the errors in the radiative transfer model are implied in the final result.

IV The actual inversion of the integral equation of radiative transfer starts by selecting a first-guess profile in the - in principle iterative - inversion scheme.
Several choices for this first-guess profile have been tried and are currently used:

g. The so-called "regression-first-guess" describes a method by which a number of recent radiosonde-observations are combined with co-located clear-sky satellite radiances to give a set of empirical regression coefficients which produce a temperature/humidity profile given the observed radiance data. This scheme of using empirical regressions implies that observed radiances are not only cloud-cleared, but also corrected for all other factors influencing the radiative transfer: limb (angle-of-view), surface elevation surface emissivity, water vapour effects in window channels etc., before being converted into initial guess temperature/humidity profiles. These regressed temperature/humidity profiles are used as first-guess data in the inversion scheme. In practice, the empirical regression coefficients are updated on a weekly basis (NOAA/NESDIS).

h. First-guess profiles can be taken from local (or regionally averaged) climatological data. In this method climatological mean temperature/humidity profiles, with possible refinements by taking seasonal or monthly mean values, are used to start the iterative inversion scheme.

i. In a number of implementations, temperature/humidity profiles computed by numerical weather prediction models, have been used to initialize the inversion iterations. In those areas where the forecasts are good, the initial profiles as well as the retrievals will be very near to the truth, but will not add much information to the model output. On the other hand, in areas where forecasts are bad, the initial guess profile will be far from the truth, so there is little hope that the final retrieved profile will be close to reality, in which case it will add wrong information to the next model analysis.
This phenomenon is commonly known as the "incest-problem": good forecasts are hard to beat, wrong forecasts tend to generate bad retrievals.

j. The most elaborate initialization scheme currently implemented takes an entirely different approach. In this scheme the initial temperature profile (ignoring humidity influences at first instance) is taken from a carefully selected set of world-wide observed radiosonde profiles. This set has been constructed in such a way that "neighbouring" profiles differ by about 3K vertically averaged over the entire atmosphere (and an equivalent amount in humidity).

The choice of the initial profile is made on the basis of observed or cloud-cleared radiances, being compared with the radiances produced by every single temperature profile, and any ancillary data such as estimates for the surface and tropopause temperatures (obtained through physical regressions over the entire data-set of historical radiosondes) and possibly forecast-data in the lower atmospheric layers (1000-700 hPa).

Current satellite-sounder inversion schemes, either for research purposes or for operational use, encompass the following three methods:

A. The operational NOAA/NESDIS scheme used to produce the so-called SATEM messages. This method uses the following ingredients:

I.a Assessment of all atmospheric parameters in one single process.

II.c N-star cloud-clearing. If this method gives results which do not pass a number of quality control tests (i.e. if the surroundings are also cloudy) a so-called "microwave only" retrieval is attempted, giving essentially different results with far less vertical and horizontal resolution.

III.e All radiances are limb-corrected (i.e. corrected for viewing angle) before being used.

IV.g Empirical "regression-first-guess" profiles are used.

B. The methods based on the ITPP (International TOVS Processing Package, distributed by the Univ. of Wisconsin), depending on release number and individual implementation differences, use:

I.a Single step retrieval of all atmospheric parameters.

II.c N-star cloud clearing, with the option of microwave-only retrieval if quality control tests indicate failure of the N-star method.

III.e Limb-corrected radiances are used, but with the option of using straight radiance data at the expense of computational efficiency (through the way of III.f method is implemented).

IV.g, h or i I.e., first-guess profiles derived from regression coefficients, climatology or model output.

C. The 3I method (available from the Laboratoire de Météorologie Dynamique at Palaiseau, France) employs the following ingredients:

I.b Temperature retrieval is done separately from humidity and other parameters. Those parameters are retrieved in a second phase.

II.d Ψ-method cloud-clearing combining microwave and infrared data and initialization to obtain cloud-cleared radiances.

III.f The viewing angle of the satellite is taken into account.

IV.j First-guess profile selected among observed historical radiosonde profiles (so-called TIGR** data base), aided by ancillary data. Note that in the cloudy cases this selection step is repeated after the application of the Ψ-method cloud clearing, which is computationally efficient because of the use of pre-computed radiances in the TIGR data set.

Further remarks concern the computer-time involved in each of these methods. Both the NOAA/NESDIS and the ITPP-based schemes require the evaluation of the radiative transfer equation for every spectral channel, every data spot in every iteration step. These calculations can be (and are) accelerated by applying simpler models for atmospheric radiative transfer at the expense of accuracy of the computed radiances.

In the 3I scheme, which requires only one single iteration step, avoids these radiative transfer computations all together, by employing a pre-computed database, containing the radiances for all channels, all the historical temperature/humidity profiles, all viewing angles and all combinations of surface conditions.

It should be noted that recent developments in the analysis-phase of numerical weather prediction models take an entirely different approach in using radiance data. In this domain so-called "adjoint-operator" schemes are under development which - among other things - aim at the direct inclusion of radiance data in the analysis-phase, without a separate inversion scheme.

These adjoint methods require the use of cloud-free radiances, corrected for viewing angle, surface emissivity and surface elevation, and the development of radiative transfer models - highly non-linear in temperature as noted above - within the necessary linearity-assumptions of the adjoint-operator techniques. At the moment of writing the outcome of this research is unknown, and the eventual success of such methods is considered to be speculative as far as radiance data inclusion is concerned, especially when clouds are involved, for which the radiative transfer calculations are far beyond the present state of the art.

3. CONSIDERATIONS FOR IMPACT STUDIES

Based on the characteristics of retrieval schemes as discussed above, some implications for the use of TOVS data in NWP schemes can be identified.

** TIGR: TOVS Initial Guess Retrieval data set, containing historical atmospheric profiles along with the corresponding radiances.

We therefore start by recalling the nature of satellite-based radiance measurements:

- The TOVS data per data-spot are limited in number: currently 20 HIRS channels and 4 MSU channels.

  Of these 20 HIRS channels, one is in the visible wavelengths, giving albedo, one is in the ozone band (sensing total ozone content). Three channels are in window regions giving essentially surface (or cloud top) temperature and three channels are sensing water vapour. This leaves 12 channels for temperature sounding, of which 4 are dimensioned to observing thick stratospheric layers. Of the 4 MSU channels one senses surface emissivity and one concentrates on the stratosphere.

  In summary, we have 10 (8 HIRS + 2 MSU) pieces of tropospheric information and 5 (4 HIRS + 1 MSU) pieces of information for the stratosphere.

  With the advent of new instruments these numbers will vary, but not by an order of magnitude.
- By the very nature of the observational techniques involved, all TOVS radiances apply to a finite volume of air, typically 1-2 km thick in the troposphere and up to around 5 km in the stratosphere and a few tens to hundreds of kilometers across (20-60 km for HIRS and 110-320 km for MSU).

  I.e., radiance data as obtained by the TOVS instruments sense bulk, thermodynamic properties of the atmosphere: internal energy (i.e. layer thicknesses), mixing ratios for water vapour and total amounts of minor constituents like ozone.

These properties of the TOVS observations do have implications for the way they can be used in NWP schemes. For the dynamical parts of the models this translates to the availability of thicknesses of atmospheric layers (typically around 8 for the troposphere and 4 for the stratosphere) as inputs to the analysis phase.
For the model description of water vapour and liquid water/clouds the model analysis can draw on a clear/cloud indicator and mixing ratio/relative humidity information for around 3 (relatively thick) tropospheric layers.

The precise number of independent data items varies slightly with the details of the retrieval method, but the numbers indicated above can be considered around optimal with the current TOVS instrumentation.
Once again: TOVS data cannot give the vertical detail of modern radiosondes, but can give considerably better defined horizontal gradients, especially in cloud-free areas.

Many state-of-the-art analysis schemes use the optimum interpolation technique. An essential element of this technique is the so-called error-correlation matrix, describing how errors in the observational data are propagated in space (grid-points) both horizontally and vertically.

Such error-correlation matrices are intimately linked to the implementation details of the weather prediction model used and can hardly be derived from first principles. Some dependence on the inversion scheme used could be anticipated, bearing in mind how exactly the infrared and microwave observations are combined. The clear/cloudy indicator obtained in the retrieval scheme might turn out to give a rough idea on the quality of the observed/derived thickness fields. But very little experience has been gained supporting or negating this idea. In this context the limited experience with the assimilation of TOVS data in NWP models is reflected in the relative ignorance on their error-correlation characteristics. Although some rough indications on compensating effects in the vertical and on reliable horizontal gradient information has been obtained, suggesting negative error-correlations between vertically neighbouring thickness values and small correlations horizontally, final estimates for the error correlation matrices can be obtained only from a routine use of the TOVS data.

Based on the characteristics of the current retrieval techniques a relatively weak dependence of these matrices on the retrieval method is probable but the major differences are expected to be directly linked with the models' implementation details.

Now let us address the question of where and how we apriori can expect impact of TOVS data in any NWP scheme.
Let us presume that - based on the above considerations - thickness values (or their horizontal derivatives: thermal winds) are assimilated into the model-analysis for 7-8 tropospheric and possibly 3-4 stratospheric layers.
Where do we apriori expect to find impact of TOVS data in the results of the NWP scheme?
A detailed answer to this question has to address "impacts" at various levels in the model system.

3.1 Analysis

Setting aside systematic biases (being reduced by assimilating horizontal gradients), which have to be eliminated by improvements in the retrieval schemes, an impact of satellite observations can only be expected apriori under two types of situations:

- in data-sparse areas, where grossly speaking: "any information is better than no information at all". Therefore it is hardly surprising that positive effects are found in global (as well as regional) models over the southern hemisphere;
- in data-rich areas when mesoscale structure cannot be adequately resolved by the traditional observation-network. This type of impact will therefore only be observed in high-resolution models over time-periods in the range of lifetimes of mesoscale weather system: up to 1 or 2 days at most, and only then when mesoscale structures are present. Of course, this statement assumes that the retrieval technique used, is capable of delineating the

mesoscale gradients, which is a demanding requirement on retrieval schemes in terms of horizontal resolution.

This type of impact is thus expected to be found over e.g. the North East Atlantic ocean and its coastal areas.

3.2 Model initialization

The next step in the NWP process involves the initialization of the model variables in conformity with the analysis just obtained.

Impact of satellite observations in this phase is strongly model-dependent and great care has to be taken not to destroy (by any smoothing mechanisms) small and meso-scale structure, present in the observations and the analysis, when suppressing e.g. gravity waves.

A number of suggestions can be found in recent research plans (e.g. HIRLAM-2) aiming at the parameterization and/or initializaion of e.g. diabatic processes in combination with cloud observations. This is indicative of hitherto unexplored potential use of satellite observations beyond the pure sounding type data. Whether or not this will aid the model initialization step remains uncertain until this research has been done.

3.3 Forecast

What has been said above under initialization applies afortiori for the actual forecast phase: strong model dependence and a complicated relation between the (initialized) analysis signals and their consequences propagated towards the final prognosis.

Here again, positive signals have been observed, as expected, over the southern hemisphere, but over the northern hemisphere no, or even negative impacts have been obtained in global forecast models. With high-resolution regional models for 1-2 days ahead) up till now little experience allowing conclusions, has been gained.

In summary, it is concluded that impact studies, after the required error-correlation statistics have been obtained, especially for mesoscale weather situations in the northern hemisphere are welcomed as a means to assess the usefulness of TOVS data in state-of-the-art NWP schemes. Such impact studies preferably should be designed in such a way, that the propagation of atmospheric signals in the TOVS observations through the various NWP steps can be separately monitored and quantified. Satellite-only maps, designed to show the presence or absence of mesoscale weather systems may turn out to be a sensible baseline in such experiments.

4. CURRENT TOVS WORK AT KNMI

4.1 Inversion scheme

Considering the time requirements for the automatic forecasting system (section 1), and the characteristics of retrieval schemes as outlined in section 2, we have decided as a first step to implement the 3I retrieval system developed at the Laboratoire de Météorologie Dynamique in Palaiseau, France. For the preprocessing of the real-time raw radiance data, received locally, we have developed a number of programs using the relevant ITPP programs developed at and distributed by the University of Wisconsin as templates. At the time of writing the whole chain except for the inversion program proper, is running autonomously under control of orbital information extracted from the daily TBUS messages received over the GTS.

The workload is distributed over two VAX type computers: a MicroVax II for reception of the HRPT data stream, the extraction of the low bit-rate TIP data and processing of AVHRR data and a VAX 3200 for preprocessing of TOVS data, the inversion and postprocessing and transmission of the results to the CONVEX array processor where the FMLAM is implemented.

We expect to have the full program suite to be pre-operational by the end of June 1989.

4.2 Impact studies

With the considerations on impact studies (section 3) in mind, we will perform a series of parallel runs with the FMLAM with and without TOVS data being available as input. These parallel runs will be done during a couple of months in late 1989. The impact will be measured in two ways: quantitatively comparing the analyses and the forecasts objectively and qualitatively by assessing the usefulness of the analyses and forecasts subjectively by experienced operational forecasters.

The results of these impact studies will be reported in spring 1990. Preliminary ideas exist for tests with a "satellite-only" model, combining TOVS data and surface observations as input to the forecast system. Whether or not this line of approach will materialise, depends on the outcome of the above impact study.

5. FURTHER OUTLOOK

For the future we envisage research on those physical parameterizations in the FMLAM and later HIRLAM models which have potential to be initialized from satellite data.

On the one hand this implies research on enhancements and extensions of the 3I inversion scheme, drawing on the imagery expertise in our institute, in close cooperation with Chedin and co-workers at LMD. On the other hand research is envisaged on liquid water and cloud representations for use in HIRLAM-type models, both for forecasting and - later on - for climate research.

Detailed plans for this research will be developed over the next twelve months. Finally ideas are being developed around the question how close the initial guess profile must be to the real atmospheric conditions in order to find correct retrievals. The outcome of this research can be applicable in two ways:

- it can be used in the selection step of the 3I algorithm, when making a choice between "close" but differing (in the meteorological sense) profiles
- it can shed light on the problems to be encountered in direct radiance assimilation schemes when distinguishing between possible but different solutions.

In both cases results on the error structure around the "exact" solution are anticipated.