Consortium for Small-Scale Modelling

A Description of the

Nonhydrostatic Regional COSMO-Model

Part III

Data Assimilation

C. Schraff and R. Hess

COSMO 6.00 September 2021

www.cosmo-model.org

DOI: 10.5676/DWD_pub/nwv/cosmo-doc_6.00_III;

Publisher
Deutscher Wetterdienst
Business Area "Research and Development"
Frankfurter Straße 135
63067 Offenbach
www.dwd.de

Editors
Ulrich Schättler, DWD
Ulrich.Schaettler@dwd.de

Contents

Section 1

Overview on the Model System

1.1 General Remarks

The *COSMO-Model* is a nonhydrostatic limited-area atmospheric prediction model. It has been designed for both operational numerical weather prediction (NWP) and various scientific applications on the meso-β and meso-γ scale. The COSMO-Model is based on the primitive thermo-hydrodynamical equations describing compressible flow in a moist atmosphere. The model equations are formulated in rotated geographical coordinates and a generalized terrain following height coordinate. A variety of physical processes are taken into account by parameterization schemes.

Besides the forecast model itself, a number of additional components such as data assimilation, interpolation of boundary conditions from a driving host model, and postprocessing utilities are required to run the model in NWP-mode, climate mode or for case studies. The purpose of the *Description of the Nonhydrostatic Regional COSMO-Model* is to provide a comprehensive documentation of all components of the system and to inform the user about code access and how to install, compile, configure and run the model.

The basic version of the COSMO-Model (formerly known as *Lokal Modell (LM)*) has been developed at the *Deutscher Wetterdienst* (DWD). The COSMO-Model and the triangular mesh global gridpoint model ICON form – together with the corresponding data assimilation schemes – the NWP-system at DWD. The subsequent developments related to the COSMO-Model have been organized within COSMO, the *Consortium for Small-Scale Modeling*. COSMO aims at the improvement, maintenance and operational application of a nonhydrostatic limited-area modeling system, which is now consequently called the COSMO-Model. The meteorological services participating to COSMO at present are listed in Table 1.1.

For more information about COSMO, we refer to the web-site at `www.cosmo-model.org`.

The COSMO-Model is available free of charge for scientific and educational purposes, especially for cooperational projects with COSMO members. However, all users are required to sign an agreement with a COSMO national meteorological service and to respect certain conditions and restrictions on code usage. For questions concerning the request and the agreement, please contact the chairman of the COSMO Steering Committee. In the case of a planned operational or commercial use of the COSMO-Model package, special regulations

Table 1.1: COSMO: Participating Meteorological Services

DWD	Deutscher Wetterdienst, Offenbach, Germany
MeteoSwiss	Meteo-Schweiz, Zürich, Switzerland
ITAF-ReMet	Ufficio Generale Spazio Aero e Meteorologia, Roma, Italy
HNMS	Hellenic National Meteorological Service, Athens, Greece
IMGW	Institute of Meteorology and Water Management, Warsaw, Poland
ARPA-SIMC	Agenzia Regionale per la Protezione Ambientale dellÂŕ Emilia-Romagna Servizio Idro Meteo Clima Bologna, Italy
ARPA-Piemonte	Agenzia Regionale per la Protezione Ambientale, Piemonte, Italy
CIRA	Centro Italiano Ricerche Aerospaziali, Italy
ZGeoBW	Zentrum für Geoinformationswesen der Bundeswehr, Euskirchen, Germany
NMA	National Meteorological Administration, Bukarest, Romania
RosHydroMet	Hydrometeorological Centre of Russia, Moscow, Russia
IMS	Israel Meteorological Service, Bet-Dagan, Israel

will apply.

The further development of the modeling system within COSMO is organized in Working Groups which cover the main research and development activities: data assimilation, numerical aspects, upper air physical aspects, soil and surface physics aspects, interpretation and applications, verification and case studies, reference version and implementation and predictability and ensemble methods. In 2005, the COSMO Steering Committee decided to define *Priority Projects* with the goal to focus the scientific activities of the COSMO community on some few key issues and support the permanent improvement of the model. For contacting the Working Group Coordinators or members of the Working Groups or Priority Projects, please refer to the COSMO web-site.

The COSMO meteorological services are not equipped to provide extensive support to external users of the model. If technical problems occur with the installation of the model system or with basic questions how to run the model, questions could be directed via email to `cosmo-support@cosmo-model.org`. If further problems occur, please contact the members of an appropriate Working Group. We try to assist you as well as possible.

The authors of this document recognize that typographical and other errors as well as dis-

crepancies in the code and deficiencies regarding the completeness may be present, and your assistance in correcting them is appreciated. All comments and suggestions for improvement or corrections of the documentation and the model code are welcome and may be directed to the authors.

1.2 Basic Model Design and Features

The nonhydrostatic fully compressible COSMO-Model has been developed to meet high-resolution regional forecast requirements of weather services and to provide a flexible tool for various scientific applications on a broad range of spatial scales. When starting with the development of the COSMO-Model, many NWP-models operated on hydrostatic scales of motion with grid spacings down to about 10 km and thus lacked the spatial resolution required to explicitly capture small-scale severe weather events. The COSMO-Model has been designed for meso-β and meso-γ scales where nonhydrostatic effects begin to play an essential role in the evolution of atmospheric flows.

By employing 1 to 3 km grid spacing for operational forecasts over a large domain, it is expected that deep moist convection and the associated feedback mechanisms to the larger scales of motion can be explicitly resolved. Meso-γ scale NWP-models thus have the principle potential to overcome the shortcomings resulting from the application of parameterized convection in current coarse-grid hydrostatic models. In addition, the impact of topography on the organization of penetrative convection by, e.g. channeling effects, is represented much more realistically in high resolution nonhydrostatic forecast models.

In the beginning, the operational application of the model within COSMO were mainly on the meso-β scale using a grid spacing of 7 km. The key issue was an accurate numerical prediction of near-surface weather conditions, focusing on clouds, fog, frontal precipitation, and orographically and thermally forced local wind systems. Since April 2007, a meso-γ scale version is running operationally at DWD by employing a grid spacing of 2.8 km. Applications with similar resolutions are now run by most COSMO partners. We expect that this will allow for a direct simulation of severe weather events triggered by deep moist convection, such as supercell thunderstorms, intense mesoscale convective complexes, prefrontal squall-line storms and heavy snowfall from wintertime mesocyclones.

The requirements for the data assimilation system for the operational COSMO-Model are mainly determined by the very high resolution of the model and by the task to employ it also for nowcasting purposes in the future. Hence, detailed high-resolution analyses have to be able to be produced frequently and quickly, and this requires a thorough use of asynoptic and high-frequency observations such as aircraft data and remote sensing data. Since both 3-dimensional and 4-dimensional variational methods tend to be less appropriate for this purpose, a scheme based on the observation nudging technique has been chosen for data assimilation from the beginning of the development. But in March 2017 the nudging scheme has been replaced by a new, more modern, ensemble-based method, called KENDA: Km-scale ENsemble Data Assimilation. Note, that KENDA, unlike the nudging scheme, is not available within the source code of the COSMO-Model.

Besides the operational application, the COSMO-Model provides a nonhydrostatic modeling framework for various scientific and technical purposes. Examples are applications of the model to large-eddy simulations, cloud resolving simulations, studies on orographic flow

systems and storm dynamics, development and validation of large-scale parameterization schemes by fine-scale modeling, and tests of computational strategies and numerical techniques. For these types of studies, the model should be applicable to both real data cases and artificial cases using idealized test data. Moreover, the model has been adapted by other communities for applications in climate mode (CCLM) and / or running an online coupled module for aerosols and reactive trace gases (ART).

Such a wide range of applications imposes a number of requirements for the physical, numerical and technical design of the model. The main design requirements are:

(i) use of nonhydrostatic, compressible dynamical equations to avoid restrictions on the spatial scales and the domain size, and application of an efficient numerical method of solution;

(ii) provision of a comprehensive physics package to cover adequately the spatial scales of application, and provision of high-resolution data sets for all external parameters required by the parameterization schemes;

(iii) flexible choice of initial and boundary conditions to accommodate both real data cases and idealized initial states, and use of a mesh-refinement technique to focus on regions of interest and to handle multi-scale phenomena;

(iv) use of a high-resolution analysis method capable of assimilating high-frequency asynoptic data and remote sensing data;

(v) use of pure Fortran constructs to render the code portable among a variety of computer systems, and application of the standard MPI-software for message passing on distributed memory machines to accommodate broad classes of parallel computers.

The development of the COSMO-Model was organized along these basic guidelines. However, not all of the requirements are fully implemented, and development work and further improvement is an ongoing task. The main features and characteristics of the present release are summarized below.

COSMO-ICON Physics

In the last months, several physical packages have been unified with their counterpart in ICON, to reduce the maintenance work for having two different versions of one parameterization. We refer to this developments as the *COSMO-ICON Physics*.

A major technical change to implement this unification was, to use the ICON data structure for the variables in the physics. This structure does not reflect a horizontal field with two dimensions, but collects the grid points in a vector (or a block). This is in contrast to the COSMO-Model, which uses the `(i,j)`-structure for horizontal fields.

This *blocked data structure* is explained in more detail in Appendix A of Part II, the Physical Parameterizations.

Not all options for the parameterizations have been ported to the blocked data structure. More details are given below in the *Physical Parameterizations*.

Dynamics

- **Model Equations** – Nonhydrostatic, full compressible hydro-thermodynamical equations in advection form. Subtraction of a hydrostatic base state at rest.
- **Prognostic Variables** – Horizontal and vertical Cartesian wind components, pressure perturbation, temperature, specific humidity, cloud water content. Optionally: cloud ice content, turbulent kinetic energy, specific water content of rain, snow and graupel.
- **Diagnostic Variables** – Total air density, precipitation fluxes of rain and snow.
- **Coordinate System** – Generalized terrain-following height coordinate with rotated geographical coordinates and user defined grid stretching in the vertical. Options for (i) base-state pressure based height coordinate, (ii) Gal-Chen height coordinate and (iii) exponential height coordinate (SLEVE) according to Schär et al. (2002).

Numerics

- **Grid Structure** – Arakawa C-grid, Lorenz vertical grid staggering.
- **Spatial Discretization** – Second-order finite differences. For the two time-level scheme also 1st and 3rd to 6th order horizontal advection (default: 5th order). Option for explicit higher order vertical advection.
- **Time Integration** – Two time-level 2nd and 3rd order Runge-Kutta split-explicit scheme after Wicker and Skamarock (2002) and a TVD-variant (Total Variation Diminishing) of a 3rd order Runge-Kutta split-explicit scheme. Option for a second-order leapfrog HE-VI (horizontally explicit, vertically implicit) time-split integration scheme, including extensions proposed by Skamarock and Klemp (1992). Option for a three time-level 3-d semi-implicit scheme (Thomas et al. (2000)) based on the leapfrog scheme.
- **Numerical Smoothing** – 4th-order linear horizontal diffusion with option for a monotonic version including an orographic limiter. Rayleigh damping in upper layers. 2-d divergence damping and off-centering in the vertical in split time steps.

Initial and Boundary Conditions

- **Initial Conditions** – Interpolated initial data from various coarse-grid driving models (ICON (and former GME), ECMWF, COSMO-Model) or from the continuous data assimilation stream (see below). Option for user-specified idealized initial fields.
- **Lateral Boundary Conditions** – 1-way nesting by Davies-type lateral boundary formulation. Data from several coarse-grid models can be processed (ICON (and former GME), IFS, COSMO-Model). Option for periodic boundary conditions.
- **Top Boundary Conditions** – Options for rigid lid condition and Rayleigh damping layer.
- **Initialization** – Digital-filter initialization of unbalanced initial states (Lynch et al. (1997)) with options for adiabatic and diabatic initialization.

Physical Parameterizations

- **Subgrid-Scale Turbulence** – Prognostic turbulent kinetic energy closure at level 2.5 including effects from subgrid-scale condensation and from thermal circulations. Option for a diagnostic second order K-closure of hierarchy level 2 for vertical turbulent fluxes (not ported to the blocked data structure). Option for calculation of horizontal turbulent diffusion in terrain following coordinates (3D Turbulence; tested in artificial setups).
- **Surface Layer Parameterization** – A Surface layer scheme (based on turbulent kinetic energy) including a laminar-turbulent roughness layer. Option for a stability-dependent drag-law formulation of momentum, heat and moisture fluxes according to similarity theory This option has not been ported to the blocked data structure. (Louis (1979)).

- **Grid-Scale Clouds and Precipitation** – Cloud water condensation and evaporation by saturation adjustment. Precipitation formation by a bulk microphysics parameterization including water vapour, cloud water, cloud ice, rain and snow with 3D transport for the precipitating phases. Option for a new bulk scheme including graupel. Option for a simpler column equilibrium scheme.
- **Subgrid-Scale Clouds** – Subgrid-scale cloudiness is interpreted by an empirical function depending on relative humidity and height. A corresponding cloud water content is also interpreted. Option for a statistical subgrid-scale cloud diagnostic for turbulence. This option has not been ported to the blocked data structure.
- **Moist Convection** – Tiedtke (1989) mass-flux convection scheme with equilibrium closure based on moisture convergence. Option for the current IFS Tiedtke-Bechtold convection scheme.
- **Shallow Convection** – Reduced Tiedtke scheme for shallow convection only.
- **Radiation** – δ two-stream radiation scheme after Ritter and Geleyn (1992) short and longwave fluxes (employing eight spectral intervals); full cloud-radiation feedback.
- **Soil Model** – Multi-layer version of the former two-layer soil model after Jacobsen and Heise (1982) based on the direct numerical solution of the heat conduction equation. Snow and interception storage are included.
- **Fresh-Water Lake Parameterization** – Two-layer bulk model after Mironov (2008) to predict the vertical temperature structure and mixing conditions in fresh-water lakes of various depths.
- **Sea-Ice Scheme** – Parameterization of thermodynamic processes (without rheology) after Mironov and Ritter (2004). The scheme basically computes the energy balance at the iceâĂŹs surface, using one layer of sea ice.
- **Terrain and Surface Data** – All external parameters of the model are available at various resolutions for a pre-defined region covering Europe. For other regions or grid-spacings, the external parameter file can be generated by a preprocessor program using high-resolution global data sets.

Data Assimilation

- **Former Method** – Continuous four-dimensional data assimilation based on observation nudging (Schraff (1996), Schraff (1997)), with lateral spreading of upper-air observation increments along horizontal surfaces. Explicit balancing by a hydrostatic temperature correction for surface pressure updates, a geostrophic wind correction, and a hydrostatic upper-air pressure correction.
- **Actual Method** – Ensemble data assimilation based on the LETKF (Local Ensemble Transform Kalman Filter) (Schraff et al. (2016))
- **Assimilated Atmospheric Observations** – Radiosonde (wind, temperature, humidity), aircraft (wind, temperature), wind profiler (wind), and surface-level data (SYNOP, SHIP, BUOY: pressure, wind, humidity). Optionally RASS (temperature), radar VAD wind, and ground-based GPS (integrated water vapour) data. Surface-level temperature is used for the soil moisture analysis only.
- **Radar derived rain rates** – Assimilation of near surface rain rates based on latent heat nudging (Stephan et al. (2008)). It locally adjusts the three-dimensional thermodynamical field of the model in such a way that the modelled precipitation rates should resemble the observed ones.
- **Surface and Soil Fields** – Additional two-dimensional intermittent analysis:
  - **Soil Moisture Analysis** – Daily adjustment of soil moisture by a variational method (Hess (2001)) in order to improve 2-m temperature forecasts; use of a Kalman-Filter-like background weighting.

- **Sea Surface Temperature Analysis** – Daily Cressman-type correction, and blending with global analysis. Use of external sea ice cover analysis.
- **Snow Depth Analysis** – 6-hourly analysis by weighted averaging of snow depth observations, and use of snowfall data and predicted snow depth.

Code and Parallelization

- **Code Structure** – Modular code structure using standard Fortran constructs.
- **Parallelization** – The parallelization is done by horizontal domain decomposition using a soft-coded gridline halo (2 lines for Leapfrog, 3 for the Runge-Kutta scheme). The *Message Passing Interface* software (MPI) is used for message passing on distributed memory machines.
- **Compilation of the Code** – For all programs a Makefile is provided for the compilation which is invoked by the Unix *make* command. Two files are belonging to the Makefile: `ObjFiles` is a list of files that have to be compiled and `ObjDependencies` contains all file dependencies. In addition it reads the file `Fopts`, which has to be adapted by the user to specify the compiler, compiler options and necessary libraries to link.
- **Portability** – The model can be easily ported to various platforms; current applications are on conventional scalar machines (UNIX workstations, LINUX and Windows-NT PCs), on vector computers (NEC SX series) and MPP machines (CRAY, IBM, SGI and others).
- **Model Geometry** – 3-d, 2-d and 1-d model configurations. Metrical terms can be adjusted to represent tangential Cartesian geometry with constant or zero Coriolis parameter.

1.3 Single Precision Version

From the beginning of the development, the COSMO-Model had been designed to be able to run in both precisions: single and double precision. Therefore, the real variables are all defined using a `KIND`-parameter, named `wp` (means: working precision) in the module `kind_parameters.f90` (earlier, this `KIND`-parameter was named `ireals`). Other `KIND`-parameters are `sp` (for single precision) and `dp` (for double precision). Before compiling the model, the user has to decide whether `wp` will be set to `sp` or to `dp`. This can be done with the compiler pragma `-DSINGLEPRECISION`. If this pragma is set, single precision will be used, otherwise double precision.

But in the first years of the COSMO-Model, only the double precision version was developed and tested, nobody ever used or tried a single precision run.

But single precision programs run faster on computers, because of less memory traffic, therefore MeteoSwiss tested to run the COSMO-Model also in single precision. Which did not work in the first instance. Some effort had to be put in adapting the model to work for single precision.

The main changes are:

- Epsilons, which are used in comparisons or to make divisions safe, are adapted to work in both precisions. Variables `repsilon` and `rprecision` have been introduced in module `data_constants.f90`.
- New variables `imp_single` and `imp_double` are added to specify an appropriate MPI data type.

- To avoid automatic conversions by the compiler, all (!) real constants (as `2.0`, `0.5`, etc.) are now written with the kind parameter as suffix: `2.0_wp`, `0.5_wp`, etc.). Further developments should follow this rule!
- The pragma `SINGLEPRECISION` is now used to choose single precision for the COSMO-Model during compilation. If it is not set, double precision is used.

It turned out, that the radiation cannot be run in single precision (at least the routines `coe_th`, `inv_th`, `coe_so`, `inv_so`). Therefore it was decided to run the subroutine `fesft` and all routines called below in double precision. The necessary variables are defined with the `KIND`-parameter `dp`.

1.4 Organization of the Documentation

For the documentation of the model we follow closely the *European Standards for Writing and Documenting Exchangeable Fortran 90-Code*. These standards provide a framework for the use of Fortran-90 in European meteorological organizations and weather services and thereby facilitate the exchange of code between these centres. According to these standards, the model documentation is split into two categories: external documentation (outside the code) and internal documentation (inside the code). The model provides extensive documentation within the codes of the subroutines. This is in form of procedure headers, section comments and other comments. The external documentation is split into seven parts, which are listed in Table 1.2.

Table 1.2: COSMO Documentation: A Description of the Nonhydrostatic Regional COSMO-Model

Part I:	Dynamics and Numerics
Part II:	Physical Parameterization
Part III:	Data Assimilation
Part IV:	Special Components and Implementation Details
Part V:	Preprocessing: Initial and Boundary Data for the COSMO-Model
Part VI:	Model Output and Data Formats for I/O
Part VII:	User's Guide

Parts I - III form the scientific documentation, which provides information about the theoretical and numerical formulation of the model, the parameterization of physical processes and the four-dimensional data assimilation. The scientific documentation is independent of (i.e. does not refer to) the code itself. Part IV will describe the particular implementation of the methods and algorithms as presented in Parts I - III, including information on the basic code design and on the strategy for parallelization using the MPI library for message passing on distributed memory machines (not available yet). The generation of initial and boundary conditions from coarse grid driving models is described in Part V. This part is a description of the interpolation procedures and algorithms used (not yet complete) as well as a User's Guide for the interpolation program INT2LM. In Part VI we give a description

of the data formats, which can be used in the COSMO-Model, and describe the output from the model and from data assimilation. Finally, the User's Guide of the COSMO-Model provides information on code access and how to install, compile, configure and run the model. The User's Guide contains also a detailed description of various control parameters in the model input file (in NAMELIST format) which allow for a flexible model set-up for various applications. All parts of the documentation are available at the COSMO web-site (http://www.cosmo-model.org/content/model/documentation/core/default.htm).

Section 2

Introduction to the Data Assimilation for the COSMO-Model

In the context of numerical weather prediction, the requirements for any data assimilation system are closely related to the purpose and main characteristics of the model for which that system is to provide the initial conditions. As for the operational configurations of the COSMO-Model (which is also denoted briefly just as "COSMO"), the model is characterized by its very high resolution on a limited domain, and its main purpose is to deliver accurate short-range and even more time-critical very short-range forecasts.

Hence, detailed high-resolution analyses have to be able to be produced frequently and efficiently. This requires a thorough use of asynoptic and high-frequency observations such as aircraft data and remote sensing data. Note that the synoptic scales are largely determined by the lateral boundary conditions provided by the steering model, and the main task of the assimilation scheme is to analyze the meso-scales.

By design, 3-dimensional analysis methods such as optimum interpolation (OI) or 3-dimensional variational analysis (3DVAR) tend to be less appropriate for this purpose. They do not allow to account for the exact observation time of asynoptic data, and they make it necessary to neglect most of the high-frequent data unless the analysis scheme is applied very frequently. This, however, would increase the computational costs and could result in problems at asynoptic analysis times when the data density may become very low and inhomogeneous. Moreover, the geostrophic approximation, usually a key ingredient of such schemes as used e.g. for the global model GME of DWD, is of limited validity in the meso scale. Therefore, 4-dimensional methods offer potential advantages since they include the model dynamics in the assimilation process directly. As variational methods allow to compare observations with the model state in observation space, they appear, in principle, best suited for the use of many types of remote sensing data, at least as long as these data are not directly linked to physical bifurcation processes such as convection. Yet, the 4-dimensional variational (4DVAR) method has been too expensive in the past for operational application of COSMO considering the small amount of time available to produce the analyses and forecasts.

a) Observation Nudging

Therefore, a scheme based on the observation nudging technique has been developed to define the atmospheric fields. It is based on an experimental nudging assimilation scheme which had been developed for the former hydrostatic model DM and its Swiss version SM (Schraff (1996); Schraff (1997)) and which compared favorably with the at that time operational Optimum Interpolation analysis of DM in a number of test cases. The scheme for COSMO has then been adapted to the nonhydrostatic modelling framework, refined and extended in various aspects, and runs on distributed memory machines using domain decomposition. It is presented in Section 3.

b) Latent Heat Nudging

Radar-derived precipitation rates can be assimilated by an extra Latent Heat Nudging scheme. It computes additional temperature and humidity increments at each model column independently from each other. It is tuned and should be used only for convection-permitting model configurations (with horizontal mesh widths of $\leq 3\,\mathrm{km}$). The observation input is gridded precipitation rates read in the form of extra Grib files. Further Grib files can be read optionally, containing a blacklist, and radar beam height maps utilised for bright band detection.

A comprehensive and detailed scientific description as part of this documentation is still lacking. However, the main aspects of the scheme are described in Stephan et al. (2008).

c) Analysis of surface and soil fields

Since the data assimilation by nudging-types schemes takes place during the forward integration of the model, also the unobserved variables including those at the surface and in the soil are modified and, in principle, adapted. In addition to that, a set of 2-dimensional intermittent schemes for explicit analysis of some of the surface and soil fields can also be applied in a full data assimilation cycle for the COSMO model. Namely, this set comprises of three analysis schemes:

- a variational soil moisture analysis that uses daytime 2-m temperature observations to derive optimized soil moisture fields; this is described in Section 4;
- an analysis of the depth of the snow cover, see Section 5;
- a sea surface temperature (SST) analysis including an analysis of sea ice cover, see Section 6.

Some other surface and soil fields are specified as external parameters, and are either constant in time or updated once per day at 0 UTC. They are addressed in COSMO Documentation Part II on Physical Parameterization.

Section 3

Analysis of Atmospheric Fields: Nudging-Based Data Assimilation

3.1 Concept and Basic Ideas

Nudging or Newtonian relaxation consists of relaxing the model's prognostic variables towards prescribed values within a predetermined time window. Detailed descriptions of the technique can be found e.g. Anthes (1974), Davies and Turner (1977), and Stauffer and Seaman (1990). In the present scheme, nudging is performed towards direct observations, which is more appropriate for asynoptic data (Stauffer and Bao (1993)) and high-resolution applications than nudging towards 3-dimensional analyses (Stauffer and Seaman (1994)). A relaxation term is introduced into the model equations, and the tendency for the prognostic variable $\psi(\mathbf{x},t)$ is given by

$$\frac{\partial}{\partial t}\psi(\mathbf{x},t) \;=\; F(\boldsymbol{\psi},\mathbf{x},t) \;+\; G_\psi \cdot \sum_{k_{(obs)}} W_k(\mathbf{x},t)\cdot[\psi_k^{obs} - \psi(\mathbf{x}_k,t)] \tag{3.1}$$

F denotes the model dynamics and physical parameterizations, ψ_k^{obs} the value of the k^{th} observation influencing the grid point $\mathbf{x}$ at time t, $\mathbf{x}_k$ the observation location, G_ψ a constant called nudging coefficient (currently set to $12\cdot10^{-4}s^{-1}$ for surface pressure and $6\cdot10^{-4}s^{-1}$ for the other assimilated quantities), and W_k an observation-dependent weight. This weight always takes values between 0 and 1 (except for surface pressure, cf. Section 3.5.1, Eq. (3.36)) and is explained further below. The difference (in the square bracket) between observed and model value is called observation increment, and the complete additional so-called nudging term determines the analysis increment which is defined as the change explicitly imposed on the model value by the nudging. Neglecting the dynamics and physics and assuming a single observation with a constant weight W_k equal to 1, the model value at the observation location relaxes exponentially towards the observed value with an e-folding decay rate of $1/G_\psi$ corresponding to about half an hour. Thus, the so-called nudging equation (3.1) describes a continuous adaptation of the model values towards the observed values during the forward integration of the model (cf. Figure 3.1).

In practical applications, the nudging term should and usually does remain smaller than the largest term of the dynamics. This situation is related to the basic idea of the method that the model fields are to be relaxed towards the observed values without significantly disturbing

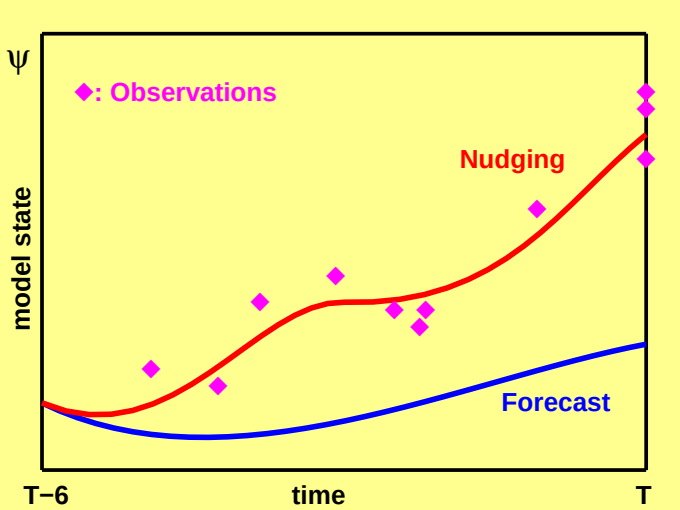

Figure 3.1: Conceptual illustration of the effect of nudging.

the dynamic balance of the model. The coupling between the innovations of the mass and the wind field is primarily imposed implicitly by the model dynamics. If the assimilation process is successful the model fields will be close to dynamic balance at the beginning of the forecast. Without requiring an initialization step, spin-up effects are reduced in comparison to 3-dimensional analysis methods where the coupling between mass and wind innovations usually relies on explicit diagnostic relationships such as the geostrophic approximation.

The factors W_k determine the relative weights given to the different observations at a specific grid point. For a single observation, this weight (w_k) comprises of the quality (and representativeness) of the observation (ϵ_k) and of weights which depend on the horizontal (w_{xy}) or vertical (w_z) distance respectively temporal (w_t) difference between the observation and the target grid point,

$$w_k \quad = \quad w_t \cdot w_{xy} \cdot w_z \cdot \epsilon_k \tag{3.2}$$

If an increasing number of observations influence the grid point the total nudging weight should be limited to prevent the nudging term from becoming dominant over the dynamics. This is achieved by complementing the individual weight w_k with a relative weight for multiple observations:

$$W_k \quad = \quad \frac{w_k}{\sum_j w_j} \cdot w_k \tag{3.3}$$

This simple approach is designed to improve the gradients of the analyzed fields (Benjamin and Seaman (1985)) and has therefore been widely used in nudging schemes. Due to the squaring of the weights, the fit to an isolated observation e.g. over the ocean is less impaired in its near environs by the simultaneous assimilation of a remote group of coastal data. In rather data dense areas, equation 3.3 automatically reduces moderately the effective scale of the horizontal spreading (for which the explicit functions will be described in section 3.7). This improves e.g. the analysis of small-scale cyclones by assimilation of surface pressure data from a rather data-dense network. Note however, that the weighting for multiple observations according to equation 3.3 has the undesirable property that with increasing number of observations, the total sum of the weights from all observations together tends to decrease instead of (slightly) increase. For instance, if there is one observation with individual weight

$w_k = 1$, then adding an increasing number of further observations with $w_k = 0.1$ will effectuate the total sum of the weights to reduce from 1 towards 0.1 (!).

Therefore, an alternative relative weighting for multiple observations,

$$W_k = \frac{w_k + 1}{\sum_j w_j + 1} \cdot w_k \tag{3.4}$$

is also implemented which is not so scale selective as equation 3.3, but neither has its undesired property mentioned. In the above example, the total sum of the weights would increase from 1 towards 1.1, and for an increasing number of observations with $w_k = 1$, the sum of the weights would converge towards a value of 2 which provides an absolute upper limit.

Note furthermore, that the approach adopted here (both using equation 3.3 or 3.4) only accounts for the data density at the target grid points and neglects the relative positions between the observations themselves in contrary to optimum interpolation. In order to take the relative positions into account as well, the data density at the observation locations would have to be determined at first and included in the quality factors ϵ_k (Lorenc et al. (1991)) at a considerable computational cost.

Besides the simplicity of the concept, the fact that the method can be implemented relatively easily, and the essential advantages of the nudging method outlined in the introduction, potential difficulties and disadvantages of the method should also be mentioned. Firstly, in contrast to optimum interpolation (OI) and 3- or 4-dimensional variational (3DVAR, 4DVAR) methods, there is no mathematical formalism to determine a theoretically optimal solution to the analysis problem. Therefore, there are several free parameters, and theoretical considerations can only provide rough estimations for their optimal specification. More appropriate values have to be found by means of relatively expensive tuning experiments. Note that on the other hand, OI, 3DVAR, and 4DVAR schemes are known to be very sensitive to the (correct) specification of the model error covariances. Results from tuning experiments suggest that this sensitivity is reduced in the nudging scheme (at least for the relatively data-dense area over Europe). This is likely due to the direct inclusion of the model dynamics in the assimilation process which modifies the effective influence of the observations during the assimilation period (Schraff (1996)).

The second and probably most important disadvantage in comparison to 3D- and 4DVAR is related to the fact that, similarly to OI, observation increments have to be expressed in model space rather than observation space. This means that for an observational information of any kind, observation increments have to be derived always in terms of the prognostic model variables in order to be used in the nudging. (This limitation can be relaxed to the extent that the increments can be expressed in terms of variables each of which can be mapped to one model variable by a function for which strict monotony is a necessary condition and linearity a sufficient condition.) That limitation does not apply to variational schemes. In order to assimilate e.g. radar reflectivity observations in a variational scheme, reflectivity values can be derived from the model fields in order to compute observation increments of reflectivity which can then be used directly for the analysis. For nudging, in contrast, it is required to deduce increments of temperature, humidity, and/or wind, etc. from the observed reflectivity (or from precipitation data). This operation has more degrees of freedom, requires more assumptions with less confidence, and is therefore more prone to errors.

Thirdly, correlations of observation errors which typically may occur with remote sensing data cannot be taken into account. Unless additional balancing steps are added, this also

applies to most types of cross-correlations of model errors such as error correlations between the wind and the mass field. It is noted however, that deploying the latter types of correlations is advantageous only for those scales for which the diagnostic relationships (e.g. geostrophy) that are usually used to specify these correlations are good approximations. In the framework of nudging, the direct inclusion of the model dynamics in the assimilation process leads to an implicit coupling between the resulting changes of the model fields. In general, this kind of coupling is incomplete, but it includes all the scales present in the model. Depending on the situation, this can be even advantageous, particularly in the meso-scales (as outlined in the introduction) and data-rich areas. Furthermore, as indicated above, explicit coupling between the increments of different model fields can partly be included by adding extra balancing steps, and this is done in the current scheme.

In fact, three types of balancing are applied to the analysis increment fields before they are added to the model fields. First, a hydrostatic upper-air temperature correction balances the pressure analysis increments at the lowest model layer. Secondly, a geostrophic wind correction balances the wind field with respect to the mass field increments as induced by the surface pressure nudging including the temperature correction. This yields a (weak) explicit coupling between the mass and wind field increments. Finally, an upper-air pressure correction balances the total analysis increments of the mass field hydrostatically.

Equation (3.1) indicates that apart from the balancing steps, the scheme consists of two main steps. After the observations have been made available to the scheme in an appropriate form in the observation processing (Sections 3.3 and 3.4), the first step is to compute the observation increments (Section 3.5). This is usually related to some sort of interpolation, and is complemented by a quality control of the observations (Section 3.6). In the second step, the weights are computed, and the increments provided with the weights are spread to the target grid points for each observation (Section 3.7). These weighted increments are then summed up to form the analysis increment fields. Before this is addressed in detail for the different variables (Section 3.9), additional increments are derived in extra balancing steps (Section 3.8). Prior to these detailed descriptions of the whole process from the observational input up to the final analysis increments, the general discretized formulation of the nudging equation is derived in the following Section 3.2.

3.2 Discretized Formulation

This section addresses the implementation of the nudging terms in the finite difference form of the set of extended dynamic equations (3.1). In finite time differences, the nudging is carried out as the last operation within a timestep except for the relaxation towards the lateral boundary fields and for the saturation adjustment.

The finite difference form of Eqs. (3.1) read

$$\frac{\psi^{n+1} - \psi^{n_{old}}}{\Delta\tau} \quad = \quad F(\psi^n) \quad + \quad G_\psi \cdot \sum_{k_{(obs)}} W_k \cdot [\psi_k^{obs} - \psi_{x_k}^m] \tag{3.5}$$

Here, $\psi^{n_{old}}$, ψ^n, and ψ^{n+1} denote the values of the model variable ψ at some grid point at an old, the current, respectively the next timestep. $\psi_{x_k}^m$ is the model value at the observation location at timestep m which has yet to be defined. $\Delta\tau$ is the interval between time levels n_{old} and $n+1$, and the other variables are as in Eqs. (3.1). For a two-timelevel scheme

such as the Runge Kutta time integration scheme, $n_{old} = n$ and $\Delta\tau = \Delta t$ where Δt is the length of the (advective) timestep. For the three-timelevel leap-frog time stepping, in contrast, $n_{old} = n-1$ and $\Delta\tau = 2 \cdot \Delta t$.

The timestep m is chosen to be equal to $n+1$, following the argument that the distance of the observation to the model value at the next rather than current timestep should be decreased by the relaxation. This renders (3.5) a set of implicit equations for $\psi^{n+1}_{(x_k)}$.

Each of the dynamic equations without nudging term can be written as

$$\frac{\psi_F^{n+1} - \psi^{n_{old}}}{\Delta\tau} = F(\psi^n) \tag{3.6}$$

ψ_F^{n+1} denotes the value of ψ after the addition of dynamic and physical tendencies and prior to the application of the nudging tendency within timestep $n+1$. For each of Eqs. (3.5), replacing $F(\psi^n)$ by (3.6) renders

$$\psi^{n+1} = \psi_F^{n+1} + \Delta\tau\, G_\psi \cdot \sum_{k_{(obs)}} W_k \cdot [\psi_k^{obs} - \psi_{x_k}^{n+1}] \tag{3.7}$$

An explicit form for this implicit equation can only be derived in a strict way at the grid point which coincides with the location of a single observation. This allows to omit $\sum$ and subscript $_{x_k}$, and Eq. (3.7) can be written as

$$\psi^{n+1} = \psi_F^{n+1} + \frac{\Delta\tau\, G_\psi\, W_k}{1 + \Delta\tau\, G_\psi\, W_k} \cdot [\psi_k^{obs} - \psi_F^{n+1}] \tag{3.8}$$

This equation is then generalized to any (other) grid point $\mathbf{x}$ by taking into account that it is the value of the model field at the observation location $\mathbf{x}_k$ that should be used for comparison to the observed value. Hence, the model value at $\mathbf{x}_k$ is used to compute the observation increment, and the generalized equation reads

$$\psi^{n+1}(\mathbf{x}) = \psi_F^{n+1}(\mathbf{x}) + \frac{\Delta\tau\, G_\psi\, W_k}{1 + \Delta\tau\, G_\psi\, W_k} \cdot [\psi_k^{obs} - \psi_F^{n+1}(\mathbf{x}_k)] \tag{3.9}$$

Finally, this is further generalized to multiple observations by reintroducing the summation $\sum$, i.e. W_k is replaced by $\sum W_k$ and $W_k \cdot [\psi_k^{obs} - \psi_F^{n+1}(\mathbf{x}_k)]$ by $\sum(W_k \cdot [\psi_k^{obs} - \psi_F^{n+1}(\mathbf{x}_k)])$. After replacing the weights W_k by the right side of Eq. (3.3) or (3.4), the general equation used in this nudging scheme then reads

$$\psi^{n+1}(\mathbf{x}) = \psi_F^{n+1}(\mathbf{x}) + \frac{\Delta\tau\, G_\psi\, \mu}{1 + \Delta\tau\, G_\psi\, \mu} \cdot \overline{\Delta\psi} \tag{3.10}$$

where

$$\overline{\Delta\psi} = \frac{\sum_k ((w_k^2 + c_w\, w_k) \cdot [\psi_k^{obs} - \psi_F^{n+1}(\mathbf{x}_k)])}{\sum_{k'} (w_{k'}^2 + c_w\, w_{k'})} \tag{3.11}$$

$$\mu = \frac{\sum_k (w_k^2 + c_w\, w_k)}{\sum_{k'} w_{k'} + c_w} \tag{3.12}$$

where $c_w = 0$ when using Eq. (3.3) and $c_w = 1$ when using Eq. (3.4).

$\overline{\Delta\psi}$ is a square weighted mean of observation increments. By analogy of Eq. (3.10) to Eq. (3.9), $\overline{\Delta\psi}$ can also be regarded as a net observation increment at the target grid point.

Similarly, the weighted mean of nudging weights μ can be seen as a net nudging weight for the net increment $\overline{\Delta\psi}$.

By definition, the analysis increments are the values that are added to the model fields by the nudging within one timestep. They are given by the complete fractional term on the right side of Eq. (3.10) (if the additional balancing steps are neglected). In order to determine these analysis increments, the following 3 terms must be computed: $\sum_k w_k$, $\sum_k {w_k}^2$, and $\sum_k(({w_k}^2 + c_w\, w_k) \cdot [\psi_k^{obs} - \psi_F^{n+1}(\mathbf{x}_k)])$.

Separate Computation of Preliminary Analysis Increments for Different Observation Types

It has already been pointed out that in order to maintain a reasonable balance of the model fields, the size of the nudging term has to be limited rather stringently, and that therefore, an additional weighting for multiple observations has to be introduced, here in the form of Eqs. (3.3) or (3.4). However, this has the effect of reducing rather drastically the effective weight and therefore the impact of e.g. a single radiosonde (profile) observation if a data-dense set of other observations such as satellite, radar wind, or GPS IWV retrievals are added. One way to address this could be to strongly reduce the quality weight ϵ_k (or almost equivalently the nudging coefficient) for these additional observations, but this would make the assimilation of these data inefficient also far away from the radiosonde observation.

Therefore, an option has been introduced to first compute the net increments (preliminary analysis increments) for different (sets of) observation types independently and afterwards combine these net increments and net weights to obtain the final analysis increments. In this case, the radiosonde increment enters the final weighting for multiple observation (types) only together with one other net increment (with similar weight) from the other observation type even if there are many observations of that type. As a result, the final weight of the radiosonde observation is far less reduced, but at the same time, the assimilation of the other data can remain efficient further away from it.

In order to accomplish this, the net increments and corresponding net weights are first computed independently for each set of observation types m by Eqs. (3.11) and (3.12) (substituting $\overline{\Delta\psi}$, μ, c_w, k, and k' by $\overline{\Delta\psi}^m$, μ_m, $c_{w(m)}$, k_m, and k'_m respectively) using only the observations $\psi_{k_m}^{obs}$ of the corresponding set of observation types. As Eq. (3.10) takes the same form as Eq. (3.9), the net increment $\overline{\Delta\psi}^m$ and net weight μ_m for each set of observation types m can then be regarded and further treated as a simple single observation with its individual weight. Therefore, in order to combine these increments and weights for the computation of the final analysis increments, Eq. (3.3) or (3.4) for the relative weighting of multiple observations can be re-applied:

$$\psi^{n+1}(\mathbf{x}) \;=\; \psi_F^{n+1}(\mathbf{x}) \;+\; \frac{\Delta\tau\; G_\psi\; \mu_{tot}}{1 + \Delta\tau\; G_\psi\; \mu_{tot}} \cdot \overline{\Delta\psi}^{tot} \tag{3.13}$$

where

$$\overline{\Delta\psi}^{tot} \;=\; \frac{\sum_m((\mu_m^2 + c_{w(tot)}\,\mu_m) \cdot \overline{\Delta\psi}^m)}{\sum_{m'}(\mu_{m'}^2 + c_{w(tot)}\,\mu_{m'})} \tag{3.14}$$

$$\mu_{tot} \;=\; \frac{\sum_m(\mu_m^2 + c_{w(tot)}\,\mu_m)}{\sum_{m'}\mu_{m'} \;+\; c_{w(tot)}} \tag{3.15}$$

Note that $c_{w(m)}$ can be chosen to take different values (0 or 1) for different observation types m. For instance, $c_{w(1)} = 0$ could be set for conventional in-situ data (being set 1 of

observation types) and $c_{w(2)} = 1$ for (a dense set of) satellite retrievals. In such a case, μ_1 would be less than 1 even close to the observation locations, whereas μ_2 would be typically about 1.5. In order not to give too much weight to the observation type 2 purely due to the different choice of $c_{w(2)}$, the quality weights (in practice: the nudging coefficients) for observation type 2 should be reduced by about a factor of 2. In order to prevent the nudging term from becoming too large, a similar adaptation should also be made for all nudging coefficients if $c_{w(m)} = 1$ is chosen for all observation types.

For $c_{w(tot)}$, a value of 0 seems less appropriate. This is firstly because the weights μ_m will not usually vary as strongly as the observation density itself, which renders the advantage of applying Eq. (3.3) less relevant. Secondly, the total weight for all observation types together would result to be smaller than the (largest of the) weights for the single (sets of) observation types, and this is not desirable. Therefore $c_{w(tot)} = 1$ is recommended.

3.3 General Observation Processing

In this documentation, the term 'observation processing' is used for the process of making the observations available to the nudging-type scheme in an appropriate form within the model code. The first step is to read the observation reports from files – either NetCDF observation input files (see COSMO Documentation Part VII – User's Guide) or a single AOF file (Analysis Observation File – a separate documentation is available from christoph.schraff@dwd.de). GPS data can also be read from a separate ASCII file in COST-716 format. Apart from the reading, the observation processing also includes assigning the observations temporally and spatially to the model space, exploiting quality flags and other attributes, putting different TEMP (or PILOT) radiosonde parts together in single complete profiles, applying bias corrections, and performing gross error and redundancy checks. These issues are addressed in the current Section 3.3, and additional processing steps specific to aircraft observations are presented in Section 3.4.

Except for some minor aspects mentioned in the following section, the observation processing, however, does not include any issues for which the observed values have to be related to the time-dependent model values, as is the case e.g. for the computation of observation increments or for performing the threshold quality control. While the latter types of issues have to be redone at each timestep in principle (or at least frequently, cf. Section 3.9), the observation processing is performed only once for each observation at the timestep when the observation is read from the files.

3.3.1 Temporal Aspects

As a consequence of the previous remarks, the feature characteristic to the observation processing is that it deals with computations and tests that are basically independent from time and from the time-dependent model fields. Yet, there are a few exceptions to this, where model values are used to complement missing pieces of information for some of the observational reports:

- for PILOT and SATOB data, model pressure is used to assign reported height to pressure levels;

- aircraft data are set passive if their pressure level (reported or derived from flight level using the ICAO standard atmosphere) is less than 3 hPa (about 25 m) above the model surface pressure level - this is done with the intention to exclude data that are too close to the ground;
- for surface-level reports without pressure observation, model pressure at observation height is used to assign the report in the vertical for the pressure-level dependent blacklist and temperature gross error checks;
- model layer thickness is used to define minimum vertical correlation scales for aircraft data (cf. Section 3.4.4).

While the time dependency is very weak for the two last items, it can have an influence for the first two ones. It is due to this that the results may depend on when and how often data are read from the AOF resp. the NetCDF observation input files. (Another reason is the fact that the construction of piecewise profiles from aircraft data (Section 3.4.2) is repeated whenever new observations from the same aircraft are read.)

In the current (operational) implementation, data are read hourly (except for the GPS data which are read all at once at the beginning, provided that they are not read from a NetCDF observation input file but from a special ASCII flle according to COST-716 format specifications). In the case of the AOF file, reading in hourly batches requires that the observations are sorted in the file according to observation time.

The temporal weighting (see Section 3.7.6) as used in the operational setups implies that radiosonde data must be available 3 hours prior to and up to 1 hour after the observation time. As a result, the data from -1 to +4 hours relative to the initial model time must be provided at the first time step. Later on, data are always read 3 to 4 hours ahead of the observation time. Hence within the observation processing, the temporal misfit of the model values to the observation time is up to 4 hours. This applies to all types of data, if read from an AOF file, but is less for other types of data, if they are read from NetCDF files and shorter temporal weighting functions are used.

Data which are too old from the beginning or have become too old during the course of the model integration are removed. The same applies to reports, for which a quality flag indicates that reported date or time is suspicious.

3.3.2 Observation Types Used

Operationally (e.g. at DWD), the following data are used:

- radiosonde observations:
  - upper-air wind (for observation report types TEMP and PILOT) temperature (TEMP only): all mandatory-level data up to 10 hPa all significant-level temperature data up to 100 hPa and all significant-level wind data up to 200 hPa
  - upper-air humidity (TEMP only): all mandatory-level and significant-level data up to 300 hPa
  - surface-level wind, temperature, and humidity: used separately in the same way as SYNOP data (see below)
  - geopotential (TEMP only): used only to derive 1 single pressure observation increment at the height of the lowest model level (see Section 3.5.1)

- aircraft observations (of report type AIREP, AMDAR, or ACARS):
  - all wind and temperature data (except for data less than about 50 m above the ground, cf. Section 3.3.3)
- wind profiler observations: upper-air wind , used at all levels
- surface-level observations (of type SYNOP, SHIP, DRIBU (drifting buoys), or TEMP):
  - station pressure (or with lower priority: pressure reported at an alternative level as derived by extrapolation): used from most stations (for restrictions, see Section 3.3.3)
  - 10-m wind: used only from selected stations (see Section 3.3.3)
  - 2-m humidity: used from most stations (see Section 3.3.3)

2-m temperature observations are not used operationally. Nudging of these data has been shown to have potentially adverse effects on the stability of the planetary boundary layer (Stauffer et al. (1991)). In pre-operational tests, it has also tended to degrade the low-tropospheric thermal structure in COSMO. Note, however, that daytime 2-m temperature data are used in the soil moisture analysis, where they have the potential to modify the surface fluxes and to improve the prediction of 2-m temperature throughout the forecast period (Section 4).

Nevertheless, the nudging scheme itself has an option to use 2-m temperature observations. Altogether, the following types of data can be used optionally in the nudging scheme:

- 2-m temperature from SYNOP or TEMP reports
- upper-air wind from Doppler radar VAD (Velocity Azimuth Display) reports
- upper-air virtual temperature from RASS (Radio Acoustic Sounding Systems) or SODAR (Sonic Detection and Ranging) systems
- integrated water vapour (IWV) derived from ground-based GPS total zenith delay
- Scatterometer 10-m wind data (of type ASCAT and OSCAT) over water areas –

  Unlike other observation types, the scatterometer data have to be pre-processed by a separate program. This reads 10-m wind products from KNMI (for the ASCAT wind product, CMOD5.5 is currently used to retrieve the wind vectors valid at 10 m above the sea surface from the microwave radiance data), performs abiguity removal, thinning (for ASCAT gridded data: from a resolution of 12.5 or 25 km down to 50 km; no thinning of OSCAT data with 50 km resolution), and quality control. Winds with velocities < 3.5 m/s or > 25 m/s are less reliable and therefore discarded.

- SATOB data: upper-air single-level atmospheric motion vector (AMV) horizontal wind data derived from sequences of satellite images – note that the use of these data is not (well) tested

Additional data are read and processed or derived only in order to be written to a file for verification purposes without any influence on the data assimilation. These are the following:

- from SYNOP (incl. SHIP, and partly from TEMP):
  - 3-hour pressure tendency
  - precipitation, aggregated over different periods
  - (horizontal) visibility
  - total cloud cover

  - cloud base height of lowest cloud
  - low cloud cover, derived from observed cloud, visibility, and weather information
  - mid-level or high cloud cover, derived from observed cloud information
  - general cloud group and individual cloud layer groups (which include cloud base height, cloud amount, and cloud type)
  - code for present weather
  - maximum and minimum 2-m temperature during past 12 hrs
  - maximum wind speed of gusts over 1-hour or 6-hour periods
  - total snow depth
  - state of ground
- from aircraft :
  - degree of turbulence
  - maximum derived vertical gust
- from wind profiler and radar VAD reports :
  - vertical wind speed
  - standard deviation of the horizontal wind
- from wind profiler and RASS reports : signal to noise ratio

Furthermore, precipitation rate derived from radar reflectivity is assimilated by means of latent heat nudging scheme. This is used operationally (only) in the convection-permitting configuration(s) of COSMO, but it is not described in this chapter.

An experimental version of COSMO (based on version V4_18) exists which allows to use additionally AMSU-A and / or SEVIRI satellite radiances from polar orbiting satellites respectively Meteosat. Since indirect observations such as radiances cannot be used directly with the nudging technique, retrievals of temperature and humidity profiles are first derived here by a 1DVAR approach and then used in the nudging in a similar way as conventional profile data.

3.3.3 Spatial Aspects, Assignment of Reports

For a high-resolution model such as COSMO which is usually run with a mesh width of $\Delta x = 7\,\mathrm{km}$ or less it is not considered necessary to interpolate the model values horizontally to the exact observation location. A horizontal shift of 1 to 2 grid lengths of that size appears to be acceptable, all the more that grid point models cannot correctly represent wave lengths of $2\,\Delta x$ or less. Instead, it is considered more important that the vertical representativeness of the model values used for comparison with the observations is good. For the computation of surface-level observation increments, large differences between station height and corresponding model orography can lead to large extrapolation errors (e.g. for pressure) or representativeness errors. The occurrence of the latter is particularly obvious for humidity observations from mountain stations which lie above a low-level inversion and which would be compared to a model-derived value below the inversion, or vice versa. Similarly, problems must be expected at the use of sounding data within the planetary boundary layer, if the model orography deviates strongly from the height of the observing station.

As a consequence, the observation location is assigned horizontally to an appropriate model grid point. Upper-air reports not related to a surface station such as single-level aircraft reports are assigned to the nearest grid point. For the other reports, a distinction is made between sea observations and land observations. Reports are considered sea reports here, either if all of the four model grid points surrounding the station location are sea grid points,

or if at least one of these four points is a sea point and the reporting station is labelled a sea station like a drifting buoy or a ship issuing surface-level or radiosonde reports. The sea reports are assigned to the nearest sea grid point. Reports from sea stations are rejected if all of the four neighbouring grid points are land points. The rejection, however, is not applied to the upper-air part of ship radiosonde reports if one of the four land points has a water fraction greater than 1 % .

Land reports are assigned to the nearest land grid point only if the (geometrical) distance to it is less than half of the latitudinal mesh width Δy in the horizontal and less than 40 m in the vertical. Otherwise, the land point with the minimum effective vertical distance Δz_e to the station height is selected out of a group of grid points. Specifying this group to be the four surrounding grid points, as is often done in verification packages, would render a maximum horizontal assignment error of $\sqrt{2}\,\Delta y$. This promotes the choice adopted here, that the group consists of all the grid points within a horizontal search radius of $\sqrt{2}\,\Delta y$ from the observation location. By doing so, the number of candidate points is increased from 4 to up to 9 near the equator of the rotated model grid and to even more points far away from it. Thus, the chance to find an appropriate grid point with a small vertical distance to the observing station is enhanced without increasing the maximum assignment error.

Instead of geometrical distances $|\Delta z|$, scaled distances $\Delta z_e = f_D \cdot |\Delta z|$ are used here as 'vertical distances'. The reason is that for deriving surface pressure increments (see Section 3.5.1), vertical extrapolation errors tend to be much smaller for negative height differences $\Delta z = z_{mo} - z_{obs}$ between model grid point and observing station than for positive differences. The reason is that for negative differences, the extrapolation of full (observation) values can be replaced by an interpolation of full values followed by an extrapolation of increments. While f_D is always set to 1 for positive differences, it attains different values for negative differences. For surface pressure, it is set to 0.25 (e.g. for deriving quality weights, cf. Section 3.5.1). For surface-level wind, temperature, and humidity, it is set to 1 because there is no difference between upward and downward extrapolation in this case (cf. Section 3.5.2). For the horizontal grid point assignment search, it is set to 0.5 as a compromise.

Operationally, surface pressure observations are rejected if the corresponding scaled distance exceeds 150 m (corresponding to geometric differences of more than -600 m or +150 m), because the extrapolation errors are considered to be too large then. The rejection limit for 2-m humidity is 150 m, and for 10-m wind, it is set to 100 m . In addition, 10-m wind data are currently not used from stations above 100 m above sea level (which should be regarded as a temporary, very rough criterion to select only stations over flat terrain). If reduced pressure, e.g. at mean sea level, is reported instead of station pressure, the vertical extent of the reduction is added to the scaled extrapolation distance, if the reduced level lies between the observation station and the model orography. If the reduced level lies either below or above both the station height and the model orography, then the maximum of the (scaled) extrapolation distance between reduced level and either station height or model orography is used as final extrapolation distance.

In general, SYNOP reports or surface-level data from radiosondes are rejected if station altitude is missing or flagged to be suspicious. And any type of report is fully neglected if longitude or latitude is attributed to be probably bad.

Furthermore, observations located within 5 grid rows from the lateral boundaries of the COSMO domain (or located even outside of the COSMO domain or outside of an optional user-defined area) are rejected. Also, observations from further inside of the domain are never

assigned to a grid point of these 5 rows. The reason is as follows. Within the lateral boundary zone, the model fields are also relaxed towards boundary fields. If the latter disagree with an observation assigned to that zone and if that observation was being nudged, the model would tend to attain a value between the observed and the boundary value without further approaching either of them after a while. At each model timestep, then there would be, say, a negative observation increment and hence a negative analysis increment, and a positive boundary relaxation increment. While this is not a problem within the boundary zone itself, the negative increments would accumulate in time in the inner-domain part of the area of influence of the observation (by lateral spreading, see Section 3.7) without being compensated by positive boundary relaxation increments.

3.3.4 Other Aspects

Each observational report consists of a report header containing information on the report as a whole and of a report body containing the individual observed values and properties related to them. The report header is evaluated first, and most of the Sections 3.3.1 to 3.3.3 are part of this evaluation. It is complemented by the following steps:

- A station identifier is read for each report. Aircraft identifiers starting with '***', 'XXX', '???', '///', or ' ' are considered bad, and reports with such station identifiers are neglected.
- Blacklisted ship reports are also neglected, if read from an AOF file. If the observations are read from NetCDF files, blacklisting in done within the observation processing, see below.
- If the observations are read from NetCDF observation input files, a flag word is determined in the form of a bit pattern which indicates all the reasons why a complete report has been set passive or rejected.
- A station characteristics indicator is compiled which is not used outside the observation processing. It contains various information such as an instrument specification, flags indicating whether and why a report is set passive (i.e. rejected), and a station correction indicator. The latter is used in the redundancy checking (Section 3.3.7).
- If the observations are read from NetCDF observation input files, data category, international data sub-category, originating centre, originating sub-centre (mainly used for GPS data), and update sequence number (correction indicator) are also read.

Then, the report body is evaluated. This includes

- selecting and rejecting data according to Sections 3.3.2 and 3.3.3;
- rejecting upper-air observation levels if pressure at TEMP reports or both pressure and height at other report types is missing or flagged to be probably bad;
- providing pressure at upper-air observation levels at which only height is reported; this is done by using the ICAO standard atmosphere for aircraft reports (for which reported 'height' is in fact flight level instead of geometrical height) respectively the current model profile for PILOT reports (cf. Section 3.3.1);
- rejecting non-surface-level data if they are below the model orography or the station height, or rejecting aircraft data if they are within about 25 m (3 hPa pressure) of the model orography;

- rejecting data flagged by the data provider to be probably bad or to have large errors (e.g. based on the roll angle flag for aircraft wind data, low precision of aircraft or RASS temperature, aircraft mixing ratio quality flag, low signal to noise ratio for wind profiler and RASS data, large standard deviation of radar VAD wind speed, etc.);
- rejecting data from blacklisted stations or stations which are missing on whitelists – the exact processing is described in COSMO Documentation Part VII: User's Guide;
- doing consistency checks, e.g. rejecting humidity if the reported relative humidity deviates from that derived from dewpoint or mixing ratio by more than 4 %;
- rejecting zero winds from drifting buoys or scatterometers (because they are considered suspect), rejected radar VAD wind data with wind speed less than 3 m/s;
- providing observation errors as derived from rmse tables and sometimes enhanced to account for extrapolation errors; these errors are derived for active observations only, but currently they are not used thereafter except as active data indicators;
- sorting upper-air observation levels, checking for multiple surface levels and rejecting levels below the (selected) surface level;
- converting reported wind direction and speed into rotated wind components;
- converting observed humidity into model-compatible (relative) humidity (see below);
- applying bias corrections to humidity-related observations.
- deriving low cloud cover from observed cloud, visibility, and weather information, and mid-level and high cloud from observed cloud information;

This is complemented by performing bias correction (Section 3.3.5), gross error and consistency checks (Section 3.3.6), redundancy checks (Section 3.3.7, putting together TEMP parts A – D into complete profiles (Section 3.3.8), and additional processing steps specific to aircraft reports (Section 3.4). Furthermore, a considerable amount of diagnostic output and statistics on the observations processed and rejected is produced (see also COSMO Documentation Part VII: User's Guide).

Converting Observed Humidity into Model-Compatible Humidity

If a model version without prognostic cloud ice is deployed, then, with regard to water in the atmosphere, the model can only distinguish between water vapour and liquid water in the dynamic equations. Hence, saturation vapour pressure is always computed over water. In reality, however, ice cloud could and often would form at lower values of vapour pressure, i.e. observed saturation pressure is measured over ice below freezing (assuming equilibrium). Nudging towards a saturated (i.e. 'cloudy') observation (over ice) would then lead to subsaturation (i.e. dissolution of cloud) in the model if vapour pressure was used for the computation of the observation increment without modification. Therefore, observed relative humidity over ice is compared to model relative humidity over water. Equivalently, the observed vapour pressure e_{ob} can be multiplied with an 'ice-to-water' correction to render model compatibility, i.e.

$$e_{ob}^{corr} = e_{ob} \cdot \frac{e_{sat}^{water}(T < T_0)}{e_{sat}^{ice}(T < T_0)} \quad , \quad e_{sat}(T) = 610.78[Pa] \cdot e^{b\frac{T-T_0}{T-T_R}} \tag{3.16}$$

where e_{ob}^{corr} is the corrected value of vapour pressure, T the observed temperature (in [K]), $T_0 = 273.16$[K], and $b^{water} = 17.27$, $b^{ice} = 21.87$, $T_R^{water} = 35.86$[K], and $T_R^{ice} = 7.66$[K] according to the Magnus formulae.

In cloudless cases of supersaturation over ice where supercooled water clouds may be about to form or have just dissipated this may result in a 'bias' of cloudiness, and the corrected 'observed' values of relative humidity may exceed 100 % . Then, this bias is reduced by cutting these values down to 100 % .

3.3.5 Bias Corrections for Humidity

For most types of radiosondes, observed humidity is known to have a dry bias near saturation in general. Hence, saturation is assumed for observed (and possibly ice-to-water adjusted) relative humidity values greater than 96 % irrespective of temperature.

This assumption, however, is not made for data which are read from AOF files and have already been comprehensively bias-corrected before being written to the AOF. Such a bias correction is optionally applied to the reported observation values of relative humidity (over water) $U_{ob\,V80}$ from Vaisala RS80 sondes in two steps (Leiterer et al. (2000)). The first step called weather screen ground check correction,

$$U_{ob}^{corr,1} \;=\; U_{ob\,V80} \;+\; 0.056 \cdot U_{ob\,V80} \tag{3.17}$$

is applied at any temperature, whereas the temperature-dependent second step

$$U_{ob}^{corr,2} \;=\; U_{ob}^{corr,1} + \frac{U_{ob}^{corr,1} \;\cdot\; (0.005{\cdot}(T{-}T_0)^2 \;+\; 0.112{\cdot}(T{-}T_0) \;+\; 0.404)}{100 \cdot e_{sat}^{ice}/e_{sat}^{water} \;-\; (0.005{\cdot}(T{-}T_0)^2 + 0.112{\cdot}(T{-}T_0) + 0.404)} \tag{3.18}$$

is applied only if temperature $T < T_0 - 12[\mathrm{K}]$.

Bias Corrections for Vaisala RS92 Humidity Observations

Relative humidity measurements from Vaisala RS92 radiosondes, which are very widely used in Europe, are well known to have a dry bias at daytime. (Miloshevich et al. (2009)) compared such measurements with reference measurements from cryogenic frost point hygrometer profiles, microwave radiometer data, and calibrated surface-level data. They derived an empirical bias correction in the form of polynomials for the nighttime plus an additional correction for clear-sky conditions at daytime with a reference solar elevation angle $\alpha_{ref} = 62\,\mathrm{deg}$. The second terms corrects for the solar radiation error. The polynomials describe the mean percentage bias (i.e. the relative relative humidity error rather than the absolute relative humidity error) as a function of pressure down to 75 hPa at night and 100 hPa at day. Another empirical function (see Fig. 10 in Miloshevich et al. (2009)) was derived to determine the ratio of the solar radiation error correction at any solar elevation angle α to the correction at α_{ref} .

In the COSMO model, 3 options are available: (0) no bias correction, (1) correction of the solar radiation error only, (2) correction of the full bias. Operationally (at DWD), the full bias correction is applied to humidity observations $< 96\,\%$, i.e. values which have not already been bias corrected to saturation previously. There are different (coefficients for the) polynomials for different given measured relative humidity values. Miloshevich et al. (2009) derived many different polynomials for very low values of relative humidity which have not all been implemented here. For simplicity, the nighttime polynomial for 12 % observed relative humidity is used for all observed relative humidities below 12 % , and similarly, the

solar radiation error polynomial for 5 % humidity is used for all humidities below 5 % . This results in (typically positive) errors in (absolute) relative humidity of at most 0.8 % .

The solar radiation error derived by Miloshevich et al. (2009) is valid only for clear-sky conditions. Therefore, a cloud correction f_{cloud} is computed which scales the clear-sky solar radiation correction error. It is equal to the diffuse transmissivity T^D of the clouds above the observation level for which the correction is computed. $T^D = 1 - R - A$, where R is the cloud reflectivity and A the absorptivity. Compared to the cloud reflectivity, the absorptivity is nearly one order of magnitude smaller (see Fig. 1 in A. (1989)) and is therefore neglected here. This is justified by the fact, that there are very large uncertainties anyway in the cloud estimation based on the radiosonde observations. The transmissivity at zero solar zenith angle can be approximated by $\exp(-LWP/60)$ where LWP is the total liquid water path in the vertical direction in $[\mathrm{g/m^2}]$. For other solar zenith angles ϕ, $T^D = 1-(R_0+(1-R_0)\cdot(\sin^2(\phi/2))$ (where R_0 is the cloud reflectivity at zero solar zenith angle), so that for observation level j,

$$f^j_{cloud} \;=\; T^D_j \;=\; \exp\left(-\sum_{j'=1}^{j-1} LWP_{j'} \;/\; 60\right)\cdot\left(\cos^2(\phi/2)\right) \tag{3.19}$$

where LWP_j is the vertical liquid water path between observation levels j and $j+1$. LWP_j is approximated by

$$LWP_j \;=\; \overline{q_c\cdot\rho}^{\,j,j+1}\cdot\Delta z \;\approx\; {}^1\!/_2\left(q_c^j + q_c^{j+1}\right)\cdot\Delta p \;/\; g \tag{3.20}$$

where q_c is the liquid water content, ρ the air density, Δp and Δz the pressure resp. height difference between observation levels j and $j+1$, and g the acceleration due to the gravity. A given observation level is assumed to be cloudy and contain 'liquid' water content, if the observed vapour pressure exceeds the saturation vapour pressure. Below freezing, a mixed phase region is assumed as in the Tiedtke convection parameterisation, i.e. the water fraction f_w for the mixed phase depends on the temperature T between $T_0 = 273.16\,\mathrm{K}$ and $T^m_{min} = 250.16\,\mathrm{K}$ in the following way: $f_w = ((T - T^m_{min})/(T_0 - T^m_{min}))^2$. For water clouds, i.e. if the observed vapour pressure exceeds saturation above freezing temperature T_0, a cloud liquid water content of 0.1 g/kg is assumed. The same value is used for convective or lower tropospheric ice clouds, whereas for cirrus cloud, a value of 0.01 g/kg is reasonable. Furthermore, it is assumed that the fraction of cirrus cloud relative to the total amount of ice cloud equals 1 above the 300 hPa level, decreases linearly in pressure to zero at 700 hPa and is 0 further below. It results that

$$q_c \;=\; f_w\cdot 0.1 \;+\; (1-f_w)\cdot\left(0.01 + 0.09\cdot\min\left(1,\max\left(0,\frac{p-300}{400}\right)\right)\right) \tag{3.21}$$

where q_c is the cloud liquid water content in [g/kg], and p is the pressure in [hPa].

3.3.6 Gross Error and Consistency Checks

Apart from trivial checks, the following gross error checks are done for individual observations, i.e. observations are rejected if the following conditions are met:

- pressure $p > 1060\,\mathrm{hPa}$
- |pressure tendency| $> 40\,\mathrm{hPa}$ / 3h
- wind speed $> 150\,\mathrm{m/s}$, or wind speed $> 90\,\mathrm{m/s}$ for data below 700 hPa

- temperature $T < -90°C$, or $T > 60°C$,
  or $T > 20°C$ above 700 hPa ,
  or $T > 5°C$ above 500 hPa ,
  or $T > -5°C$ above 400 hPa
- dewpoint $T_d < -150°C$ for upper-air data,
  $T_d < -90°C$ for surface-level data, or $T_d > 40°C$ for any type of data,
  or reported (i.e. uncorrected) relative humidity $U > 120\,\%$

The following consistency checks which are performed after the redundancy checking relate to multi-level data of 1 variable each, and they are based on the ECMWF pre-processing.

Superadiabatic Lapse Rate Check

This check examines whether observed multi-level temperature decreases at unrealistically high rates with increasing height. Considering both significant and standard levels, temperature is converted into potential temperature Θ. If

$$\Theta_{j_1} \quad < \quad \Theta_j^{thr} \quad \doteq \quad \Theta_j - \Theta^{supcor} \tag{3.22}$$

then the layer from j upwards to j_1 is designated to be superadiabatic. The superadiabatic correction Θ^{supcor} is given by Table 3.1 and allows that the dry adiabatic lapse rate may be exceeded particularly at low levels.

The test defined by Eq. (3.22) is applied to any two adjacent temperature observation levels. $\Theta_{j+1} < \Theta_j^{thr}$ denotes that the layer between level j and level $j+1$ is superadiabatic, and further tests are performed to determine whether Θ_j or Θ_{j+1} is in error. If the layer between levels $j-1$ and $j+1$ is superadiabatic and the layer from j to $j+2$ is not superadiabatic, then the observation T_{j+1} is considered to be erroneous. In the opposite case, T_j is rejected, and if no definite conclusion can be drawn then both T_j and T_{j+1} are rejected.

Note that no inversion check is done to identify any unrealistic temperature increase with height. Due to the occurrence of strong inversions particularly during winter, such a check would have to deploy very large threshold values, and it would not result in additional rejection of data compared to the quality control that is performed after the observation processing (see Section 3.6).

Wind Speed Shear Check and Directional Shear Check

Both wind shear checks relate to pairs of adjacent standard levels m and are applied to radiosonde data only. If either of these checks is not passed, then the wind observations at the two standard levels and all the significant levels in between are rejected. The shear of

pressure	>1000	>850	>700	>500	>400	<400
Θ^{supcor}	4.5	3.5	2.5	1.5	1.0	0.5

Table 3.1: Superadiabatic correction Θ^{supcor} in [K] as a function of pressure (in [hPa]) at the lower boundary of the examined layer. (From ECMWF Met. Bull. M1.4/3, 1990 .)

$\lvert d_m - d_{m+1}\rvert$ [degrees]	>30	>40	>50	>60	>70	>80	>90	f^d_{lim}
base level: 700 - 200 hPa	110	84	77	70	63	52	50	7.825
base level: ≥ 850 or ≤ 150 hPa	72	61	57	53	49	46	41	4.5625

Table 3.2: Limit values for the sum of wind speeds $(f_m + f_{m+1})$ in [m/s] for particular directional shear $\lvert d_m - d_{m+1}\rvert$. For f^d_{lim}, see text. (From ECMWF Met. Bull. M1.4/3, 1990.)

wind speed f is considered to be in error if

$$|f_m - f_{m+1}| \;>\; 20.6 \;+\; 0.275 \cdot (f_m + f_{m+1}) \tag{3.23}$$

This means that for stronger winds, larger values for speed shear are accepted. For the directional shear check, a maximum permitted sum of speeds $(f_m + f_{m+1})$ for a particular directional shear $|d_m - d_{m+1}|$ is given by Table 3.2. Note that for wind speeds less than the limit f^d_{lim} (cf. Table 3.2, right column), failing the directional check always implies failing the speed check, and the directional shear check can be omitted ($f^d_{lim} = (f_c \cdot (1-0.275) - 20.6)/2$, where f_c is the maximum permitted sum of speeds for directional shear of 90 degrees).

3.3.7 Redundancy Checking

The redundancy checking is first applied to single-level reports, and this includes the surface-level reports derived from radiosonde TEMP reports. The processing consists of three main steps. Firstly, the conditions for redundancy are evaluated for each pair of reports. In case of redundancy, it is then decided which out of the two reports is redundant. Finally, missing data of the active report may be replaced by available data of the redundant report before the latter is rejected.

The requirements for redundancy are met in case of generalized collocation of two reports, which is specified to be true if

- the two reports are assigned to the same model grid point horizontally,
- the difference in observation time is ≤ 15 min for aircraft reports resp. ≤ 9 min for other reports,
- the vertical distance is < 5 hPa for aircraft reports resp. ≤ 10 m for other reports,
- their station identity is equal; this condition is dropped for GPS reports, since reports from the same station and based on the same raw measurement may have different station identities if processed by different centres,
- and their observation type is equal; this condition is dropped if one report is a TEMP and the other one a PILOT radiosonde report; on the other hand for remote-sensing profile reports (i.e. wind profiler, RASS, and radar VAD), also the code type has to be equal in order to meet the conditions of generalized colocation.

A generalized collocation of two reports with different status (active or passive) is diagnosed only after it has been checked that no generalized collocation exists with a report with equal status. In case of generalized collocation of two reports, one report is selected to be redundant as soon as in the given order, one of the following criteria is satisfied:

- the report has previously been set passive whilst the other report is active,
- for single-level reports only: the other report is of type SYNOP and the current report is not,

- for multi-level reports only: the other report is of type TEMP and the current report is not,
- for multi-level aircraft reports only: the lowest observation level has a higher pressure value in the other report then in the current one,
- for GPS reports only: a zenith total delay value exists for the other report in contrast to the current report,
- for GPS reports only: the processing center of the other report is preferred over the processing center of the current report (according to a preference list given by namelist input),
- in contrast to the current report, the other report is a station correction (according to a flag or update sequence indicator),
- for multi-level radiosonde reports only: the other report has more vertical levels than the current one,
- for GPS reports only: the other report has a smaller (but finite) observation error of total zenith delay,
- the other report has been read prior to the current report from the NetCDF resp. AOF files.

Provided that both reports (i.e. the redundant and the other one) have been active or both have been passive prior to the redundancy checking, then any missing value or observation with passive observation status in the non-redundant report (hereafter called 'active' report) is replaced by the corresponding observation from the redundant report if that observation has active observation status. As exceptions, pressure is never replaced in a single-level report if the redundant report is a surface-level TEMP report, and missing or flagged data of GPS reports are never replaced at all.

For multi-level reports, the process of replacing missing values of the active report by available observations from the redundant report is extended. Specifically, the active report is complemented by the following observation levels or single observations from the redundant report:

- complete observation levels (i.e. including all data from these levels), which are not closer to the nearest observation levels of the active report than 25 hPa for radiosonde reports resp. 5 hPa for aircraft reports,
- all observations with active status from mandatory levels which are more than 1 hPa away from each observation level of the active report,
- other levels which are more than 1 hPa away from the nearest observation levels of the active report – in this case, only those observations are included which have active status in the redundant report and are missing or flagged passive both at the nearest active level above and the nearest active level below, provided that the distance to these levels is less than 25 hPa for radiosonde reports resp. 5 hPa for aircraft reports,
- single data from a level which is not more than 1 hPa away from an active observation level; if there are several levels in the redundant report within this distance to the active level, then the most distant of these levels is selected, and if this level is mandatory then the resulting level is assigned to the mandatory-level pressure; data from the selected level of the redundant report replace the original active values if the latter are missing or their quality flag indicates inferior quality.

Furthermore, if the original active report contains two quasi-collocated levels within $\leq 1\,\text{hPa}$, then the upper level is removed. By applying the whole redundancy-checking algorithm for each multi-level report to itself at first (in which case the criteria for redundancy are always satisfied), quasi-collocated observation levels are eliminated very simply without loosing relevant data. Also, quasi-collocated levels are not created by the complementation of the active report with levels from the redundant report.

Surface levels in multi-level reports are handled in a separate way. If the active report lacks a surface level, the surface level from the redundant report is added. Any levels below the surface level are discarded. The surface level is never complemented with observations from non-surface levels. Furthermore, it is not used to limit the complementation of the active report with (upper-air) levels or observations from the redundant report. As a result, an upper-air level may exist within $\leq 1\,\text{hPa}$ above the surface level.

It is finally noted that the rather extensive manner of complementing active multi-level reports with observations levels and data from redundant reports may be important particularly in cases when a TEMP report containing only temperature and humidity data and a PILOT report have been derived and disseminated from one and the same radiosonde ascent.

3.3.8 Putting TEMP or PILOT parts A, B, C, D into a single profile

TEMP and PILOT radiosonde profiles are often available in the form of 4 different reports, which are also called 'parts' in this context. Parts A and B are for mandatory-level respectively significant-level data from the surface up to $100\,\text{hPa}$, and parts C and D for the corresponding data thereabove. For the data assimilation (and verification), these part have to be put together into a single multi-level report containing the complete profile. This is accomplished by re-running the redundancy check algorithm for radiosonde reports with a revised setting: In the first condition mentioned in Section 3.3.7 for the extended complementation of the active multi-level report with observation levels from the redundant report, the limit of $\geq 25\,\text{hPa}$ is simply replaced by a limit of $> 1\,\text{hPa}$. (Note that this makes the 2^{nd} and 3^{rd} of the four conditions obsolete.)

3.4 Observation Processing of Aircraft Data

A feature common to all of the subsequent processing steps for aircraft data is that they relate only to sets of reports with the same station (i.e. aircraft) identity. Hence, the reports are grouped according to their station identity prior to these steps. Note that on distributed memory machines, this requires the collection of such sets at the same node. And if the processing steps are to be worked in parallel for the different sets, communication steps will be required where each node sends different data to various other nodes and at the same time receives different data from other nodes.

3.4.1 Flight Track Checking

The flight track checking is a model-independent quality control step. Given a complete set of reports with the same aircraft identity, the reports are sorted in a unique way according to their probable chronological order. This is done by sorting them according to time at first.

Groups of 'simultaneous' reports, i.e. reports assigned to same observation time (reported in units of minutes), are then sorted vertically such that the vertical distance to the report prior to the group minus the distance to the subsequent report will increase from one to the next report in this group. Finally, simultaneous and vertically collocated reports are sorted horizontally in an analogous fashion.

Next, the flight phases are determined according to the following criteria. Within a phase to be labeled descent phase, interspersed ascents e.g. related to waiting loops before landing must be smaller than 50 hPa . Consequently, ascents of at least 50 hPa are considered ascent phases. Conversely, interspersed descents must be smaller than 10 hPa within an ascent phase, and descents of at least 10 hPa are considered descent phases. Level flight phases are defined to be whenever there are pairs of subsequent reports above 350 hPa which are at most 3 hPa apart from each other, and they include other reports above 350 hPa between such pairs of reports.

The subsequent checks are performed only if the aircraft identifier is unique, i.e. does not begin with ***, XXX, ???, ///, or BBX.

The first check is about exaggerated horizontal collocation of the reports which may be caused by instrument or transmission errors. All the reports from an aircraft are rejected if more than 50 % of them have the same horizontal coordinates and at least 3 of the collocated reports are more than 12 minutes apart from each other. The latter condition helps to avoid the rejection of aircrafts (e.g. from Lufthansa) which issue identical coordinates for the frequent reports in the first 2 or 3 minutes after take-off (up to about 800 hPa). Exaggerated vertical collocation is diagnosed if at least 50 % or 75 % of the reports have an identical pressure value greater than 500 hPa respectively 350 hPa . These cases also impose the rejection on all reports. No limitation to the vertical collocation is done for reports above 350 hPa to avoid the rejection of flight tracks dominated by the level flight phase.

The actual flight track check is based on an idea following the ECMWF pre-processing (Met. Bull. M1.4/3) to derive a confidence by comparing the reported horizontal position with estimates extrapolated from previous positions. In the present scheme, however, the estimates are computed from previous reports not only with respect to the chronological order, but also to the reverse order. The two resulting confidences are then combined to make the final decision. In this sense, the flight track check is performed both in forward and backward direction. This significantly improves the reliability to reject the erroneous reports and at the same time reduces the probability to reject correct reports due to previous erroneous reports. Furthermore, the scheme is complemented by an analogous, independent vertical check, and an iterative check for missing sign at the reported longitude. The latter is introduced to diagnose more reliably this error, which is relatively frequent with reports that should be located at a longitude between 0 and -1 degrees.

The details are as follows. Given the current position and time c, the positions and times of previous reports with good confidence c_1, c_2, c_3, the previous positions and time with any confidence p_1, p_2, then for each of the combinations (c_2, c_1), (c_3, c_2), (c_3, c_1), (p_2, p_1), an estimate e_i for the current position is computed by temporal linear extrapolation from the pair of predictors. (If the predictors have the same report time (in minutes), then the time differences used for the extrapolation are increased by 1 minute). From the distances d_i^h and d_i^v of each estimate to the current location, horizontal and vertical 'single-estimate' confidences f_i^h resp. f_i^v are computed:

$$f_i^h \quad = \quad \max\left(90 - 40 \cdot d_i^h / d_{ref\,i}^h \;,\; 0\right) \tag{3.24}$$

$$f_i^v = \max\left(90 - 30 \cdot d_i^v / d_{ref\,i}^v\ ,\ 0\right) \tag{3.25}$$

where the reference distances (in units of km resp. hPa) defining the tolerance are given by

$$d_{ref\,i}^h = 30 + 5 \cdot c_i^x + 12.5 \cdot (d_{mi}^t + 1) + d_{mi}^h \tag{3.26}$$

$$d_{ref\,i}^v = 50 + 25 \cdot c_i^x + 50 \cdot (d_{mi}^t + 1) \tag{3.27}$$

d_{mi}^h and d_{mi}^t are the horizontal and temporal distances (in units of km resp. minutes) between the estimate and the more recent of the two predictors. The temporal term allows for a tolerance increasing at a speed of about 550 km/h in the horizontal and 50 hPa per minute in the vertical. c_i^x is equal to $\max\left(d_{mi}^t / d_{mmi}^t\ ,\ 3/(d_{mmi}^t + 1)\right)$ where d_{mmi}^t denotes the temporal distance between the two predictors (if d_{mmi}^t is zero then c_i^x is equal to $\max\left(d_{mi}^t + 1\ ,\ 3\right)$). This term tends to enhance the tolerance if the extrapolation is not well conditioned in the sense that the reported temporal distance between the two predictors is very small in itself or at least small compared to the distance to the current report. The constant term is important for small reference distances, and it is further enhanced by 50 hPa at ascent phases to allow for a 150 hPa decrease within (less than) one minute.

Next, the final horizontal and vertical confidences F^h resp. F^v are defined as a weighted mean of the 'single-estimate' confidences. The weights w_i are smaller than 1 only if the extrapolation is not well conditioned or if it extends over a very long period:

$$w_i = \min\left(\ 1\ ,\ 60/(2d_{mi}^t + d_{mmi}^t)\ ,\ 2/(1 + c_i^x)\ \right) \tag{3.28}$$

If the minimum of the two resulting confidences is decreased by the contribution of the 'single-estimate' confidence deduced from the two possibly erroneous predictors (p_2, p_1) then the weight w_4 given to that 'single-estimate' confidence is set to zero, and the confidences F are recomputed. This reduces the probability of rejecting correct reports due to position estimation from erroneous reports. In an analogous way, a set of confidences is also computed for the backward trajectory, and horizontal 'reversed-sign' confidences for the forward and for the backward trajectory are computed by reversing the sign for every reported positive longitude value.

To further reduce the probability of rejecting correct reports, iterative checks for missing sign at the reported longitude are performed first. Reports that are set passive in previous iterations are not used in the subsequent iterations, and this may alter the confidences derived for the other reports. Specifically, reports are set passive if their confidence is $< 50\,\%$ in the first iteration and $< 65\,\%$ in subsequent iterations and their 'reversed-sign' confidence $> 80\,\%$, until no active report meets these conditions.

Finally, a non-iterative check for other location errors is done separately in the horizontal and the vertical. Provided that both the forward and the backward confidence are available, a report is set passive if one of the two confidences F_{low} is $< 60\,\%$ and the other confidence is $< 85 + (60 - F_{low})/10\,\%$, or if both confidences are $< 65\,\%$. If only one confidence is available (e.g. for the first two resp. the last two reports of the flight), the limit is set to a confidence of $60\,\%$.

3.4.2 Construction of Piecewise Vertical Profiles

During the ascent and descent flight phases, the sequences of single-level reports from individual flights often have a high vertical resolution, at least over western and central Europe.

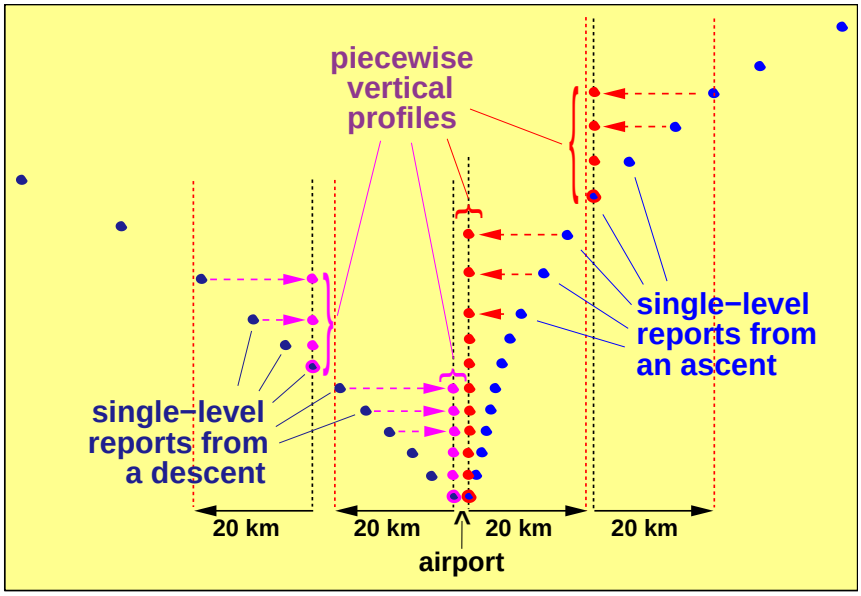

Figure 3.2: Conceptual illustration on constructing piecewise vertical profiles.

In many cases, it is about 10 to 20 hPa in the lower troposphere. Due to the vertical spreading of the single-level increments by applying the vertical weight function, a model grid point can be influenced by many observations from one aircraft, and this results in a vertical smoothing of the observational information. Therefore, most of the single-level reports are grouped into piecewise vertical profiles as part of the observation processing in order to take better advantage of the high vertical resolution. The extent of the resulting multi-level reports is limited such that the temporal and horizontal position errors introduced are not considered significant. The profiles are then assimilated in a way analogous to the nudging of radiosonde profiles.

The practical implementation is as follows (cf. Figure 3.2). Given the collection of all active reports from the same aircraft, the report with the highest pressure value is selected as base report. If there are several reports at the same pressure, the report temporally or horizontally closest to the next report further above will be selected. The observation time and horizontal location of this base report is assigned to the whole multi-level report in construction. Within a temporal radius of 15 minutes and a selectable horizontal radius (set to 20 km in the current operational implementation), at least 3 further reports with strictly decreasing pressure are added, until no reports can be found within the given time interval and horizontal area and within 55 hPa of the previously added report. If no more than 2 reports meet the criteria to be added to the base report the latter will be assimilated as a single-level report. Then the whole process is repeated without all the previous base reports and reports already used for the multi-level report until no report is left.

With the horizontal tolerance radius being set to 20 km, the vertical extent of the lowest multi-level report close to the ground is typically about 100 hPa for the descent phase and 150 hPa for the steeper ascent phase, but it can reach more than 400 hPa if the aircraft reverses its direction during the ascent. Further above, there are often 1, 2, or even more shorter multi-level reports. As a result, most aircraft data are assimilated as part of a multi-level report below 700 hPa and as original single-level report above 400 hPa. Note that the scheme allows to assimilate all aircraft data as single-level reports by setting the tolerance

radius to zero. However, the assimilation as piecewise vertical profiles is computationally far more efficient due to the inherent limitation of the vertical spreading.

3.4.3 Thinning of Aircraft Reports

In order to reduce both the computational costs and the variation of the data density related to quasi-collocated data, sequences of vertically quasi-collocated single-level reports from the same aircraft are thinned horizontally. Specifically, all reports within less than $\pm 5\,\mathrm{hPa}$ and at most 4 minutes from a previous active report are set passive. This mainly applies to data from aircrafts which issue very frequent reports even at flight level. Assuming a regular time interval between these reports, the resulting interval between the active reports after thinning is between 5 and 9 minutes. For a maximum flight speed of $1000\,\mathrm{km/h}$, this corresponds to a horizontal distance of at most $150\,\mathrm{km}$ which is of the order of the 2-folding decay length of the horizontal correlation functions used to spread the observational temperature or wind information in the mid-troposphere and further above. Thus, the data coverage remains very good, and the number of assimilated single-level aircraft reports is reduced by about $10\,\%$.

3.4.4 Reduction of the Vertical Correlation Scale

As already noted, a data source with high vertical resolution should be exploited without much smoothing to allow to correct e.g. the position of inversions. Therefore, sequences of aircraft reports which are at least $5\,\mathrm{hPa}$ apart from each other are not thinned. However, the variation of the effective data density (as well as the computational costs) can be reduced by decreasing the pre-defined vertical correlation scale s_p by a factor w_c for active single-level or multi-level reports which are close to each other. Such a factor

$$w_c \;=\; 1 \;-\; \left[1 - \min\left(\frac{s_p^{red}}{s_p}\;,\;1\right)\right] \cdot e^{-\left[\left(\frac{\Delta_r t}{s_t}\right)^2 + \left(\frac{\Delta_r r}{s_r}\right)^2\right]} \tag{3.29}$$

is computed for each pair of reports. Δ_r denotes the distance between the two reports, s_r the 2-folding decay length of the horizontal correlation function, and s_t half the period of the decreasing part of the temporal nudging weight function. The exponential term is a measure for the horizontal and temporal overlap between the areas of influence of the two reports. If this overlap is small, the factor w_c will approach 1, and if the overlap is large, the correlation scale will tend towards the minimum of s_p and s_p^{red}.

s_p^{red} as given by

$$s_p^{red} \;=\; \frac{\max\left(\Delta_r p\;,\;\Delta_m p\right)}{2\;\overline{p}\;\;\sqrt{ln2}} \tag{3.30}$$

is a reduced correlation scale such that the 2-folding decay height of the related vertical correlation function is equal to the maximum of half the vertical distance between the 2 reports and half the model layer thickness $\Delta_m p$ (in pressure units) at the reports (cf. Figure 3.3). The inclusion of the latter term prevents the correlation scale from becoming smaller than the grid scale even if the vertical distance between the reports is very small. $\overline{p}$ denotes the average of the two reported pressure values.

The reduction factors finally applied to s_p for the upward and downward spreading of a specific report are equal to the minimum of all the factors w_c derived from pairs which

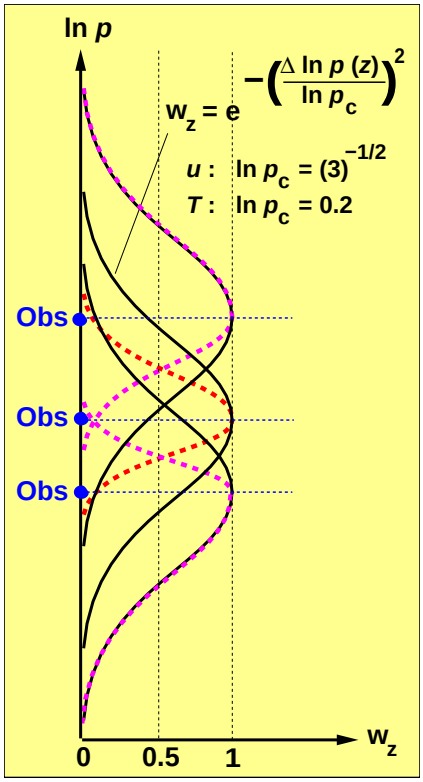

Figure 3.3: Schematic illustration of reducing the vertical correlation scales. The solid black graphs denote the predefined vertical correlation or weight functions w_z as given by the equation included in the figure (cf. Section 3.7.1), whereas the colored dotted graphs show the corresponding functions with the adaptively reduced scales (in pink for the upper resp. lower observation, in red for the observation in between).

consist of this report and another report further above resp. below. Thus, the correlation function used for spreading an observation increment upwards may differ significantly from the function used for spreading it downwards, and it may also vary from report to report. Note again, that this procedure is not only applied to the active single-level reports, but also to the piecewise profiles derived from the same flight.

Related to the reduction of the vertical correlation scale, there is also a reduction (in terms of height or pressure) of the vertical cut-off radius by a factor of $(0.1+0.9\cdot w_c)$. The constant part which provides a lower limit to the cut-off radius in terms of the original cut-off radius helps to prevent the influence of the observations from becoming smaller than the model layer thickness (e.g. if the original cut-off radius is smaller than s_p).

3.5 Observation Increments and Quality Weights

In order to infer corrections to the model fields from observation information, the observations have to be compared to the model state. This comparison is expressed in terms of observation increments, i.e. differences between a value representing the observation and a value representing the model. Either of these two values may be the result of a kind of interpolation, extrapolation, and possibly non-linear function of other values. This section does not only address the derivation of these increments but also the specification of their quality in terms of weights. As the observation increments form the basis for the model corrections, and e.g. determine the direction (sign) in which the model is corrected, attention is given to fair accuracy in the computation of these increments.

Prior to this computation, two lists are produced of all the multi-level resp. single-level reports

which contain observations that are being used for nudging or being quality controlled at the actual model timestep. If the selected temporal weighting is linear interpolation (see Section 3.7.6), then no more than two reports per station are used, i.e. the latest report in the past and the earliest report in the future with respect to the actual model time. In this case, the list of active reports is replaced by a list of active stations, and for each station, there is one or two active reports.

Then, the observation increments are determined separately for the following type of data: surface pressure data, other (local) surface-level data, multi-level data, and upper-air single-level data. They are expressed in terms of pressure, rotated horizontal wind components, temperature, and relative humidity (cf. Section 3.9). For surface-level data, there is also an option for the use of specific humidity increments. From virtual temperature measured by RASS profilers, virtual temperature increments are first computed and then converted into temperature increments after vertical interpolation using model humidity. Thus, perfect humidity is assumed here, and all the errors in (observed and model) virtual temperature are projected onto errors in temperature.

Virtual temperature measured by RASS profilers is converted into temperature

3.5.1 Surface Pressure Increments

As surface pressure is not a prognostic model variable, observational information of station pressure and geopotential has to be provided in terms of observation increments of pressure at the lowest model level. In this and subsequent sections, these increments are often denoted surface pressure increments, and no distinction is made between station pressure and reduced surface pressure data. As extrapolation of full observed or model values to a target level generally tends to be prone to significant errors, it is replaced by extrapolation of increments in combination with interpolation of full values wherever possible.

With respect to single-level data, the situation is considered first where the observation height z_{ob} , for which the observation is valid, exceeds the height z_{k_s} of the lowest model level k_s . In this case, an interpolated model value p_{vi} and hence an observation increment $(\Delta_{ob}p \doteq p_{ob} - p_{vi})$ can be computed at the observation height z_{ob} by interpolating the logarithm of the model pressure linearly in height from the two nearest model levels (k_a , k_b) above and below z_{ob} , i.e.

$$\ln p_{vi} \quad = \quad \ln p_{k_b} \; + \; (\ln p_{k_a} - \ln p_{k_b}) \cdot \frac{z_{ob} - z_{k_b}}{z_{k_a} - z_{k_b}} \tag{3.31}$$

In a second step, the increment has to be transfered from the observation height down to the target level by means of a height correction. Such a correction is necessary provided that the final observation increment at the lowest model level is to yield the increment $\Delta_{ob}p$ at z_{ob} , and that these increments are in hydrostatic balance. Hence, the correction is derived from the hydrostatic equation. Assuming that the simulated virtual temperature is correct between the target level and the observation level, it follows that $\delta_z \ln p$ is identical for the model profile and the 'observed' profile (δ_z denoting the variation with respect to a height interval). Thus,

$$\ln p_{vi} - \ln p_{k_s} \quad = \quad \ln p_{ob} - \ln p_{vob} \tag{3.32}$$

where $\ln p_{vob}$ denotes the observed value extrapolated to the lowest model level. This can

be rearranged to $\ln(p_{vob}/p_{k_s}) = \ln(p_{ob}/p_{vi})$ or

$$\Delta p \doteq p_{vob} - p_{k_s} = \frac{p_{ob}}{p_{vi}} \cdot p_{k_s} - p_{k_s} \cdot \frac{p_{vi}}{p_{vi}} = \frac{p_{k_s}}{p_{vi}} \cdot (p_{ob} - p_{vi}) \doteq \frac{p_{k_s}}{p_{vi}} \cdot \Delta_{ob} p \tag{3.33}$$

so that the height correction f_z is given by

$$f_z \doteq \frac{\Delta p}{\Delta_{ob} p} = \frac{p_{k_s}}{p_{vi}} \tag{3.34}$$

In the other case, where the observation level lies below the lowest model level, an extrapolation of full values cannot be avoided. Thus, the observed value is extrapolated to the lowest model level k_s. With respect to the temperature profile used for this, it is assumed that the model (virtual) temperature at level k_s is correct, and that temperature increases with the logarithm of pressure at the same constant rate β as in the model base state. Then, the relationship used in Part I to derive the base state pressure can be applied, so that

$$p_{vob} = p_{ob} \cdot e^{\frac{T_{v_{k_s}}}{\beta} \cdot \left(1 - \sqrt{1 - \frac{2\,\beta\,g \cdot (z_{ob} - z_{k_s})}{R\;T_{v_{k_s}}^2}}\right)} \tag{3.35}$$

Radiosonde TEMP reports provide the only upper-air data source for geopotential (as a function of pressure) that is used by the nudging scheme. Assuming hydrostatic balance for the total increments (see Sections 3.1 and 3.8.4) and given a profile of temperature increments, there is no degree of freedom for pressure increments except at one level. Hence, the geopotential TEMP data are used to derive only one single pressure increment at the lowest model level.

If there are geopotential observations above and below that level (which implies that the observing station releasing the sonde lies below that level), then the logarithm of the reported pressure is interpolated linearly in height to the height of the lowest model level. Thus, the increment at the target level is obtained directly by a simple interpolation. If the observed profile is very incomplete and there are only observations below the target level, an extrapolation analogous to that for single-level data has be performed. Finally, if there are only observations above the target level (which is the case if the observing station lies above that level), then the increment is computed at the lowest observation level by interpolation of the nearest model values and then transfered to the lowest model level by means of the height correction (3.34).

Quality Weights

Currently, the statistical error related to the observation (type) respectively the quotient of this error to the model error is not taken into account. This is also true for any extrapolation involved in the determination of the observed value. Hence, any station pressure or surface pressure observation of type SYNOP, SHIP, BUOY, or TEMP is assigned the same quality weight in the first place (as long as the observation is accepted by the observation processing checks and passes the quality control). Yet, interpolation errors related to the determination of the increments are roughly accounted for by a Gaussian in a scaled interpolation distance (cf. Section 3.3.3). This scaled distance is defined to be the geometric distance between observation level and the lowest model level scaled by a factor f_D. This factor is equal to 1 except when it is set to 4 for the case that a full value has to be extrapolated, i.e. that the single-level observation lies below the lowest model level. The Gaussian radius z_c is set to 400 m.

An additional weight factor is included for observations from those (SYNOP) reports which also provide three-hourly pressure tendency $\Delta_t p_{ob}$. The factor is 1 for tendencies up to $\pm 3 \frac{hPa}{3h}$ and increases linearly to 1.5 for tendencies beyond $\pm 25 \frac{hPa}{3h}$. It has a positive impact in cases of extreme mesoscale cyclones as it tends to increase the influence of the observations near the cyclone center and in the area where the dynamic forcing is very strong. Note that within the whole scheme, this is the only contribution to a weight that may cause the weight to become larger than 1.

As a result, the total quality weight for surface pressure reads

$$\epsilon \;=\; e^{-\left(\frac{f_D \cdot (z_{ob} - z_{k_s})}{z_c}\right)^2} \;\cdot\; \left[1 \,+\, 0.5 \cdot \min\left(1\,,\, \max\left(0\,,\, \frac{|\Delta_t p_{ob}| - 3\frac{hPa}{3h}}{22\frac{hPa}{3h}}\right)\right)\right] \tag{3.36}$$

3.5.2 Surface-Level Increments

10-m wind and 2-m humidity observations are compared directly to the corresponding screen-level model values as diagnosed by the surface parameterization scheme. In principle, this also applies to 2-m temperature except that the model-derived value is corrected according to a lapse rate and the height difference between the observing station and the model orography. The lapse rate that is used here is derived from the simulated temperature difference between two model levels which are specified to be located about 150 hPa resp. 300 hPa above the ground over low terrain. The observation increments valid for the screen level are then simply assigned to the lowest model level (where they are needed) without any further correction.

(Note that this concept is applied in the current operational version. Yet, a different strategy is adopted for an optional, more simple surface parameterization. With some simplifying assumptions, the observed values can be extrapolated from the screen level to the height of the lowest main model level (relative to the orography) by inverting the parameterization. In this way, the increments are designed to be valid at the target model level. Hence, they should be more realistic as long as the simulated boundary layer is realistic. However, the extrapolation also involves a higher risk of larger errors if e.g. the stability of the simulated boundary layer is significantly erroneous. Generally, it always tends to be less risky to extrapolate increments (in the operational version by simple assignment) than full observed or model values.)

Quality Weights

Operationally, the quality weights are set to 1 for all the active surface-level observation increments. An optional weight for 2-m humidity data is given by a Gaussian in the corresponding 2-m temperature observation increments. Its purpose is to reflect the potential of adverse effects if only humidity is adjusted in the presence of large errors in the simulated near-surface thermal structure.

3.5.3 Upper-Air Increments

The determination of observation increments for upper-air single-level data can be regarded as a special case of that for multi-level data which is addressed here first. Observation increments can either be computed at the pressure level of the observations by interpolation

of model values or at the model levels by interpolation of observed values. The use of these increments depends on the selected type of lateral spreading and other parameters.

Vertical Interpolation to Observation Levels

Increments at the observation levels are always required at the beginning for several reasons. They are deployed for observation quality control (cf. Section 3.6), which should be done prior to any further use of the observations. In addition, they may be needed to compute the increments at model levels, namely for humidity or in the absence of significant-level data.

They are obtained by a simple linear interpolation in logarithm of pressure (denoted log-pressure) of the pair of nearest model values above resp. below the observation level. For quality control purposes only, the two lowest model values are linearly extrapolated in log-pressure to observation levels which are located below the lowest model level. Analogous extrapolations are performed for observation levels above the top model level.

Related to humidity, model values of generalized relative humidity based both on water vapour and cloud water content are interpolated. This may allow the interpolated value to indicate saturation even if the model is saturated only at one out of the two neighbouring levels. A maximum of 100 % relative humidity is finally enforced for quality control and after the interpolation to model levels is finished.

Vertical Interpolation to Model Levels

For the case of a complete TEMP radiosonde report with significant levels, the observed atmospheric state between two adjacent reported observation levels is defined approximately by an linear interpolation in log-pressure of the two observed values. Thus, for the purpose of deriving observation increments, observation values can be interpolated to model levels by such a simple linear interpolation. Yet, there is an alternative method subsequently called vertical scale adjustment. It adjusts the value representing the observations to the same vertical scale for which the model value is representative (Woodage (1985)). That scale is given by the thickness of the model layer. To achieve the scale adjustment, a continuous observation profile, which can be thought to exist by means of linear interpolation between the (mandatory and) significant levels, is averaged over the layer thickness with respect to log-pressure. In other words, the interpolated value ψ_{vob} is a logarithmically mass-weighted sum of arithmetic means of adjacent observed values within the model layer, i.e.

$$\psi_{vob} = \sum_{j=j_b}^{j_a-1} \left[\frac{1}{2} (\psi_{j+1} + \psi_j) \cdot \frac{\ln p_{j+1} - \ln p_j}{\ln p_{j_a} - \ln p_{j_b}} \right] \tag{3.37}$$

where additional 'observed values' ψ_{j_b}, ψ_{j_a} at the lower and upper model layer boundary are always provided by linear interpolation. If a model layer at the top or base of the profile is only partly covered by the continuous observation profile, then the scale adjustment is replaced by linear interpolation for this layer. Note that if the model temperature equals the scale-adjusted observed temperature, then the hydrostatic model pressure difference between the model layer base and top height will be the same as for the observed profile (provided that the model humidity is also correct). In this sense, the vertical scale adjustment method is hydrostatically consistent.

With regard to moisture, relative humidity rather than specific humidity is interpolated. During the interpolation process, any observed values of 100 % relative humidity are replaced

by the corresponding available interpolated model values of generalized relative humidity (including cloud water) whenever they are greater than 100 % . This should yield better estimates of the true total moisture in the model layers and helps to prevent spurious drying (e.g. if the top of a cloud is within a model layer). Finally, a maximum of 100 % is enforced both on the interpolated values representing the observations and on the model values for the purpose of determining the observation increments.

In the absence of significant-level data, care has to be taken with the interpolation of observed values. For perfect observations and model fields, which should imply near-zero increments, and assuming that the true profile deviates significantly from the linearized profile between adjacent standard observation levels, interpolation of full observation values can yield large erroneous observation increments at the model levels in between. Nudging by using these increments tends to linearize the model profiles between the observation levels. This is not the case, if the observation increments at the model levels are obtained by direct interpolation of the observation increments available at the observation levels. Therefore, without modifying the interpolation methods in other respects, interpolation of full observation values is replaced by interpolation of increments if one of the following criteria is met. Either the report is an aircraft report, or there is a big gap (larger than 200 hPa in the troposphere, or 80 hPa above the 220-hPa level) between two arbitrary adjacent observation levels, or there are no significant-level data at all, or there is a big gap (larger than 300 hPa , or larger than 150 hPa below the 650-hPa level) between two adjacent significant levels. Note that the criteria are applied to each variable (wind, temperature, humidity) separately, and that in particular the last criterion may be met for one variable while it is not for another one.

Use of Multi-Level Increments in the Present Model Set-Up

Here, the selection of observation increments to be conveyed to the further processing steps is addressed with a view to the operational version. At first, the vertical interpolation method of observations to model levels has to be specified. At the horizontal location of a multi-level report, the target model grid points (see Section 3.1) coincide with the model levels. Hence, increments are required and considered only at the model levels there. As has already been pointed out, the vertical scale adjustment is hydrostatically consistent. This allows to control the upper-air hydrostatic pressure (and equivalently the geopotential on isobaric surfaces) by nudging of temperature at all levels in combination with nudging of pressure at the lowest model level. With respect to the balancing between wind and mass fields, the vertical scale adjustment appears to be beneficial not only for temperature, but also for wind, and it is therefore used for both variables. For humidity, however, different considerations apply. In view of the ability to introduce sharp humidity gradients and thin cloud layers into the model, vertical smoothing should be kept to a minimum. As the vertical scale adjustment includes averaging, linear interpolation is deployed for relative humidity.

The increments being defined, their selection depends on the specified type of lateral spreading. Where it is along the model levels, i.e. in the stratosphere, only increments at model levels are required. An exception is at the top and base of incomplete profiles, where one observation-level increment is also used (for the vertical spreading). For purely horizontal spreading as applied operationally within the troposphere, the final increment spread laterally to a target grid point is obtained by vertical interpolation of the available increments at the observation location (cf. Section 3.7.2). Using temperature increments at model levels only, this would introduce vertical smoothing in addition to that caused by the vertical scale adjustment. This might be disadvantageous e.g. in the environs of temperature inversions.

On the other hand, the use of observation-level increments could hamper the hydrostatic consistency. Hence, these increments are specified to be used in addition to the model-level increments from the ground up to 800 hPa only, i.e. to the area where most inversions occur. This limitation keeps the possible hydrostatic inconsistencies small. For wind, which tends to have weaker vertical (changes of) gradients, the upper limit is set to 850 hPa . For humidity, finally, it is set to 700 hPa , and furthermore, model-level increments are even discarded below 900 hPa , since thin cloud layers tend to be captured better by observation-level information.

Quality Weights

In preparation of the vertical interpolation weight (Section 3.7.2), which can be seen as a quality weight, an effective interpolation distance Δz_{ie} is assigned to each observation increment. It is computed from the two individual distances Δz_a , Δz_b between the target model (or observation) level, at which the increment is defined, and the nearest observation (resp. model) level above resp. below. The relationship between these distances is the same as that between a total resistance and two parallel resistances, i.e.

$$\Delta z_{ie} \;=\; \left(\frac{1}{\Delta z_a} + \frac{1}{\Delta z_b} \right)^{-1} \tag{3.38}$$

This means that the larger the individual distances and in particular the larger the minimum of these two distances is, the larger the effective distance will get, and the smaller the quality weight related to the observation increment will become. In the absence of significant-level data, when increments instead of observations are interpolated to the model levels, the individual distances themselves are equal to the geometric distances plus the effective distances which have previously been determined at the observation levels using Eq. (3.38). For spreading along isentropic surfaces, the geometric distances are replaced by potential temperature differences, and above the approximate tropopause level (derived from the standard atmosphere), these differences are scaled by a factor of 0.25 to allow for reasonably large quality weights within the strongly stable stratosphere.

Upper-Air Single-Level Data

For upper-air single-level data, increments are provided only at the observation level pressure by a simple linear interpolation in log-pressure of the two nearest model values above resp. below. For observations within the lowest half model layer, an analogous extrapolation is performed. Upper-air observations below the model surface pressure level are discarded. With regard to moisture, generalized relative humidity including the cloud water content is interpolated analogously to multi-level data.

3.5.4 Increments of Integrated Water Vapour

Integrated water vapour (IWV) can be derived from zenith total delay D_t as 'measured' by ground-based Global Positioning System (GPS) stations. While IWV itself is reported by some of these stations, it has to be computed from reported zenith total delay for other stations. This is done by using an algorithm described by Bevis et al. (1994),

$$Q_{ob} \;=\; \frac{D_v}{R_v \cdot (k_s + k_l/T_m)} \qquad , \qquad D_v \;=\; D_t - \frac{2.2765 \cdot p}{1-\epsilon} \tag{3.39}$$

where Q_{ob} is the IWV given in [kg/m^2] ($\cong$ [mm]) and D_t in [mm], D_v is the zenith wet delay (i.e. the path delay due to the atmospheric moisture), p the pressure at the GPS station

in [hPa], $T_m = 70.2 + 0.72 \cdot T$, T the temperature at the GPS station in [K], R_v the gas constant for water vapour, $k_s = 2.21{\cdot}10^{-4}$, $k_l = 3.739$, $\epsilon = 2.66{\cdot}10^{-3}\cos(2\,\phi) - 2.8{\cdot}10^{-7} z$, and ϕ and z the GPS station latitude resp. height. Pressure p and temperature T are interpolated or extrapolated from the model values of the assimilation run to the height of the GPS station.

Provided that the model is run without prognostic cloud ice, an 'ice-to-water' correction is then applied to Q_{ob} similarly to the correction done for direct (point-wise) humidity observations in the observation processing (see Section 3.3.4). The corrected value is given by

$$Q_{ob}^{corr} \quad = \quad Q_{ob} \cdot \frac{Q_{mo}^{water}}{Q_{mo}^{ice}} \tag{3.40}$$

where Q_{mo}^{water} is the IWV value corresponding to the model specific humidity profile (over water). The water-to-ice adjusted value Q_{mo}^{ice} is determined by

$$Q_{mo}^{ice} \quad = \quad \sum \rho{\cdot}\Delta z \cdot q_v(U{\cdot}e_{sat}^{ice}) \tag{3.41}$$

where the sum goes over the model layers, ρ is the (moist) air density, and Δz the thickness of the model layer (resp. its fraction above the GPS station; if the station lies below the model orography then Δz for the lowest model layer is set equal to the layer thickness plus the height difference between orography and station). Specific humidity q_v is converted here from a vapour pressure value that is equal to the model relative humidity U (over water) multiplied by saturation vapour pressure e_{sat}^{ice} that is determined over ice below freezing and over water otherwise.

It is noted that (in contrast to the ice-to-water correction for point-wise humidity data as given by Eq. (3.16)) the correction for IWV as above is not based on saturation vapour pressure alone, but also takes the vertical model profile of relative humidity into account. This is advantageous provided that the vertical structure of simulated relative humidity has smaller errors than a profile of constant relative humidity. For instance, assuming a perfect model state, a perfect IWV observation, and no moisture at all in the layers below freezing, this will result in a zero ice-to-water correction and hence a zero observation increment as wanted.

Finally, the bias correction (Eqs. (3.17), (3.18)) is added optionally. The threshold quality control (Section 3.6.1) is then applied to the resulting IWV 'observation'. Observed IWV of less than $2\,\mathrm{kg/m^2}$ is neglected as in such cases, relatively small observation errors, and in particular an underestimation of IWV, may have strong adverse effects on isolated thin wet layers e.g. related to low stratus.

Use of IWV data

As IWV is not a model variable, the IWV observation $Q_{ob}^{(corr)}$ has to to be converted into 3-dimensional humidity. In other words, the vertically integrated observational information has to be distributed in the vertical. This is done by defining a profile of specific humidity 'observations' $q_{v_{ob}}$ by a simple scaling of a guess for a profile of specific humidity $q_{v_{gs}}$, i.e.

$$q_{v\,ob} \quad = \quad q_{v\,gs} \cdot \frac{Q_{ob}^{(corr)}}{Q_{gs}} \tag{3.42}$$

where the first guess is given by the model values (i.e. $q_{v_{gs}} = q_{v_{mo}}$, $Q_{gs} = Q_{mo}$). If some of the resulting specific humidity values exceed saturation with respect to the model temperature, they are reduced to the saturation specific humidity. The resulting q_v profile and IWV are then used as a new guess for a second application of Eq. (3.42), and so on. The process stops after 20 iterations or when the IWV corresponding to $q_{v_{ob}}$ deviates by less than $0.1\,\%$ from Q_{ob} .

Quality Weights

Very thin or very cold model layers can contribute only little to IWV even if they are saturated. This means that observed IWV gives little indication of the true (relative) humidity in such layers. Consequently, they should not be influenced by IWV information as strongly as thick and warm layers. Quality weights ϵ proportional to saturation specific humidity $q_{v_{sat}}$ and layer thickness Δz are therefore introduced at each model level k such that

$$\epsilon_{\,k} \;=\; \frac{\Delta z_{\,k} \cdot q_{v_{sat}\,k}}{\max_{k'}\left(\Delta z_{\,k'} \cdot q_{v_{sat}\,k'}\right)} \;\leq\; 1 \tag{3.43}$$

3.6 Quality Control

Whereas the preliminary quality checks for observations done within the observation processing (Sections 3.3 and 3.4) are independent from the model, the quality control procedures presented here make use of the model fields by comparing the observations to them. Since the model fields vary in time and each observation is used for relaxation during a certain period, the result of such a comparison may vary in time. Hence, each observation is quality controlled several times. Specifically, it is done at the beginning of the relaxation period of the individual observation, and it is also done for all reports once every 12 minutes. As a result, an observation may pass the quality control at some times and be used in some parts of the relaxation period while it may fail at other times and be discarded correspondingly in other parts of that period.

At a given time, the individual observations are first checked independently from each other. Then, using the result of these tests, some of the observations are subject to further checks which relate to several observations at one time.

3.6.1 Quality Control of Individual Observations

In this check, the deviation of each observation from the model fields is individually compared to a certain threshold value ψ^{thr} . If the deviation given by the corresponding observation increment exceeds the threshold, i.e.

$$\left|\psi_k^{obs} - \psi_F^{n+1}(\mathbf{x}_k, t)\right| \;>\; \psi^{thr} \tag{3.44}$$

then the observation ψ_k^{obs} is considered bad and set passive (i.e. rejected). This concept assumes that the model fields can be regarded as a fair estimate of the truth. The increment used in Eq. (3.44) is always defined at the observation level, except for surface pressure data, where it is approximated by the increment computed at the lowest model level. Related to

pressure	1000	850	700	500	400	300	250	200	150	100	70	50
TEMP v	11.5	11.5	12.5	15.0	17.5	18.5	17.5	17.5	17.0	16.5	16.0	16.0
aircraft v	10.0	10.0	12.0	14.0	16.0	16.0	16.0	16.0	16.0	16.0		
TEMP T	12	10	7	4	4	5	5	6	7	8	8	9
aircraft T	12	10	7	5	5	6	6	7	8	9		

Table 3.3: Operational quality control thresholds for upper-air observations at 50 hPa and below. Pressure levels are in [hPa], wind v in [m/s], temperature T in [K].

wind, the absolute value or speed of the horizontal wind vector increment is checked against a scalar threshold.

It is noted that for frequent observations, this type of check is appropriate only for sudden occurrences of rather large errors. It will usually fail to detect a gradual increase of bias errors from one observation to the next e.g. due to an instrument drift. This is because the increment in Eq. (3.44) remains limited then, unless other observations or strong dynamical or physical forcing prevents the model from adjusting to the gradually increasing error.

As the observations are compared to model values instead of true values, the quality control thresholds should not only take into account the expected observation errors, but also the uncertainty about the model values. The thresholds for upper-air wind and temperature are height-dependent and given by Table 3.3 for TEMP radiosonde and aircraft data. The thresholds for PILOT winds are 90 % of those used for TEMP winds. Zero wind (speed) from aircraft is also rejected if the increment exceeds 5 m/s. As can be seen in Table 3.3, the wind thresholds are largest around the upper-tropospheric jet level. In contrast, the temperature thresholds are large in the low troposphere to allow for modelling errors in cases of strong low-level inversions. Optionally, the thresholds given in Table 3.3 can be multiplied with any constant factor and complemented by non-zero height-independent constant values.

For relative humidity U, the basic threshold is given as function of observation error σ_o and model background error σ_b, which are prescribed as follows:

$$U_{thr\,(1,3)} \;=\; \min\left[\left({\sigma_o}^2 + {\sigma_b}^2\right)^{1/2} \;,\; 2\;\sigma_b\right] \cdot c_{flag\,(1,3)} \tag{3.45}$$

where $\sigma_o = 10\,\%$ ($15\,\%$ for $T_o < 233\,\mathrm{K}$, $20\,\%$ for $U_o < 20\,\%$), $\sigma_b = 10\,\%$ ($15\,\%$ south of $30\,\mathrm{N}$), and the constant $c_{flag\,(3)} = 3.1$ ($c_{flag\,(1)} = 1.8$ for flag 1 as used in the multi-level check, see Section 3.6.4). This formulation is the same as in the global 3DVAR analysis of DWD and leads to a threshold of $44\,\%$ in conditions prevailing in the low and mid-level troposphere over Europe.

However, in strongly stable situations and in particular at inversions, model errors are known to be increased often. The assumed background error σ_b is therefore enhanced by 2 terms selectively for those humidity observations at which the observed lapse rate β to the next humidity observation further above or below is $\beta > \beta_{crit} = -0.0065\,\mathrm{K/m}$:

$$\begin{aligned} \sigma_b \;&\rightarrow\; \sigma_b \cdot (1 + f_{stable} + f_{invers}) \\ f_{stable} \;&=\; 1/4 \cdot (1 - \min(\beta, 0)\,/\,\beta_{crit}) \cdot (1 + c_s) \quad , \quad c_s \;=\; \Delta_\beta T/(1 + \Delta_\beta T) \\ f_{invers} \;&=\; 1/5 \cdot \max(\,\Delta T\,,\,0\,) \cdot (1 + c_i) \quad , \quad c_i \;=\; \min(\,2\,,\,\beta\,/\,0.05\,) \end{aligned} \tag{3.46}$$

$\Delta T = T_k - T_{k-1}$, where T_k and T_{k-1} are the temperature observations at the humidity observation level k respectively at the next level $k-1$ further below. $\Delta_\beta T = T_k^\beta - T_{k-1}$, where T_k^β is T_k extrapolated to level $k-1$ with the lapse rate β_{crit}. Both terms f_{stable} and f_{invers} increase with increasing stability and with increasing thickness of the stable layer (given by the two successive humidity observation levels). Finally, an upper limit of 70 % is imposed. As an alternative to this stability-dependent threshold used operationally, the threshold for relative humidity can also be prescribed by a simple constant.

For the surface-level observations, the threshold values are 12 m/s for SYNOP wind, 6 m/s for DRIBU wind, 12 K for 2-m temperature, and 70 % for 2-m relative humidity. For near-surface pressure, the basic threshold is 4.5 hPa for DRIBU reports and 5 hPa for the other reports except for those SYNOP reports which also issue observed 3-hourly pressure tendency. In this case, it is given by

$$p_s^{thr} \quad = \quad 4\,[hPa] \;+\; 0.6\;\left.\frac{\partial p_s}{\partial t}\right|_{k_{obs}} \left[\tfrac{hPa}{3h}\right] \tag{3.47}$$

The inclusion of pressure tendency reflects to some extent the increased variance of errors in areas of large pressure gradients. In particular, this applies to cases of explosively developing small-scale cyclones, where e.g. moderate phase errors in the model fields produce large pressure errors, and it is reasonable that the resulting thresholds can exceed 15 hPa.

For GPS-derived IWV data, the threshold is set to $0.15 \cdot Q_{sat}$, where Q_{sat} is the IWV corresponding to the model temperature profile and assuming saturation. With this choice, the threshold is expressed in terms of a relative rather than absolute quantity, in analogy to direct (point-wise) humidity data.

In general, the threshold values specified above are valid at the observation time only. As already mentioned, quality control is performed at other times too, e.g. at the beginning of the nudging time window for each observation, when the model state at the observation time is not available. Hence, the observation has to be compared to the model state of a different time, and this state is not as good an estimate for the truth at the observation time. Therefore, the thresholds are linearly enhanced by a factor $f_{\Delta t}$ for each hour of difference Δt_h between model and observation time. In the operational setting, where the equations 3.45 and 3.46 are applied, $f_{\Delta t} = 20\,\%$ for surface pressure, $f_{\Delta t} = 10\,\%$ for wind and temperature, and $f_{\Delta t}^q = 3.4\,\%$ for relative humidity. The rationale behind the small value for humidity is that due to the small-scale variability and large uncertainty of the humidity, the threshold is already large at observation time und should not be strongly increased off the observation time. If, however, equations 3.45 and 3.46 are not used (this is an option that has been used operationally in the past), $f_{\Delta t} = 20\,\%$ for all variables. As a general exception (i.e. independently from the scheme used for the humidity threshold), the pressure thresholds which include observed tendency are kept constant since they already account for time dependency. E.g. for small tendencies, the pressure field can be regarded as nearly steady and is likely to contain only moderate gradients, so that even the model state away from the observation time should provide a good estimate of the truth. Hence, a small threshold of little more than 4 hPa is justified then.

If a temperature observation fails to pass the threshold quality control, then at least either the relative humidity or specific humidity value from the corresponding humidity observation has to be assumed erroneous too. Therefore, humidity is generally rejected in such a case. Furthermore, if aircraft wind or temperature fails the threshold check the whole single-level aircraft report is rejected.

3.6.2 Spatial Consistency Checks

An additional quality control step is performed by checking the horizontal consistency for vertically integrated quantities, namely surface pressure (representing the vertically integrated air mass above the surface) and vertically integrated water vapour (IWV). It is based on the idea to find an improved estimate of the truth by correcting the model value with the help of observation increments in the vicinity. These increments do not have to be valid at exactly the same time as the observation that is quality controlled. However it is important to mention that past or future observations from the same station as the checked observation are not used to compute the improved estimate of the truth at that station location. Thus this check does not consider the temporal consistency of a sequence of observations from a particular station. If the temporal consistency were also taken into account then bad data e.g. from a drifting buoy would be even more difficult to detect in cases where the observed values drift away from the truth slowly and gradually.

Surface Pressure

For this additional check, the individual threshold quality control presented above is refined at first such that each surface pressure observation is assigned to one of four classes labelled 'good', 'probably good', 'probably bad', respectively 'bad'. The three threshold values used to distinguish between the four classes are given by $0.7 \cdot p_s^{thr}$, p_s^{thr}, and $1.5 \cdot p_s^{thr}$, where p_s^{thr} is the original threshold value specified above e.g. by Eq. (3.47). 'Good' observations are let pass the quality control automatically, whilst all the other observations are subject to the consistency check.

At the location $\mathbf{x}_k$ of each of these observations k, a weighted sum $\Delta p_{s_k}^{scc}$ of spread observation increments from all the neighbouring except 'bad' observations j is computed, i.e. (with superscript 'scc' denoting 'spatial consistency check')

$$\Delta p_{s_k}^{scc} = \frac{\sum\limits_{j \neq \{k,bad\}} w_{kj}^2 \cdot \left(p_{s_j} - p_s(\mathbf{x}_j, t) \right)}{\max\left(\sum\limits_{j \neq \{k,bad\}} w_{kj}^2 \;,\; 1 \right)} \tag{3.48}$$

This means that (only) the 'bad' observations are not used for trying to improve the estimate of the truth. Apart from two exceptions related to temporal aspects, the weights w_{kj} are computed in exactly the same manner as for the spreading of pressure increments in the context of computing the analysis increments (see Section 3.7). Firstly, the correlation scale s (see Section 3.7.3) is held constant at 85 km (instead of varying between 70 km at observation time and up to 100 km at other times, cf. Section 3.7.5). Secondly, the asymmetric temporal weight function (see Figure 3.8) used for the relaxation is replaced here for quality control by a symmetric linear function covering a time window of ± 2 hrs. Thus, simultaneous observations obtain a temporal weight of $w_t = 1$, whilst observations from one hour before or later get a weight of $w_t = 1/2$ (i.e. a relative weight of $w_t = 1/4$ after squaring w_{kj} in Eq. (3.48)).

The weighted sum $\Delta p_{s_k}^{scc}$ is used as a correction or bias to the model value to provide a better estimate of the truth. The threshold quality control for individual observations can then be repeated with the new estimate and a modified threshold value, i.e.

$$\left| p_{s_k} - \left(p_s(\mathbf{x}_k, t) + \Delta p_{s_k}^{scc} \right) \right| > p_s^{thr_{scc}} \tag{3.49}$$

On the one hand, the threshold is reduced depending on the sum of weights used to compute the correction $\Delta p_{s_k}^{scc}$. This sum reflects the additional content of information in the new estimate. If it is small, the reduction of the threshold should be small, and vice versa. On the other hand, a large correction or bias $|\Delta p_{s_k}^{scc}|$ tends to indicate increased uncertainty about the estimate. Therefore, the threshold is enhanced by a fraction of the bias and finally reads

$$p_s^{thr_{scc}} \;=\; \left(1 \,-\, 0.2 \cdot \min\left(\tfrac{1}{2} \sum\nolimits_j w_{kj}^2 \;,\; 1\right)\right) \cdot p_s^{thr} \;+\; 0.5 \cdot \Delta p_s^{scc} \tag{3.50}$$

where p_s^{thr} is the time-dependent threshold used in the check for individual observations (Section 3.6.1). It follows that the smallest threshold values of about 3.2 hPa will occur in data-dense areas with small pressure tendency and small increments at the neighbouring stations.

Integrated Water Vapour

The check for integrated water vapour is very similar to the one for surface pressure. Its purpose is to detect a general offset in a radiosonde humidity sounding, or an error in a ground-based GPS zenith path delay measurement. As a first step, observation increments of IWV are derived from radiosonde humidity profiles and (optionally) also from ground-based GPS zenith path delay data. At the location of each IWV 'observation' Q_k, an 'analysis increment' ΔQ_k^{scc} is then computed using only the neighbouring observations $Q_{j \neq k}$:

$$\Delta Q_k^{scc} \;=\; \frac{\sum_{j \neq k} \; w_{kj}^2 \cdot \frac{Q^{sat}(\mathbf{x}_k,t)}{Q^{sat}(\mathbf{x}_j,t)} \cdot \; (\, Q_j \;-\; Q(\mathbf{x}_j,t)\,)}{\max\left(\sum_{j \neq k} \; w_{kj}^2 \;,\; 1\right)} \tag{3.51}$$

Here, $Q^{sat}(\mathbf{x}_k,t)$ is the IWV derived from the model temperature profile at the observation location assuming saturation. The Q^{sat} term scales the observation increment, mainly in order to account for differences in orographic height. This is important since the main contribution to IWV comes from the humidity at low levels. Such a term has been neglected in the pressure check. The weight w_{kj} consists of a horizontal weight (equal to that used for the nudging of radiosonde humidity data at 850 hPa respectively for GPS data), and of a temporal weight (given by a linear function of time within $\pm 2\,h$ respectively $\pm 1\,h$ from the observation time).

The spatial consistency check of IWV is again a revised first guess check, in which the model background is corrected by the above 'analysis increment' in order to obtain a better estimate of truth. The complete humidity profile of the sounding k (or the GPS IWV 'observation' k) is rejected if

$$|\; Q_k \;-\; (Q(\mathbf{x}_k,t) \;+\; \Delta Q_k^{scc})\;| \;\;>\;\; Q_k^{thr_{scc}} \tag{3.52}$$

This check corresponds to a first guess check of IWV if there are no neighbouring observations influencing the observation location $\mathbf{x}_k$. This usually applies approximately if GPS data are not used. The basic threshold Q_k^{thr} depends on temperature and is set to (in [mm]; for the definition of $f_{\Delta t}^q$, Δt_h, see Section 3.6.1):

$$Q_k^{thr} \;=\; \left(\; 1 \;+\; 0.15 \cdot Q^{sat}(\mathbf{x}_k,t)\right) \cdot f_{\Delta t}^q \Delta t_h \tag{3.53}$$

In the presence of many neighbouring IWV observations, however, the check addresses the spatial consistency between them. The more observations are used for the 'analysis increment', the more accurate the estimate of truth, and the smaller the threshold $Q_k^{thr_{scc}}$ should

be set. On the other hand, the larger the 'analysis increment' and hence the disagreement between model and observations, the more uncertain the estimate of truth, and the larger the threshold should be. Therefore, the following correction is applied to Q_k^{thr}:

$$Q_k^{thr_{scc}} \quad = \quad \left(1 \; - \; 0.2 \cdot \min\left(0.2 \cdot \sum\nolimits_j w_{kj}^2 \;\; , \;\; 1\right)\right) \cdot Q_k^{thr} \quad + \quad \Delta Q_k^{scc} \tag{3.54}$$

3.6.3 Check of Surface Pressure against Lateral Boundary Fields

It is already pointed out in section 3.6.1 that during active nudging, the threshold quality control using the model fields as a proxy for the truth often fails to detect a gradual increase of errors from one observation to the next which may occur e.g. due to an instrument drift. This is because the model fields themselves are affected by the increasing errors as a result of being drawn continuously towards the erroneous observations. This process prevails until the dynamics and physics forcing counteracts the erroneous nudging tendencies to a sufficient degree so that the difference between model field and observations become larger than the quality control threshold. Once an observation has been rejected, the errors in the model fields decrase because of the lacking nudging tendendy, and subsequent erroneous data from that station are usually also rejected.

Such a behaviour has been found about once a year on average in the operational COSMO-EU configuration of DWD. Its model domain includes fairly large areas over sea, and each event was caused by a sequence of erroneous surface pressure observations from a single drifting buoy. Since there were no other observations nearby the respective faulty buoy, the spatial consistency check was also not successful enough in detecting these observation errors, even though this was the check that eventually detected the errors first. As a result, surface pressure analysis errors amounted to one or even several tens of hPa. At the same time, operational global analysis systems such as the 3-dimensional variational (3DVAR) scheme of DWD for the global model GME were found to be largely unaffected by this problem. The main reason is that in these systems, which do not deploy observation nudging, the model is not continuously drawn towards the observations. In the free forecast of 3 hours or more from one analysis time to the next, the model dynamics and physics have then enough time to let the model fields evolve away from the observation at the next analysis time, if the observation error is large enough. Moreover, the weight of an observation in the analysis is already reduced by a variational quality control even if the deviation of the observed value from the model first guess field is still somewhat below the limit of the threshold quality control (called first guess check in this context).

It follows that if the fields which provide the lateral boundary conditions for the COSMO run (hereafter called 'LBC fields') are derived from an NWP system using variational data assimilation, they are typically less affected by this type of erroneous observations. As a result, the LBC fields should be a better proxy for the truth than the COSMO fields in this situation.

Therefore, and exclusively for (all) surface pressure data, an additional quality control step is performed, where the model field is replaced by the LBC field both in the individual threshold quality control and the spatial consistency check. By default, the constant part of the thresholds for the original checks is enhanced by a factor of 1.4. For DRIBU reports, this implies a threshold of 6.3 hPa instead of 4.5 hPa at observation time.

This additional quality check tends to tide the COSMO initial conditions somewhat to the

solution of the steering (global) model. However, it is noted that the COSMO model, even with data assimilation, still has the freedom to develop its own, very different solution compared to the global one. The restriction is, that with the additional check, it can do this only by itself (because of different dynamics, physics, resolution, etc.) or by assimilation of observations other than surface pressure. Only where surface pressure data would impose large changes to the analysis, this check tides the COSMO initial conditions closer to the global ones. The increased thresholds in the LBC checks leave some room for the possibility that (particularly for cases of small-scale cyclones with strong pressure tendencies), COSMO can continue to use correct surface pressure data where the global system is likely to reject them.

3.6.4 Multi-level Check and Hydrostatic Thickness Check

The checks presented here relate to multi-level reports only.

Multi-level Check

The so-called multi-level check is applied to each variable (horizontal wind, temperature, humidity, and height) separately. It relies on the following consideration. If a certain amount or fraction of observations has been rejected by the gross error checks or the individual threshold quality control then it is fairly probable that a general problem, e.g. related to the measuring instrument, exists and affects all data of that type. Consequently, the whole profile or at least a part of it is rejected when certain criteria are met.

The rules are as follows. All data (always with respect to one variable) are rejected if at least $50\,\%$ out of totally at least 3 observations have failed in any of the previous checks (in the threshold quality control, but also the observation processing (except for the blacklisting). For radiosondes, furthermore, the same happens if all or 4 subsequent mandatory-level data have failed previously. And if at least $50\,\%$ of all data within 3 subsequent mandatory levels have not passed the checks, then all data within this height range are rejected.

If the stability-dependent humidity quality control thresholds are used (operational setting, see Section 3.6.1) then the following check is added (for wind, temperature, humidity, and height):

- Analysis layers are defined equal to the standard layers except below $700\,\mathrm{hPa}$, where the thickness of the analysis layers is reduced to $50\,\mathrm{hPa}$ and below $800\,\mathrm{hPa}$ to $25\,\mathrm{hPa}$.
- Criterion: If 4 or all consecutive standard layers contain observations with flag ≥ 1, then these standard layers are set to 'rejected'. Each analysis layer within those rejected standard layers is set to 'rejected' if it contains observations with flag ≥ 1. All observations within these rejected analysis layers are rejected.

For wind, temperature, and height, flag ≥ 1 is equivalent to being rejected in the quality check for individual observations, i.e. the thresholds given in Section 3.6.1 are deployed, and there is no distinction between flag set to 1 or to 3 in contrary to humidity. Finally note that rejection of temperature data always implies rejection of corresponding humidity observations.

Hydrostatic Height and Thickness Check

p	1000	850	700	500	400	300	250	200	150	100	70	50	30	20	10
σ_z	4.3	4.4	5.2	8.4	9.8	10.7	11.8	13.2	15.2	18.1	19.5	22.5	25	32	40

Table 3.4: Standard deviation σ_z of radiosonde height error in [m] on standard pressure levels (p in [hPa]). (From ECMWF Met. Bull. M1.5/2, 2000 .)

This check is designed to identify large biases of multi-level temperature data relative to the model fields over considerably large vertical extents. In that sense, it can be regarded as a hydrostatic quality control of indirect observational information on geopotential and thickness. It is applied only to reports which extend over at least $50\,\mathrm{hPa}$.

In order to anticipate the hydrostatic effect of the assimilation of a multi-level temperature observation profile on model geopotential, the profile is extended vertically. Apart from the fact, that only the increments at model levels are used here, it is done in the same way as for the spreading (cf. Section 3.7.1) later on. Specifically, the increment at the base is spread downwards as far as to the cut-off height and is assigned vertical weights. The analogous procedure is applied at the top. Using the exact model-consistent formulation of the hydrostatic equation, the profile of temperature increments is then converted into a profile of pressure increments. If a pressure increment at the lowest model level derived from reported geopotential data is available, it is also accounted for. Next, the pressure increments are converted into a profile of height increments using the standard formulation of the hydrostatic approximation. Finally, a profile of height increments at the temperature observation levels is obtained by linear interpolation in log-pressure.

For long temperature profiles, which are defined not to be enclosed by any three consecutive standard pressure levels (cf. Table 3.4), a hydrostatic threshold quality control of height is done first. This means that the individual height increments are compared against thresholds using Eq. (3.44) as in Section 3.6.1 . The thresholds z^{thr} are given by

$$z^{thr} = \sqrt{\sigma_{z_b}^2 + \sigma_z^2} \cdot \sqrt{L_i} \cdot f_{\Delta t} = \sigma_z \cdot \sqrt{3\,L_i} \cdot f_{\Delta t} = \sigma_z \cdot \sqrt{3\,L_i} \cdot (1 + 0.2 \cdot |t - t_{obs}|) \quad (3.55)$$

This makes use of the assumption that the error variance of the model fields $\sigma_{z_b}^2$ is twice as large as the observation error variance σ_z^2 . σ_z depends on the pressure level of the observation and is given by Table 3.4. $f_{\Delta t}$ denotes the time dependency and describes a linear increase of $20\,\%$ for each hour of difference Δh between model time t and observation time t_{obs} (in [hours]). $L_{i=\{1,2\}}$ is a constant and takes two different values. Only if any height 'observation' fails to pass the check with $L_1 = 10$, then the whole profile will be subject to the thickness check subsequently. The lowest observation that does not pass the check with the use of $L_2 = 20$ is labelled 'probably bad'.

Within the vertical extent of long profiles, thickness is then checked for layers limited by two standard pressure levels with exactly one standard pressure level in between. Two exceptions are made in the lower troposphere so that the lowest four intervals are given by [1000 , 700], [850 , 700], [700 , 500], and [500 , 300] hPa . Additionally, the layer between the lowest temperature observation level and the nearest but one standard level further above is also checked (this standard level is set to $700\,\mathrm{hPa}$ whenever the lowest observation level is below $850\,\mathrm{hPa}$). The same applies to an analogous layer at the top of the profile. To compute thickness increments between standard levels, height increments are interpolated linearly in log-pressure from the observation levels. The thresholds are specified analogous to Eq. (3.55) with σ_z replaced by $\sigma_{\Delta z}$. Thickness error variance $\sigma_{\Delta z}$ can be expressed as a function of

p_m	1000	850	700	500	400	300	250	200	150	100	70	50	30
$C^{z_m,z_{m+1}}$	.716	.733	.573	.851	.814	.935	.919	.895	.861	.929	.951	.878	.881
$C^{z_m,z_{m+2}}$	.276	.183	.268	.480	.601	.738	.678	.597	.649	.782	.710	.598	.426
$C^{z_m,z_{m+3}}$	.029	.055	.077	.288	.364	.458	.361	.375	.460	.480	.412	.192	

Table 3.5: Height error correlation $C^{z_m,z_{m'}}$ between standard pressure levels m and m' of pressure p_m , $p_{m'}$ (in [hPa]). (From ECMWF Met. Bull. M1.5/1, 3rd Ed., 1992 .)

height error variances σ_{z_u} , σ_{z_l} at the upper respectively lower limit of the layer and the correlation C^{z_u,z_l} between the two height errors, so that

$$\Delta z^{thr} \;=\; \sqrt{3\,L_i}\cdot f_{\Delta t}\cdot\sigma_{\Delta z} \;=\; \sqrt{3\,L_i}\cdot f_{\Delta t}\cdot\sqrt{\sigma_{z_u}^2 \;+\; \sigma_{z_l}^2 \;-\; 2\;\sigma_{z_u}\,\sigma_{z_l}\;C^{z_u,z_l}} \tag{3.56}$$

The height error correlations between standard pressure levels m are given in Table 3.5. The correlation C^{z_m,z_j} between a standard level m and an observation level j which is located between standard levels m_1 and m_2 is obtained by linear interpolation in log-pressure of correlations $C^{z_m,z_{m_1}}$ and $C^{z_m,z_{m_2}}$. Analogously, correlations between observation levels can be computed by interpolation of correlations of type C^{z_m,z_j} . In Eq. (3.56), $L_1 = 10$ is used, yet height error variances σ_z^2 above the 100-hPa pressure level are enhanced by a factor of $L_2/L_1 = 2$ to allow for relatively larger stratospheric thickness thresholds. As a result of the choice of all these parameters, the threshold for the mean temperature error within a tropospheric layer of the order of 200 hPa varies around 3.5 K .

If the 'observed' thickness fails to pass this check in certain layers, then all temperature observations are rejected between the base height of the lowest of these layers and the top height of the uppermost failing layer. If a height observation further below has been labelled 'probably bad' in the height check, then the range of rejected observations is extended downwards accordingly. The whole temperature profile at and above such a 'probably bad' height observation is rejected, if there are no layers that fail in the thickness check. Short temperature profiles, which do not meet the criterion for 'long' profiles, are subject only to one thickness check for the layer between the lowest and uppermost active observation.

Finally, it is noted that if any temperature observations have been rejected in the hydrostatic height and thickness check the corresponding humidity (and height) observations are also rejected. And since observed temperature and humidity interpolated to the model levels has been computed previously in order to be used in this check, and since these interpolated values depend on the results of quality control, they have to be re-computed for later use.

3.7 Spreading of Observational Information

This section deals with the way in which the observational information contained in the observation increments and in the quality measures assigned to these increments is spread explicitly to the model grid points (hereafter often called target grid points).

The rationale behind the spreading is related to the fact that it is not sufficient to correct the model only at the location of an observation (see Figure 3.4). If the model value is to be corrected at that location, it must be assumed to have an error prior to the correction. The smoothness of the meteorological fields then suggests that the corresponding model field is likely to contain a similar error in the close vicinity. The confidence about this error will generally decrease with increasing distance from the observation. Hence, the same correction as at the observation location should be made in the vicinity, however it should be weighted by the confidence which is described statistically by the correlation of the errors. In data assimilation schemes, these correlations are often expressed by functions, and in the context of nudging, these correlation functions or 'structure functions' are often called weight functions. The correlation functions largely determine the explicit spatial spreading of the observational information to the target grid points.

The spatial spreading consists of two main steps. At first, incomplete profiles and single-level increments are extended vertically with the outermost available observation increments and provided with vertical weights w_z. The second step is the lateral spreading which comprises of the correct selection of the observation increments and vertical weights for each target grid point and the evaluation of the lateral correlation functions.

With respect to time, it is inherent to the concept of relaxation or nudging, that corrections should be made not only at the observation time but over a certain period of time. The temporal weight function determines the explicit temporal spreading of the observational forcing to the model time.

It should be noted that due to the relatively large uncertainty about the optimal way of spreading and specification of weight functions, very high accuracy is generally not needed in the subsequent computations. And it is mentioned that prior to the spreading, the reports are sorted in a unique order (according to their location, time, and other properties) so as to allow for reproducible results at different domain decompositions in distributed-memory computer environments.

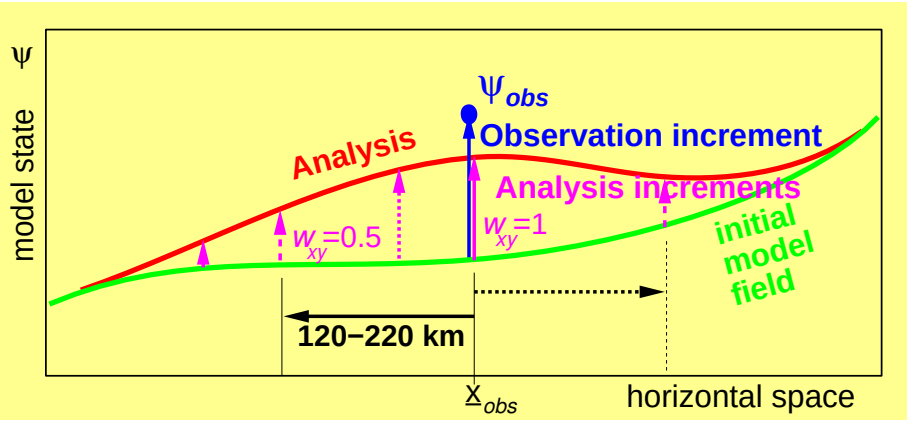

Figure 3.4: Schematic illustration on lateral spreading.

3.7.1 Vertical Spreading at the Observation Location

Upper-Air Data

The weight function deployed for the vertical spreading of upper-air wind is chosen to match well the formulation that is proposed by Barwell and Lorenc Barwell and Lorenc (1985) and used in the former analysis correction scheme of UKMO (Lorenc et al. (1991)). This was a Gaussian in log-pressure differences between the observation increment level and the target model levels. Since the vertical coordinate in COSMO is height instead of pressure, this is approximated by a Gaussian in height differences assuming that the mean virtual temperature between the two levels equals the virtual temperature at the uppermost resp. lowermost increment level, $T_v^{z_{obs}}$. The correlation function then reads

$$w_z = e^{-\left(\frac{[g/(R\,T_v^{z_{obs}})]\ \cdot \Delta z}{\ln p_c}\right)^2} \tag{3.57}$$

where $\Delta z = z - z_{obs}$, and the correlation scale $\ln p_c$ is equal to $1/\sqrt{3}$. The analysis of single-level data benefits from a cut-off on this function outside a radius of influence centred at the observation level. Thus, the vertical weights are set to zero beyond the upper resp. lower cut-off height z_{cut} given by

$$\frac{g}{R\,T_v^{z_{obs}}} \cdot \left| z_{cut} - z_{obs} \right| = c_{cut}^z = 1/2 \tag{3.58}$$

It follows that the value of the correlation function at the cut-off height is equal to $e^{-3/4}$. It is reminded that both the correlation scale and the cut-off heights are adjusted for pairs of aircraft reports which are (vertically) close to each other (see Section 3.4.4).

For upper-air temperature and humidity data, the same correlation function is used except that the scale $\ln p_c$ is set to 0.2 to make it similar to temperature correlations used in other schemes. It confines the influence of these mass field data to a smaller area in the vertical, and this supports the assimilation according to the geostrophic adjustment theory (see Section 3.8.1). By setting $c_{cut}^z = \ln p_c = 0.2$, a smaller fraction of the function is cut off.

Generally, it should be mentioned that the exact specification of the correlation functions and its parameters is not (yet) based on the statistical errors of the COSMO model itself. Although this could and possibly should have been done, the task to do this has never had high enough priority in the past. The main reason for that is that the performance of the nudging scheme is significantly less dependent on the precise shape of these functions than 3-dimensional analysis techniques such as Optimum Interpolation or 3DVAR.

It is finally noted that the vertical correlation functions are expressed as a Gaussian of potential temperature differences if the observation increments are optionally spread laterally along isentropic surfaces (see Section 3.7.2). Above the approximate tropopause level at 225 hPa, the Gaussian radius used elsewhere is enhanced by a factor of 4 to yield a relatively smooth transition of the vertical weighting from the troposphere to the stratosphere.

Surface-Level Data

In the basic formulation for surface-level data, the same type of vertical correlation function (i.e. Eq. (3.57), or a Gaussian of potential temperature differences for the optional spreading along isentropic surfaces) is used as for upper-air data. For 10-m wind, the cut-off parameter

c_{cut}^{z} is also set to $1/2$, and the correlation scale is chosen such that the cut-off height z_{cut} and range of substantial direct influence is about 850 m. For 2-m humidity, setting $\ln p_c = \sqrt{0.002}$ yields an e-folding decay height of about 300 m. However, due to problems related to the surface-layer parameterization and the limited representativeness of the data, the original cut-off height of about 600 m is reduced to the lowest model level in the current operational version.

In addition to the basic weight as a function of height, an additional weight factor reflecting the stability in the planetory boundary layer (PBL) can be introduced optionally. This 'PBL weight' consists of a Gaussian in the difference $\Delta\Theta^{PBL}$ between the potential temperature at the target level and the minimum of the potential temperature at the observation level and at the lowest model level. The inclusion of that minimum yields very small weights in strongly stable cases when the ground is largely decoupled from the atmosphere and the surface-level observations are of a very limited representativeness.

To avoid the formal spreading of increments to areas where the vertical weights are very small, the cut-off height should also be adjusted. This is done in such a way that at the new cut-off height z_{cut}^{PBL}, the total vertical weight w_z^{cut} takes the same value as the basic weight (without PBL weight) at the original cut-off height. This makes sense particularly if that value is small, i.e. if even smaller values can be neglected. The above relationship reads

$$w_z^{cut} = e^{-\left(\frac{\Delta z_{cut}^{PBL}}{z_s}\right)^2} \cdot e^{-\left(\frac{\Delta\Theta_{cut}^{PBL}}{\Theta_s^{PBL}}\right)^2} \doteq e^{-\left(\frac{c_{cut}^{s}}{\ln p_c}\right)^2} = e^{-\left(\frac{\Delta z_{cut}}{z_s}\right)^2} \tag{3.59}$$

where z_s is the correlation scale equal to $[(R T_v^{z_{obs}}/g] \cdot \ln p_c$, and Θ_s^{PBL} is the correlation scale of the PBL weight. It follows that

$$(\Delta z_{cut})^2 = \left(\Delta z_{cut}^{PBL}\right)^2 + \left(\frac{z_s}{\Theta_s^{PBL}}\right)^2 \cdot \left(\Delta\Theta_{cut}^{PBL}\right)^2 \tag{3.60}$$

This is an implicit equation for the new cut-off range Δz_{cut}^{PBL} since $\Delta\Theta_{cut}^{PBL}$ also depends on it. Δz_{cut} can be regarded here as an effective distance which is made up of the geometric distance Δz_{cut}^{PBL} and a kind of thermal distance. Equation (3.60) can be solved for Δz_{cut}^{PBL} approximately by computing this kind of effective distances $\Delta z_{eff}(k)$ at every model level k, by finding the model levels k_c and $k_c - 1$ for which $\Delta z_{eff}(k_c) \leq \Delta z_{cut} \leq \Delta z_{eff}(k_c - 1)$, and by interpolating linearly between these two levels so that the interpolated effective distance is equal to Δz_{cut}. The geometric interpolation level is then set to the new cut-off height.

As already mentioned, this approach makes sense if the original basic weight at the original cut-off height is small. That is the case for humidity for which this weight takes the value $e^{-(\Delta z_{cut}/z_s)^2} = e^{-4}$. For wind, however, it is $e^{-3/4}$, i.e. rather large, and therefore, the above computation is done by setting $\Delta z_{cut} = 2 z_s$ in Eq. (3.60). The final cut-off height is then defined to be the minimum of the resulting Δz_{cut}^{PBL} and the original Δz_{cut}. This means that the cut-off height for 10-m wind data is reduced only in rather strongly stable cases for which the final weight would become smaller than e^{-4} at the original cut-off height.

It is mentioned finally that for the optional spreading along isentropic surfaces, height z is replaced by potential temperature Θ everywhere in Eqs. (3.59) and (3.60). Thus, Eq. (3.60) can be solved analytically for the new cut-off range $\Delta\Theta_{cut}$:

$$\Delta\Theta_{cut}^{PBL} = \Delta\Theta_{cut} \cdot \left[1 + \left(\frac{\Theta_s}{\Theta_s^{PBL}}\right)^2\right]^{-1/2} \tag{3.61}$$

3.7.2 Basic Types of Lateral Spreading

As mentioned in the introductory remarks of Section 3.7, the lateral spreading comprises of two parts. The current section addresses the selection of the observation increments and vertical weights for each target grid point, whereas the subsequent Sections 3.7.3 to 3.7.5 deal with the lateral correlation functions used to define the lateral weights.

In nudging-type schemes, the conventional way to spread or extrapolate the observation increments laterally to (target) model grid points is along the model levels. For terrain-following coordinates as deployed in COSMO, this can have adverse effects near steep orography, particularly in cases with low stratus stratus associated with an inversion (Schraff (1997)). For instance, low-level radiosonde data increments can be spread along the sloping surfaces from low terrain upwards across the inversion to high areas which are largely decoupled from the observation location. Therefore, two other, more expensive types of lateral spreading are also implemented in the COSMO scheme. This is spreading along purely horizontal surfaces, and spreading along isentropic surfaces. In particular on isentropic surfaces, flow patterns are spatially and temporally more coherent, and the isotropic correlation assumption (see Section 3.7.3) is better founded (Benjamin (1989)). Both types of spreading significantly reduce the adverse effects mentioned above.

In the operational setups of COSMO, upper-air data are spread along horizontal surfaces, (partly) because isentropic spreading has turned out to be too expensive. In contrast, surface-level data, i.e. the small pieces of vertical profiles of weighted increments just above the ground as resulting from vertical spreading, are spread along the model levels. This confines the direct influence of these data to close to the ground, and hampers the spreading of observational information from elevated mountain surface stations to the free atmosphere over valleys or basins such as the Swiss Plateau or Po Valley. Note that the uppermost few model levels of the operational COSMO setups are horizontal, and spreading along model levels is exactly horizontal there.

Whenever the lateral spreading is along model levels, the observation increments and vertical weights are computed at the model levels. For each target grid point, the increment from the same model level is then selected together with its vertical, temporal, and quality weight, and spread to the target grid point by assigning a lateral weight (see Sections 3.7.3 to 3.7.5) to it.

For spreading along horizontal surfaces, the observation increment for any given target grid point is related to the increment at the same height at the horizontal observation location. This means that for a single-level observation, the vertical weight is determined by applying Eq. (3.57) to the vertical distance between the observation point and the target grid point. For surface-level observations, this can be complemented by the corresponding stability-dependent weight. The increment, the quality weight of the increment, and the temporal weight are then assigned to the target grid point together with this vertical weight and with a lateral weight.

For (incomplete) multi-level profiles of observation increments, the same procedure is applied to the uppermost or lowermost increment if the target grid point lies above the top level resp. below the base level of the profile. For any target grid point in between, the nearest observation increments above and below are interpolated linearly in height to the target height level. The vertical weight which can otherwise be regarded as an extrapolation weight is replaced here by an interpolation weight. It is determined by applying Eq. (3.57) with

the same Gaussian radius to a total effective interpolation distance Δz_{tie}. This distance is computed from the two individual effective distances between the target level and the nearest increment level above Δz_{ae} resp. below Δz_{be} in the same way as a total resistance from two parallel resistances, i.e. $1/\Delta z_{tie} = 1/\Delta z_{ae} + 1/\Delta z_{be}$. Each of these two individual effective distances is the sum of the corresponding geometric distance (in log-pressure units) plus an effective quality distance assigned to the observation increment (cf. Section 3.5.3). Note that an analogous Gaussian weight which is based only on the effective quality distance is also applied for single-level data and for spreading along model levels. A lower limit of 0.2 for the log-pressure Gaussian radius is used in the context of these effective distances. (This avoids too strong a reduction of weights due to interpolation even if the Gaussian radius used for the vertical weights is very small.)

Spreading along isentropic surfaces is basically analogous to horizontal spreading if the height levels and vertical distances are replaced by model potential temperature surfaces and distances. However, a disadvantage to isentropic coordinates is that the vertical resolution decreases as the atmospheric stability decreases. Hence in case of a near-neutral stratification either at a radiosonde location or at a target grid point, a small error in the simulated temperature could result in a large error in the height level z_Θ at which the increment is determined. Moreover, if the model atmosphere exhibits an (unconditional) instability at the radiosonde location, z_Θ may even be ambiguous because the same potential temperature values may occur several times within the profile. To alleviate these problems, the following strategy is adopted. Firstly, within the vertical interval of ambiguity, the model potential temperature profile at the radiosonde location is replaced (for the purpose of finding z_Θ only) by a stable profile where potential temperature varies linearly in log-pressure between the range of values that have been ambiguous in the original profile. Secondly, if the stability at the target grid point is below a threshold (chosen such that a temperature error of 1 K corresponds to a height error of about 500 m for typical low-tropospheric conditions), the observation increment for the target grid point is a weighted linear combination of two increments computed by spreading along isentropic resp. horizontal surfaces. For neutral or instable conditions, the spreading is performed exclusively horizontally. At high levels (i.e. above about 220 hPa), the observation increments are also spread horizontally along the model levels. In all these cases, however, the vertical weight function always remains a Gaussian in potential temperature differences.

It is also mentioned that prior to the spreading, the temperature observation increments are converted to potential temperature increments which are conserved quantities with adiabatic flow along isentropes. This would preserve the shape of the isentropic surfaces if the (lateral) weights did not vary.

p	1000	850	700	500	400	300	250	200	150	100	50
s_T , s_q	58	66	75	83	83	83	83	91	100	100	100

Table 3.6: Correlation scales for temperature s_T and humidity s_q in [km] at the observation time as a function of pressure p (in [hPa]).

3.7.3 Lateral Weights for Scalar Quantities

Basic Correlation Function

In the basic formulation, the lateral weighting for scalar quantities is assumed to be homogeneous and isotropic. The correlation function deployed is a simple second-order autoregressive function of the distance Δr between the observation location and the target grid point,

$$w_{xy} \quad = \quad (1 + \Delta r/s) \cdot e^{-\Delta r/s} \tag{3.62}$$

For small separations $(\Delta r/s)$, w_{xy} falls off a little faster with distance than a Gaussian, and this tends to yield a better resolution of small-scale details. And its longer tail is of benefit in data-sparse areas. Hence, this function has been widely used in nudging-type schemes.

The correlation scale s for upper-air data is a function of time and of pressure level. As indicated in Table 3.6, the value for temperature and humidity valid at the observation time increases from about 60 km in the boundary layer to 83 km in the middle to upper troposphere and 100 km in the stratosphere. Hence at 500 hPa, $w_{xy} = 1/2$ is at a distance of about 135 km. Optionally, s_q as used for radiosonde humidity (see Table 3.6) is scaled by a factor of e.g. 45 % for the humidity profiles derived from GPS IWV data in order to account for their high horizontal data density. Generally, all the correlation scales for upper-air data are linearly enhanced with increasing difference between observation time and model time by a maximum amount of 30 %. Since the temporal weight decreases with increasing distance from the observation time (cf. Section 3.7.6), it follows that any observation enters the assimilation with low weight affecting rather large scales. As the model time approaches the observation time, the observation affects smaller scales with high weight. This strategy which is also adopted by standard successive correction analysis schemes has the effect to speed up convergence, and tends to compensate for deficiencies in the correlation functions.

Analogous to the vertical correlation functions, the lateral correlation functions are also cut off at some distance. The cut-off radius which determines the horizontal radius of direct influence of an observation is simply a fixed number of correlation scales, namely $2 \cdot s$ for humidity and $3.5s$ for the other quantities. As a result, it is also time-dependent for upper-air observations (in terms of km).

It is noted that the area of influence is determined for each observation in a step-wise way for the sake of efficiency. At first, the minimum upper limit for the vertical range of influenced model levels is established (both on the total and on the local domain in distributed-memory applications). Then, depending on it, the minimum upper limits for the longitudinal and latitudinal range of grid points are provided for the total report, for the individual variables on all levels, and finally for the individual variables on the individual model levels prior to the exact determination of the grid points that are within the area of influence.

Non-Isotropic Correction

The problem of reduced representativeness that has been mentioned in Section 3.7.1 and occurs particularly for spreading along the sloping model levels can also be addressed by complementing (i.e. multiplying) the isotropic lateral weight with a non-isotropic correction weight w^{ni}. In the present scheme, this weight is defined to be a Gaussian in potential temperature differences between the target grid point $\mathbf{x}$ and a certain point $\mathbf{x}_k^{ni}$ at the horizontal location of the observing station k, i.e.

$$w^{ni} = e^{-\left(\frac{\Theta(\mathbf{x})-\Theta(\mathbf{x}_k^{ni})}{\Theta_s^{ni}}\right)^2} \tag{3.63}$$

In the vertical, point $\mathbf{x}_k^{ni}$ is defined to lie on the surface which is used for the lateral spreading to target point $\mathbf{x}$. This means that $\mathbf{x}_k^{ni}$ is on the same model level as $\mathbf{x}$ for spreading along model levels resp. on the same height level as $\mathbf{x}$ for horizontal spreading. While providing potential temperature at $\mathbf{x}_k^{ni}$ is simple for spreading along model levels, it requires some vertical interpolation for horizontal spreading. In the latter case, a profile of potential temperature is first computed at the observation location throughout the atmosphere at vertically equidistant points. (The equidistance is measured in approximate pressure units.) This allows to directly identify the appropriate levels used to interpolate potential temperature to the height of $\mathbf{x}_k^{ni}$. The same strategy is adopted for isentropic spreading, except that potential temperature is replaced by height e.g. in Eq. (3.63), and that point $\mathbf{x}_k^{ni}$ lies on the same isentropic surface as target grid point $\mathbf{x}$.

Compared to spreading along model levels with isotropic weights, horizontal or isentropic spreading alleviates the problem of reduced representativeness related e.g. to steep terrain by a vertical modification and shift of the area of main influence of a single observation. In contrast, the non-isotropic correction tends to alleviate it by restricting the area of main influence in the horizontal. By using potential temperature instead of height differences in Eq. (3.63), the restriction depends on the stability of the atmosphere. Hence, the spreading of information across an inversion is diminished selectively. Reduced representativeness of observations across air mass discontinuities over flat terrain can also be accounted for. However, it has been found to have a tendency to reduce too much the influence of radiosonde data in the Alpine region (Schraff (1997)). This is why in the operational version, upper-air data are spread horizontally without non-isotropic correction. However, for surface-level humidity which is spread along the model levels, the correction is applied with a Gaussian radius of $\Theta_s^{ni} = 3\,\mathrm{K}$.

Use of the Tangent Cone Projection

Given the fact that Eq. (3.62) models the true, unknown error correlations only imprecisely, the distances Δr do not have to be computed very accurately. Since the area of influence of each individual observation is rather limited, the latitude-longitude displacement $(\Delta\phi, \Delta\lambda)$ from the observation point (ϕ, λ) to any target grid point is small as long as the observation is not located too far from the equator of the rotated latitude-longitude grid of COSMO. If the model domain is large or far away from the equator, however, $\Delta\lambda$ can become large near the pole. Projection from the centre of the earth onto a cone tangent to the earth at latitude ϕ gives fairly simple, globally valid equations (Bell et al. (1996)). For $\Delta\phi \ll 1$ and $\phi = 0$, this projection is equivalent to the latitude-longitude projection.

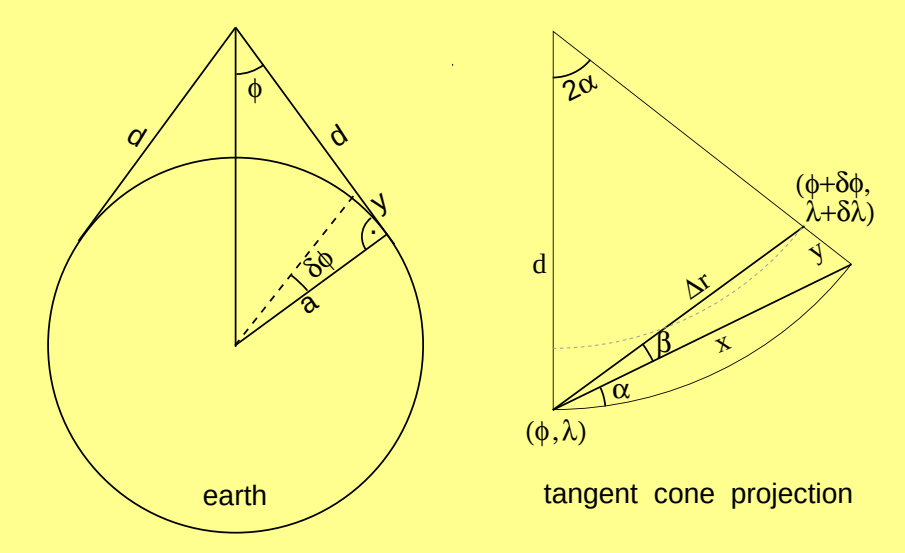

Figure 3.5: Geometry underlying the distance formulae.

The geometry relevant is portrayed in Figure 3.5. In the projection, circles of latitude become arcs with radius $d = a\,\cos\phi/\sin\phi$ where a is the radius of the earth. The arc from (ϕ,λ) to $(\phi,\lambda{+}\Delta\lambda)$ has a length of $\sin\phi\; d{\cdot}\Delta\lambda = a\,\Delta\lambda\,\cos\phi$ and subtends an angle 2α in the projection. α is given by

$$\alpha \;=\; {}^1\!/_2\; a\;\cos\phi\;\Delta\lambda\;/d \;=\; {}^1\!/_2\;\sin\phi\;\Delta\lambda \tag{3.64}$$

and the length x in the projection from (ϕ,λ) to $(\phi,\lambda{+}\Delta\lambda)$ by

$$x \;=\; 2\,\sin\alpha\;\; d \;=\; 2\,\sin\alpha\;\; a\;\;\cos\phi/\sin\phi \tag{3.65}$$

(It follows that for ϕ near 0, $\sin(\sin\phi) \cong \sin\phi$, and $x = a\cos\phi\;\Delta\lambda$ as in the latitude-longitude projection.) The arc from $(\phi\,,\,\lambda{+}\Delta\lambda)$ to $(\phi{+}\Delta\phi\,,\,\lambda{+}\Delta\lambda)$ has the length $a{\cdot}\Delta\phi$ on the sphere and $a{\cdot}\tan(\Delta\phi)$ on the projection which can be approximated by $a{\cdot}\Delta\phi$ since $\Delta\phi$ is small. I.e., the length y of the equivalent line in the projection is given by

$$y \;=\; a\;\Delta\phi \tag{3.66}$$

Finally, applying the cosine rule to the triangle in the projection with sides x , y , and Δr defining the length from (ϕ,λ) to $(\phi{+}\Delta\phi\,,\,\lambda{+}\Delta\lambda)$ yields

$$(\Delta r)^2 \;=\; x^2 \;+\; y^2 \;-\; 2\;x\,y\;\sin\alpha \tag{3.67}$$

3.7.4 2-Dimensional Horizontal Wind Correlations

If the lateral weights used for the spreading of a single horizontal wind vector observation (increment) were computed for each wind component separately by use of a one-dimensional function then each wind vector of the resulting analysis increment field would have the same direction (Figure 3.7 a). This field would contain a strong divergent component. It is desirable, however, that the observational wind information mainly corrects the rotational component of the model wind field. In optimum interpolation, rotational increments (approximated

in Figure 3.7 b) are obtained by deriving wind covariances from a streamfunction covariance. This approach is also adopted here. A small modification to the non-divergent correlations is sufficient to account for divergent structures seen in the observations (Daley (1985)).

Assuming homogeneity and isotropy, the simplest way to specify the correlations is in terms of wind components along and perpendicular to the direction from the observation point to the target grid point (Bell et al. (1996); see Figure 3.6). Assuming non-divergence, the longitudinal w_{xy}^{LL} and transverse w_{xy}^{TT} correlation components can be expressed as a function of the correlation for streamfunction $w_{xy}^{\Psi\Psi}$,

$$w_{xy}^{LL} = -\frac{1}{\Delta r}\frac{\partial}{\partial \Delta r} w_{xy}^{\Psi\Psi} \tag{3.68}$$

$$w_{xy}^{TT} = -\frac{\partial^2}{\partial(\Delta r)^2} w_{xy}^{\Psi\Psi} = w_{xy}^{LL} + \Delta r \frac{\partial}{\partial \Delta r} w_{xy}^{LL} \tag{3.69}$$

$$w_{xy}^{LT} = w_{xy}^{TL} = 0 \tag{3.70}$$

From Eq. (3.69), it follows that w_{xy}^{TT} can be expressed as a function of w_{xy}^{LL} , and that $\int_0^\infty w_{xy}^{TT}\, d(\Delta r) = 0$ (as long as $\partial_{\Delta r} w_{xy}^{\Psi\Psi}$ is zero at $\Delta r = 0$ and approaches zero for very large Δr). Hence, w_{xy}^{LL} can be regarded as the basic wind correlation function, and the second term on the right side of Eq. (3.69) can be seen as a non-divergent correction. Without this correction, the two-dimensional correlation function would collapse into two separate one-dimensional correlation functions for each wind component (at least on a longitude-latitude projection).

In many assimilation schemes, in particular those based on optimum interpolation, the errors are assumed to be geostrophic. This is adopted here in the sense that the (non-divergent) wind errors as indicated by wind observations are assumed to be in geostrophic balance to the geopotential errors. Consequently, the correlation function for streamfunction $w_{xy}^{\Psi\Psi}$ must have the same form as for geopotential resp. for surface pressure and temperature, and is therefore given by Eq. (3.62). From Eqs. (3.68) to (3.70), it then follows that

$$w_{xy}^{LL} = e^{-\Delta r/s} \tag{3.71}$$

$$w_{xy}^{TT} = e^{-\Delta r/s} - \gamma_n \cdot (\Delta r/s) \cdot e^{-\Delta r/s} \tag{3.72}$$

A small modification has been made by introducing a multiplicative, so-called non-divergence correction factor γ_n to the non-divergence correction, with $(0 \le \gamma_n \le 1)$ (Lorenc et al. (1991)). This allows to relax the non-divergence constraint if desired.

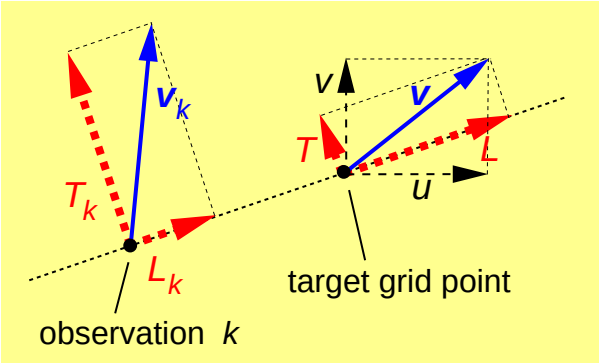

Figure 3.6: Longitudinal and transverse wind velocity components.

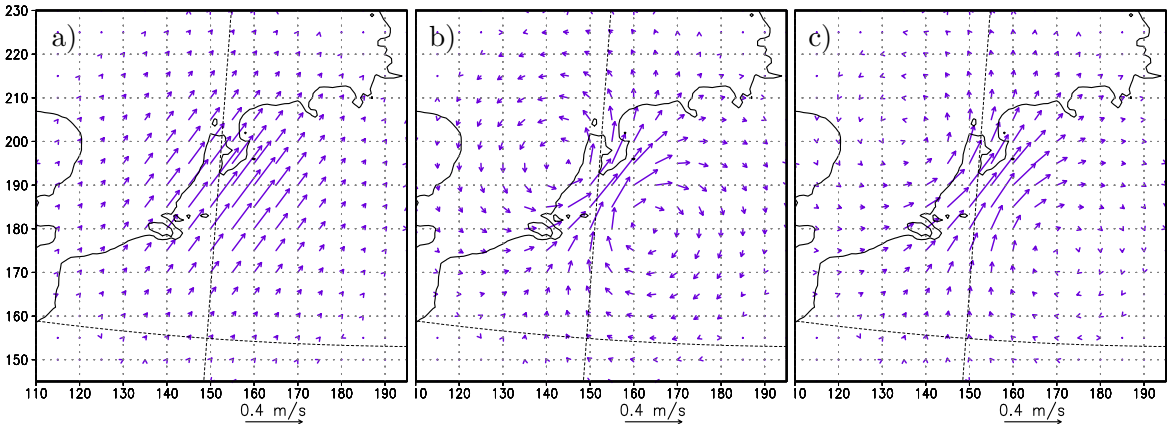

Figure 3.7: Analysis increments of horizontal wind given a single observation increment of 10 m/s at grid point (153,188) in 500 hPa. a): non-divergence correction factor $\gamma_n = 0$ (divergent); b): $\gamma_n = 1$ (non-divergent); c): $\gamma_n = 0.5$ (semi-divergent).

The correlation functions as given by Eqs. (3.70) to (3.72) relate the longitudinal and transverse wind errors or increments at the target grid point $(\Delta L,\, \Delta T)$ to the error resp. observation increment $(\Delta L_k,\, \Delta T_k)$ at the observation point. This relationship can be written in vector form as

$$\begin{aligned} \begin{bmatrix} \Delta L \\ \Delta T \end{bmatrix} &= \begin{bmatrix} w_{xy}^{LL} & w_{xy}^{LT} \\ w_{xy}^{TL} & w_{xy}^{TT} \end{bmatrix} \cdot \begin{bmatrix} \Delta L_k \\ \Delta T_k \end{bmatrix} \\ &= w_{xy}^{LL} \cdot \begin{bmatrix} \Delta L_k \\ \Delta T_k \end{bmatrix} - \gamma_n \cdot \frac{\Delta r}{s} \cdot w_{xy}^{LL} \cdot \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \Delta L_k \\ \Delta T_k \end{bmatrix} \end{aligned} \tag{3.73}$$

In the COSMO model, horizontal wind is expressed as eastward component u and northward component v. In order to use Eq. (3.73), it is necessary to transform these by a rotation of coordinates to the longitudinal and transverse components. The transformation is given by

$$\begin{bmatrix} \Delta L \\ \Delta T \end{bmatrix} = \begin{bmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{bmatrix} \cdot \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix} \doteq \mathbf{R}(\vartheta) \cdot \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix} \tag{3.74}$$

where $\mathbf{R}$ is the rotation matrix.

On the tangent cone projection, the rotation angle ϑ at the observation point (ϕ, λ) is equal to $\alpha+\beta$, where α is given by Eq. (3.64) and β is the angle between the line (ϕ, λ) to $(\phi, \lambda+\Delta\lambda)$ and the line (ϕ, λ) to $(\phi+\Delta\phi,\, \lambda+\Delta\lambda)$. Triangular geometry (see Figure 3.5) yields

$$\cos\beta = (x - y\,\sin\alpha) \,/\, \Delta r \quad ; \quad \sin\beta = y\,\cos\alpha \,/\, \Delta r \tag{3.75}$$

The rotation angle for the same kind of transformation at the target point $(\phi+\Delta\phi,\, \lambda+\Delta\lambda)$ is equal to $\beta-\alpha$. Hence, the transformation from the longitudinal and transverse components back to the model components at that point is given by $\mathbf{R}^{-1}(\beta-\alpha) = \mathbf{R}(\alpha-\beta)$. Equation (3.73) for the model wind components then reads

$$\begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix} = w_{xy}^{LL} \cdot \mathbf{R}(2\alpha) \cdot \begin{bmatrix} \Delta u_k \\ \Delta v_k \end{bmatrix} - \gamma_n \cdot \frac{\Delta r}{s} \cdot w_{xy}^{LL} \cdot \mathbf{R}(\alpha-\beta) \cdot \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \cdot \mathbf{R}(\alpha+\beta) \cdot \begin{bmatrix} \Delta u_k \\ \Delta v_k \end{bmatrix} \tag{3.76}$$

p	1000	850	700	500	400	300	250	200	150	100	50
$s_{(u,v)}$	70	80	90	100	100	110	115	120	125	125	125
γ_n	.42	.50	.50	.50	.50	.58	.62	.66	.70	.70	.70

Table 3.7: Correlation scale for wind $s_{(u,v)}$ in [km] and non-divergence factor γ_n at observation time as a function of pressure p (in [hPa]).

Using Eqs. (3.71), (3.74), and (3.75), evaluation of Eq. (3.76) is straightforward and yields

$$\begin{bmatrix}\Delta u\\ \Delta v\end{bmatrix} = \left(\begin{bmatrix} 2\cos^2\alpha - 1 & 2\sin\alpha\cos\alpha \\ -2\sin\alpha\cos\alpha & 2\cos^2\alpha - 1\end{bmatrix} - \frac{\gamma_n}{\Delta r \cdot s} \begin{bmatrix} (\Delta r^2 + y^2)\cos^2\alpha - \Delta r^2 & (y^2+\Delta r^2)\sin\alpha\cos\alpha - x\,y\,\cos\alpha \\ (y^2-\Delta r^2)\sin\alpha\cos\alpha - x\,y\,\cos\alpha & (\Delta r^2 - y^2)\cos^2\alpha\end{bmatrix}\right) \cdot e^{-\Delta r/s} \begin{bmatrix}\Delta u_k\\ \Delta v_k\end{bmatrix} \tag{3.77}$$

Note that even if the non-divergence correction γ_n is set to 1, the resulting correlations are not in exact geostrophic balance. This is because the correlation scale s for temperature varies with height, and the resulting error correlations for geopotential cannot be expressed by Eq. (3.62) with an appropriate choice of s . Moreover, the resulting analysis increment field is not even exactly non-divergent due to the squaring of weights in Eq. (3.3). However, this is not a problem since the purpose of the present scheme is to analyze mainly the meso-scales.

In the operational version, both the correlation scale s for wind and the non-divergence factor γ_n increase with height and with distance to the observation time. The values valid at observation time are given in Table 3.7. The vertical variation of the scale s is very similar to that for temperature, but the magnitude of s is about 20 % larger. Nevertheless, the correlation function for wind falls off faster due to the different shape of the longitudinal correlations w_{xy}^{LL} . The non-divergence factor which allows to specify the degree of divergence varies between 0.42 and 0.77, with typical values of about 0.5 resulting in 'semi-divergent' increments (Figure 3.7 c) in the troposphere. The linear temporal variation is 30 % for s and 10 % for γ_n .

3.7.5 Lateral Spreading of Surface Pressure

Basic Correlation Function

For the lateral spreading of station pressure or surface pressure data, the basic form of the correlation function deployed is given by Eq. (3.62) like for the upper-air scalar quantities. The correlation scale s at the observation time is set to 70 km . The scale is linearly enhanced with increasing difference between observation time and model time by a maximum amount of 43 % . The function is cut off at a radius of $3.5{\cdot}s$. There is also an option for reducing s as a function of (increasing) density of surface pressure observations.

If the hydrostatic temperature correction for surface pressure nudging (Section 3.8.1) is applied, the pressure increments are projected on temperature increments between the ground

and 400 hPa . The mean level of projection is at about 750 to 800 hPa , where the resulting geopotential to surface geopotential resp. pressure to surface pressure correlations take values of about 0.5 (cf. Figure 3.10). At those levels, the horizontal correlation scale at the observation time is also about 70 km for upper-air temperature data. This means, that the information from pressure data and from temperature data are projected on approximately the same horizontal scale.

The choice of such a rather small correlation scale for pressure data is also promoted by the high data density in most of the operational COSMO domains. It allows to correct for small scale errors, and in particular affects the analysis of mesoscale cyclones favourably. The distinct enhancement of the correlation scale away from the observation time helps to project the information from isolated observations to larger scales. This applies all the more that the large value of the nudging coefficient for pressure data (cf. Section 3.7.6) tends to shift the model adjustment forward in time to prior to the observation time of isolated data (whereas in data-dense areas, the effective forcing is reduced and the main adjustment remains at close to the observation time).

Orography-Dependent Correction

Since surface pressure is a vertically integrated quantity, it need not (and cannot) be spread in the vertical in a hydrostatic atmosphere. However, an observation increment of surface pressure (resp. of pressure at the lowest model level) relates to the mass of the atmospheric column which in general differs in the horizontal, particularly over complex terrain. A surface pressure change of 1 hPa does not produce the same geopotential change at 1000 hPa in a valley as at 700 hPa on a high mountain. Using the basic lateral correlation function only, the resulting pressure analysis increment fields exhibit orographic footprints on horizontal or isobaric surfaces. In order to obtain smooth increment fields, an orography-dependent correction in the form of an additional weight factor w^{oro} has to be applied as part of the lateral spreading such that the same relationship is used as for the vertical extrapolation of pressure increments to the upper (model) levels within a column. This relationship is given by the hydrostatic equation and, depending on its application, by the hydrostatic temperature correction (Section 3.8.1) .

Without temperature correction, the orography-dependent correction should be consistent with the hydrostatic height correction that is used to convey pressure observation increments from the observation level down to the lowest model level k_s in Section 3.5.1 . Thus, it is given by the quotient of the pressure at the target point and that at the observation point,

$$\left.\frac{\Delta p'_{k_s}}{\Delta p'^{obs}_{k_s}}\right|_{w_{xy}=1} \doteq w^{oro} = \frac{p_{k_s}}{p^{obs}_{k_s}} \tag{3.78}$$

(It follows that $\Delta\Phi_{k_s} \cong \Delta\Phi^{obs}_{k_s} \cdot T_{k_s}/T^{obs}_{k_s}$, i.e. the orography-dependent variation of geopotential increments $\Delta\Phi_{k_s}$ is approximately proportional to the temperature quotient.)

The operational COSMO setups, however, make use of the temperature correction. The requirement of smooth pressure analysis increment fields on isobaric surfaces implies that along the sloping surface, the reduction of the pressure increments with height is the same as within a vertical column. Consequently, the orography-dependent correction is defined by the pressure to surface pressure correlation function (Eq. (3.87)) as specified in the temperature correction scheme. Assuming the basic lateral weight w_{xy} being equal to one, this means

that the increment Δp_{k_s} at the target pressure level p_{k_s} is given by

$$\Delta p_{k_s} \doteq w^{oro} \cdot \Delta p_{k_s}^{obs} = C_{\eta_c}^{p,p} \cdot \Delta p_{k_s}^{obs} \tag{3.79}$$

where

$$C_{\eta_c}^{p,p} = \eta_c^{\,2}\, e^{(1-\eta_c{}^3)/8} \tag{3.80}$$

$$\eta_c = \max\left(\frac{p_{k_s}-p_{top}}{p_{k_s}^{obs}-p_{top}}\ ,\ 0\right) \tag{3.81}$$

p_{top} denotes the pressure at the top of the temperature correction. Note, however, that these equations are restricted to target grid points with $p_{k_s} \le p_{k_s}^{obs}$, i.e. $\eta_c \le 1$. For any other grid point, the requirement of smoothness implies that the orography-corrected increment at the target grid point has to be chosen such that when it is reduced up to the observation level $p_{k_s}^{obs}$ it takes the same value as the original observation increment. This means that $\Delta p_{k_s}^{obs} = C_{\hat{\eta}_c}^{p,p} \cdot \Delta p_{k_s}$, or

$$\Delta p_{k_s} \doteq w^{oro} \cdot \Delta p_{k_s}^{obs} = \frac{1}{C_{\hat{\eta}_c}^{p,p}} \cdot \Delta p_{k_s}^{obs} \quad , \quad \hat{\eta}_c = \frac{p_{k_s}^{obs}-p_{top}}{p_{k_s}-p_{top}} = \frac{1}{\eta_c} \tag{3.82}$$

For these grid points $(\eta_c > 1)$, w^{oro} can be written as

$$w^{oro} = \frac{1}{\hat{\eta}_c^{\,2}}\, e^{-(1-\hat{\eta}_c{}^3)/8} = \eta_c^{\,2}\, e^{(1-\eta_c{}^3)/(8\,\eta_c{}^3)} \tag{3.83}$$

Combining Eqs. (3.80) and (3.83) leads to the final form for the orography-dependent correction weight,

$$w^{oro} = \eta_c^{\,2} \cdot e^{\frac{(1-\eta_c{}^3)/8}{\max(1,\eta_c{}^3)}} \tag{3.84}$$

This is valid for all target grid points, and η_c is defined by Eq. (3.81).

For the alternative temperature correction function which is expressed as a geopotential to surface geopotential correlation and which is implemented as an option (see Section 3.8.1), the orography-dependent correction weight can be derived in a similar way:

$$w^{oro} = \frac{p_{k_s}}{p_{k_s}^{obs}} \cdot \frac{T_{k_s}^{obs}}{T_{k_s}} \cdot \eta_c \cdot \max(1,\eta_c) \cdot \left(\frac{1+\eta_c}{2}\right)^{\frac{1-\eta_c}{|1-\eta_c|}} \tag{3.85}$$

For $\eta_c = 1$, this reduces to $w^{oro} = T_{k_s}^{obs}/T_{k_s}$.

3.7.6 Temporal Spreading and Specification of the Nudging Coefficients

Basic Considerations

In some respects, the basic considerations for the specification of the temporal spreading differ fundamentally from those for the spatial spreading. While spatial interpolation makes it possible to compare observed and model values valid for the same location to derive the

observation increments, the observed values have to be compared to the model values from the current model time which in general disagrees with the observation time. In particular, the model values at the observation time are not known before the assimilating model run itself reaches the observation time. This compromises the argument to use error correlations to define the temporal weight function and promotes the use of a shorter function for the time dimension. Furthermore, it is inherent to the concept of relaxation that the effect of the weighting is integrated in time. Hence, the relaxation concept does not agree with the correlation concept with respect to the time dimension.

Thus, other considerations should apply for the specification of the temporal weight function. In order to reduce the discrepancy between the observed and the model values to a reasonable degree by means of nudging, it is necessary that the time integral of total temporal weight function is of the same order of magnitude and preferably somewhat greater than the e-folding decay time of the relaxation. Since this decay time is the inverse of the nudging coefficient (for a single observation with weight $w_k = 1$), the time scale of the temporal weight function (i.e. the time integral over the function) should be approximately inversely proportional to the nudging coefficient.

Previous arguments related to the mismatch of observation time and model time suggest that a time scale as small as possible would be advantageous. This would also keep temporal smoothing effects small which may occur e.g. in the presence of fast-moving mesoscale cyclones or fronts. However, a longer time window for the relaxation would help to filter high-frequency inertia-gravity waves that could be excited by the nudging. Also, a very short window would imply a high value for the nudging coefficient which could destroy the dynamic balance during the assimilation.

To conclude, theoretical considerations can provide an rough estimate for the optimal time scale of the temporal weight function and for appropriate values of the nudging coefficients. A compromise between small and large values should yield the best results. The values that are applied operationally have been (or will soon be) adjusted to some extent by means of tuning experiments.

Nudging Coefficients

For wind, temperature, and humidity, the value of the nudging coefficient is set to $6{\cdot}10^{-4}\,s^{-1}$. It is applied to radiosonde, aircraft, and surface-level synoptic data. That value corresponds to an e-folding decay time of nearly half an hour for a single observation with a (spatial, temporal, and quality) weight w_k equal to 1. Note, however, that the effective weight is reduced and the decay time is enlarged in the case of multiple observations if Eq. (3.3) is applied. If for instance four observations with the same value for the observation increment are added in the vicinity so that their individual (spatial) weight w_k is 0.25 at the location of the original observation, then the total weight (related to the 5 observations together) will be reduced from 1 to about 0.625. This means that the resulting effective e-folding decay time is almost doubled. This effect prevails particularly with the surface pressure nudging where a target grid point is usually influenced by many observations from the relatively dense European synoptic surface station network.

This promotes a larger value for the nudging coefficient for surface pressure. As a result of tuning experiments, it is set to $12 \cdot 10^{-4}\,s^{-1}$.

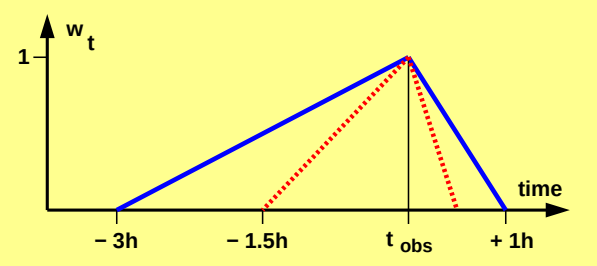

Figure 3.8: Temporal weight functions for single (i.e. non-frequent) observations. The blue solid line applies to radiosonde data, the red dotted line to other data.

Temporal Weighting

Given the relatively high data density over Europe and the nudging coefficients as above, the effective e-folding decay time for the relaxation will be typically between about 30 and 50 minutes. Hence, the appropriate time scale for the temporal weight functions should be about 1 hour or slightly larger. The shape of the function should be continuous to avoid shocks both when the forcing is turned on and increased and when it is decreased and turned off. An asymmetric shape allows more than 50 % of the model's adjustment to the data to be completed by the observation time.

As a result, the functions as shown in Figure 3.8 are designed for and applied to non-frequent observations. A more symmetrical function with appreciable forcing after the observation time would tend to hold back the model to fit 'old' data with undesirable effects. While the function valid for aircraft and surface-level data corresponds to a time scale of 1 hour, a time scale of 2 hours is currently used for radiosondes. This will make sure that the model draws closely to the radiosonde data, however, a scale of 2 hours is longer than what was suggested by theory. Also, there is evidence that the corresponding nudging time window of 4 hours causes smoothing effects in cases of fast-moving mesoscale cyclones. However, tuning experiments have shown degraded results, if the time scale was reduced (in particular for temperature and humidity).

For hourly or even more frequent data from the same stationary observing platform, the data are temporally linearly interpolated to the model time. The rationale for this is that the meteorological evolution is usually close to linear within one hour. The observations can then be regarded as a continuous stream of observational information, and an asymmetric shape of the function is not meaningful in such a case. If a platform issues frequent multi-level reports which cover different height ranges or have different quality, then the temporal weighting can be a linear combination of the weights related to a single observation resp. to temporal interpolation. In fact, this applies only to the report for which the total weight at the target grid point excluding the temporal weight is larger than for the other report. The combination depends on the fraction of these two total weights.

Wind profiler reports are usually available at intervals of 30 minutes or less. In these cases, linear temporal interpolation is applied. If the interval is larger than 30 minutes, a temporal weight function with a shape similar to the one shown in Figure 3.8 is applied, but it extends only 30 minutes into the past and 12 minutes into the future. The resulting time scale of the weight function is 21 minutes, then.

3.8 Explicit Balancing of the Analysis Increment Fields

3.8.1 Hydrostatic Temperature Correction

Basic Idea

The application of the temperature correction addressed here is motivated by the linear geostrophic adjustment theory (Okland (1970)). It assumes that inertia-gravity modes are dispersed or dissipated leaving only the non-divergent modes as solution. The smaller the horizontal scale and the larger the vertical scale of an initial perturbation which is not in geostrophic equilibrium, the more the mass field tends to adjust to the non-divergent component of the wind field. Vice versa, the smaller the vertical scale, the more the wind tends to adjust to the mass field geostrophically.

With respect to the nudging of surface pressure, this has the following implications when assuming that the increments are in hydrostatic equilibrium (by the application of the hydrostatic upper-air pressure correction as presented Section 3.8.4). Without any further correction, geopotential changes at the lowest model level due to the surface pressure nudging would impose geopotential increments throughout the atmosphere up to the top of the model (which is not physically meaningful anyway). Thus, the mass field perturbation would be very extended in the vertical and therefore tend to adjust to the unperturbed wind field. As a result, the mass field disturbance would propagate away in the form of inertia-gravity waves similarly to surface waves on a pond generated by a stone thrown into the water, and the observational information would get lost after the nudging period (cf. Figure 3.9 a).

In the real atmosphere, changes in surface pressure are always related to changes in density resp. (virtual) temperature somewhere above the ground. Hence, they are statistically correlated to temperature changes at a certain height range. For the planetary and synoptic scales, this range lies mainly in the lower stratosphere, whereas for the mesoscale which is more relevant in the context of data assimilation for COSMO, surface pressure is correlated mainly to the temperature in the lower troposphere. Therefore, temperature is corrected so as to satisfy the following conditions:

- The temperature correction approximates roughly the statistical surface pressure to temperature correlation in the mesoscale.
  (Without correction, temperature would be held constant at a given height, and a pressure increase at the surface would result in a slight cooling of the atmosphere in a way which is not consistent with the correlation.)

- The variation of the temperature correction is relatively small within the lowest about 1500 m above the ground so as to limit the resulting changes of the stability within the planetary boundary layer.

- The direct geopotential change due to the surface pressure nudging is zero above a specified pressure level p_{top} . In the current operational implementation, this level is at 400 hPa .

This implies that the vertical extent of the mass field perturbation imposed by the surface pressure nudging is reduced significantly. As a result, the wind field is far more compliant

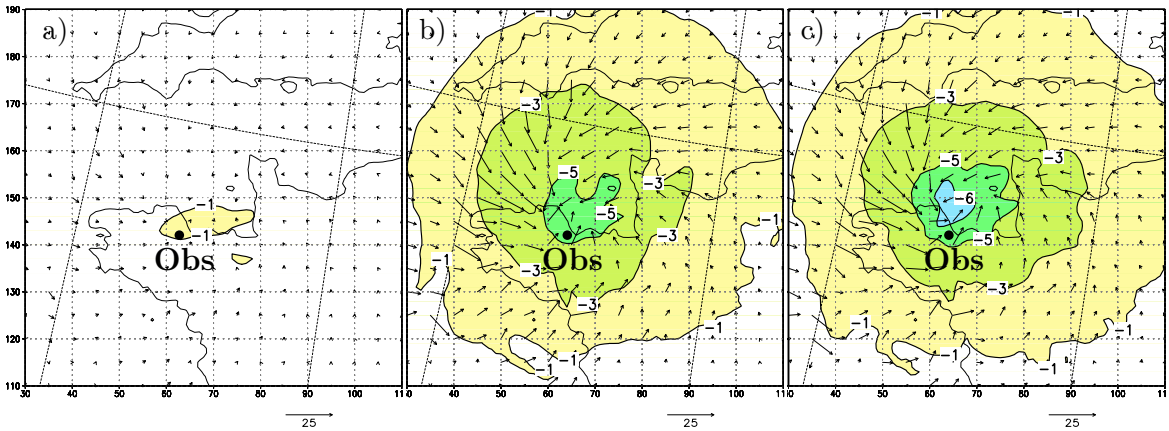

Figure 3.9: Difference fields of mean sea level pressure (isolines and shading: -1, -3, -5, -6 hPa) and of horizontal wind about 500 m above the ground (vectors) between a COSMO assimilation run and a free forecast after 3 hours of integration, valid for 24 October 1999, 3 UTC . For the assimilation run, only 2 surface pressure observations from 1 station (indicated by the blue dot) at 1 UTC and 2 UTC are used which deviate from the free forecast by about -6.1 hPa . a): Assimilation without temperature correction and geostrophic wind correction; b): with temperature correction but without geostrophic wind correction; c): with temperature correction as well as with geostrophic wind correction.

to adjust to the mass field perturbation, and the assimilation of surface pressure data is strongly improved (Figure 3.9 b).

Specification of the Correction Function

Following the ideas of Bell et al. (1996), the geopotential change at upper levels $\Delta\Phi$ is specified in terms of the geopotential change at the surface $\Delta\Phi_s$ in order to define the temperature correction. They used $\Delta\Phi_k = \frac{1}{2} B_k (1 + B_k) \Delta\Phi_s$ at model level k for a hybrid pressure-based vertical model coordinate $p_k = A_k + B_k p_s$. In COSMO, a similar function is implemented as an option by replacing B_k by the quantity η defined as

$$\begin{aligned} \eta &= \frac{p - p_{top}}{p_{k_s} - p_{top}} \quad , \quad p \geq p_{top} \\ \eta &= 0 \quad , \quad p \leq p_{top} \end{aligned} \tag{3.86}$$

This function, however, is not a good approximation to the statistical correlation, and the second condition is not satisfied at all. Hence, in the current scheme, another function is applied operationally which relates the pressure change at upper levels Δp to the pressure change (i.e. to the pressure analysis increments) at the lowest model level Δp_{k_s} :

$$\Delta p \doteq C^{p,p_{k_s}} \cdot \Delta p_{k_s} \quad , \quad C^{p,p_{k_s}} = \eta^2 \, e^{(1-\eta^3)/8} \tag{3.87}$$

(The relation between geopotential changes at upper levels and the lowest level is given by $C^{\Phi,\Phi_{k_s}} = C^{p,p_{k_s}} \cdot (p_{k_s} T_v)/(p T_{v_{k_s}})$ using the hydrostatic formulation $\Delta p = -\Delta\Phi \cdot p/(R T_v)$.) The resulting temperature correction is shown in Figure 3.10 for an initial model pressure of 1000 hPa at the lowest model level and for p_{top} set to 400 hPa . The 3 conditions are satisfied reasonably well. For a pressure change of –1 hPa at the lowest level, the temperature is increased by an almost constant value of slightly more than 0.5 K within the lowest 170 hPa . The upper-air to (near-)surface relations of the geopotential changes $C^{\Phi,\Phi_{k_s}}$ and of the pressure changes $C^{p,p_{k_s}}$ are also shown in Figure 3.10, and they give an indication of the strong reduction of the vertical extent of the resulting mass field perturbation. At 700 hPa ,

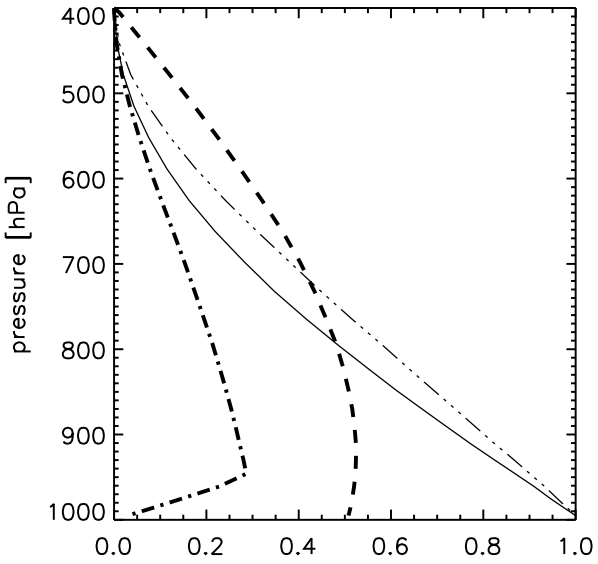

Figure 3.10: Vertical profiles related to the temperature and geostrophic wind corrections for near-surface pressure (i.e. pressure at the lowest model level) of 1000 hPa .
Thin solid line: Correlation of the pressure change aloft with the near-surface pressure change as given by Eq. (3.87).
Thin dot-dash-dotted line: Analogous correlation for the geopotential changes.
Thick dashed line: Temperature correction in [-K/hPa] (i.e. in [K] for a near-surface pressure decrease of 1 hPa).
Thick dash-dotted line: Geostrophic wind correction, expressed as a fraction of the full geostrophic wind with respect to the near-surface geopotential change.

for instance, the resulting geopotential change is only 40 % and the pressure change less than 30 % of that at the surface.

Discretized Implementation

It has been found to be difficult here to find appropriate initial conditions for the top-down integration of the hydrostatic equation that is exactly consistent with the model formulation. Also, as will become clear from below, it is not possible to obtain zero pressure increments above level p_{top} if relative humidity is to be preserved. Therefore, an quasi-iterative procedure is deployed, and a more simple form of the hydrostatic equation can be used to determine a first estimate of the temperature increments:

$$\frac{p_{k+1/2} - p_{k-1/2}}{z_{k+1/2} - z_{k-1/2}} = -\frac{g}{R} \cdot \frac{p_k}{T_{v_k}} \tag{3.88}$$

T_{v_k} , p_k , and $p_{k\pm1/2}$ denote the variable quantities virtual temperature and (hydrostatic) pressure at main model levels k resp. pressure at half levels, whereas $z_{k\pm1/2}$ are the invariable height values at the model's half levels which are fixed in physical space. (The hydrostatic pressure is derived by hydrostatic downward integration from the full model pressure at the top model level.) After the addition of the increments due to the surface pressure nudging including the temperature correction, the hydrostatic equation (3.88) reads

$$\frac{p_{k+1/2} + \Delta p_{k+1/2} - p_{k-1/2} - \Delta p_{k-1/2}}{z_{k+1/2} - z_{k-1/2}} = -\frac{g}{R} \cdot \frac{p_k + \Delta p_k}{T_{v_k} + \Delta T_{v_k}} \tag{3.89}$$

Subtracting Eq. (3.88) from Eq. (3.89) and solving for ΔT_{v_k} (using again Eq. (3.88)) yields

$$\Delta T_{v_k} = \frac{\left(z_{k-1/2} - z_{k+1/2}\right) \frac{g}{R} \cdot \Delta p_k - T_{v_k} \cdot \left(\Delta p_{k+1/2} - \Delta p_{k-1/2}\right)}{p_{k+1/2} - p_{k-1/2} + \Delta p_{k+1/2} - \Delta p_{k-1/2}} \tag{3.90}$$

Given the virtual temperature increments, there is no exact solution to determine temperature increments if relative humidity is to be preserved. Therefore, a quasi-iterative procedure is devised to split these increments into temperature and humidity increments:

1. Temperature increments are computed from the virtual temperature increments assuming constant specific humidity. Thus,

$$\Delta T_k^{(1)} = \frac{T_k}{T_{v_k}} \cdot \Delta T_{v_k} \tag{3.91}$$

   where $\Delta T_k^{(1)}$ is the temperature correction after the first iteration.

2. Specific humidity increments are optionally computed from the temperature increments assuming constant relative humidity. Note that this implies that the resulting virtual temperature increments deviate from those given in Eq. (3.90). Therefore, there are non-zero geopotential changes above p_{top} no matter which formulation of the hydrostatic equation is used.
   (This step is omitted if specific humidity instead of relative humidity is to be preserved, as is the case in the current operational set-up.)

3. A residual pressure increment at the top level of the temperature correction is computed by bottom-up integration of the hydrostatic equation using the pressure increment at the lowest model level and the upper-air temperature and specific humidity increments. This computation should be as exact as possible (because the residual increment can be regarded as a sort of observation increment). Therefore, the formulation of the hydrostatic equation used here is exactly consistent with the model formulation. It is presented in Section 3.8.4 on behalf of the hydrostatic upper-air pressure correction.

4. The upper-air residual pressure increment is converted into a residual pressure increment at the lowest model level. The exact way to do this would by top-down integration of the hydrostatic equation using pressure increments only (i.e. without temperature increments). However, as the demands for accuracy are more limited for this step (which can be regarded as a sort of spatial spreading of an observation increment), it is more convenient to apply a simple factor derived by approximating this integration (cf. Section 3.7.5: spreading of surface pressure observation increments):

$$\Delta p_{k_s}^{resid} = \frac{p_{k_s}}{p_{top}} \cdot \Delta p_{top}^{resid} \tag{3.92}$$

5. The occurrence of a non-zero residual pressure increment means that the temperature increments as above were able to equilibrate only a pressure increment of $\Delta p_{k_s} - \Delta p_{k_s}^{resid}$ instead of Δp_{k_s}. Thus, the final temperature increment can be determined from the first iteration increment by a simple scaling:

$$\Delta T_k = \frac{\Delta p_{k_s}}{\Delta p_{k_s} - \Delta p_{k_s}^{resid}} \cdot \Delta T_k^{(1)} \tag{3.93}$$

   This is equivalent to restarting the whole computation process several times with the initial Δp_{k_s} being replaced by $\Delta p_{k_s}^{resid}$ from the previous iteration.

3.8.2 Geostrophic Wind Correction

Basic Idea

The inclusion of the model dynamics in the assimilation process yields some implicit balancing between the wind and mass field perturbations that are induced by the nudging. The linear geostrophic adjustment theory (Okland (1970)) provides an estimation of the adjustment process related to the balancing. Since this process does not result in a very good assimilation of all types of observational information in general, it is advantageous to add a somewhat limited explicit balancing by use of a diagnostic relationship.

An incremental wind field which is in geostrophic balance to mass field increments can be determined easily. A fraction of this incremental geostrophic wind can then be added to the model wind field as a geostrophic wind correction. In this sense, increments from the nudging of both pressure and temperature observations can be balanced. In the current version, however, increments from temperature data are not taken into account for this correction. This is because the wind balancing such increments tends to have small-scale and rather irregular structures, and it is very sensitive to temperature observation errors with a non-zero bias. Also, spurious small-scale structures could result in areas where sloping terrain restricts the horizontal spreading of low-level temperature increments.

Therefore, the geostrophic wind correction applied here balances only the mass field increments that result from to the nudging of the surface pressure observations including the temperature correction. Thus, the top of the wind correction is at the same level p_{top} as the top of the temperature correction, i.e. at 400 hPa . It tends to have the effect to further improve the assimilation of the surface pressure data (Figure 3.9 c).

Derivation of the Full Geostrophic Increments

The continuous formulation of the geostrophic equation in COSMO reads

$$f\, v_g = \frac{1}{\rho\, a\, \cos\phi}\left(\frac{\partial p'}{\partial \lambda} - \frac{1}{\sqrt{\gamma}}\frac{\partial p_0}{\partial \lambda}\frac{\partial p'}{\partial \zeta}\right) \tag{3.94}$$

$$-f\, u_g = \frac{1}{\rho\, a}\left(\frac{\partial p'}{\partial \phi} - \frac{1}{\sqrt{\gamma}}\frac{\partial p_0}{\partial \phi}\frac{\partial p'}{\partial \zeta}\right)$$

where ζ denotes the computational vertical coordinate ($\zeta_k = k$, where k is the number of the k^{th} main model level), and $\sqrt{\gamma}$ the variation of reference pressure with ζ (i.e. $\sqrt{\gamma} = \partial p_0 / \partial \zeta$). Subsequently, the geostrophic wind increments in finite differences are derived for the v_g-component. The derivation for the u_g-component is analogous.

Equation (3.94) implies that variations in the geostrophic wind v_g can be due to variations in pressure p' or density ρ. If variables with subscript '$_F$' denote values prior to the nudging as in Section 3.2 and variables without subscript values after the nudging, the geostrophic wind increment Δv_g is given by

$$\begin{aligned} f\, \Delta v_g = {} & \frac{1}{\rho\, a\, \cos\phi}\left(\frac{\partial}{\partial \lambda}\left(p' - p'_F\right) - \frac{1}{\sqrt{\gamma}}\frac{\partial p_0}{\partial \lambda}\frac{\partial}{\partial \zeta}\left(p' - p'_F\right)\right) \\ & - \frac{1}{\rho\, a\, \cos\phi}\left(\frac{\partial p'}{\partial \lambda} - \frac{1}{\sqrt{\gamma}}\frac{\partial p_0}{\partial \lambda}\frac{\partial p'}{\partial \zeta}\right)\frac{1}{\rho}\left(\rho - \rho_F\right) \end{aligned} \tag{3.95}$$

A scale analysis reveals that close to the ground, the second term with the density variation is typically about two orders of magnitude smaller than the first term with the pressure variation. However, near the top level of the correction, the difference in size between the two terms is much reduced.

In finite differences, the geostrophic equation for the v_g-component is evaluated at the u-grid points of the Arakawa-C grid, since it results directly from the prognostic equation of the u-wind component by neglection of some of the terms. Considering the incremental iterative approach of the nudging and the limited validity of the geostrophic approximation for the scales of interest, the geostrophic equation does not need to be solved exactly for the present purpose. This allows to evaluate Eq. (3.94) at v-grid points instead of u-grid points. I.e., instead of applying the averaging operator over λ and ϕ to the left side, it is applied to the right side of the equation. The finite difference formulation of the geostrophic equation then reads

$$\frac{\overline{\sqrt{\gamma}}^{\phi} \cdot v_g \cos\phi}{\overline{\cos\phi}^{\lambda,\phi}} \cdot \overline{\left(\frac{f}{\overline{\sqrt{\gamma}}^{\lambda,\phi}}\right)}^{\lambda} = \overline{\frac{1}{\overline{\rho}^{\lambda} a \cos\phi}\left(\partial_\lambda p' - \frac{1}{\overline{\sqrt{\gamma}}^{\zeta,\lambda}} \partial_\lambda(p_0) \cdot \partial_{2\zeta} \overline{p'}^{\lambda}\right)}^{\lambda,\phi} \tag{3.96}$$

It is noted that in COSMO p', ρ, p_0, and $\sqrt{\gamma}$ are given at the centre of the grid boxes, whereas the Coriolis parameter f is given at the corners. For the increments, it follows that

$$\begin{aligned}\Delta v_g = \frac{I}{\overline{\sqrt{\gamma}}^{\phi}} \left[\overline{\left(\frac{f}{\overline{\sqrt{\gamma}}^{\lambda,\phi}}\right)}^{\lambda}\right]^{-1} \cdot \Bigg\{ & \overline{\frac{1}{\overline{\rho}^{\lambda} a_\phi}\left(\partial_\lambda\left(p' - p'_F\right) - \frac{\partial_\lambda p_0}{\overline{\sqrt{\gamma}}^{\zeta,\lambda}} \cdot \partial_{2\zeta} \overline{p' - p'_F}^{\lambda}\right)}^{\lambda,\phi} \\ & - \overline{\frac{1}{(\overline{\rho}^{\lambda})^2 a_\phi}\left(\partial_\lambda p' - \frac{\partial_\lambda p_0}{\overline{\sqrt{\gamma}}^{\zeta,\lambda}} \partial_{2\zeta} \overline{p'}^{\lambda}\right) \cdot \overline{\rho - \rho_F}^{\lambda}}^{\lambda,\phi} \Bigg\}\end{aligned} \tag{3.97}$$

where $I = \overline{\cos\phi}^{\lambda,\phi} \cdot (\cos\phi)^{-1}$ and $a_\phi = a\cos\phi$. Analogously

$$\begin{aligned}\Delta u_g = \frac{-1}{\overline{\sqrt{\gamma}}^{\lambda}} \left[\overline{\left(\frac{f}{\overline{\sqrt{\gamma}}^{\lambda,\phi}}\right)}^{\phi}\right]^{-1} \cdot \Bigg\{ & \overline{\frac{1}{\overline{\rho}^{\phi} a}\left(\partial_\phi\left(p' - p'_F\right) - \frac{\partial_\phi p_0}{\overline{\sqrt{\gamma}}^{\zeta,\phi}} \cdot \partial_{2\zeta} \overline{p' - p'_F}^{\phi}\right)}^{\lambda,\phi} \\ & - \overline{\frac{1}{(\overline{\rho}^{\phi})^2 a}\left(\partial_\phi p' - \frac{\partial_\phi p_0}{\overline{\sqrt{\gamma}}^{\zeta,\phi}} \partial_{2\zeta} \overline{p'}^{\phi}\right) \cdot \overline{\rho - \rho_F}^{\phi}}^{\lambda,\phi} \Bigg\}\end{aligned} \tag{3.98}$$

$(p' - p'_F)$ and $(\rho - \rho_F)$ are the pressure resp. density analysis increments from the nudging of the surface pressure data including the combined effect of the temperature correction and the hydrostatic upper-air pressure correction (see Section 3.8.4). Note, however, that the pressure and density increments used here are determined without accounting for latent heat to sensible heat transfer in order to prevent unrealistic small-scale wind increments in saturated regions. Yet, they optionally include a change in specific humidity to keep the relative humidity constant.

Reduction of the Increments

For different reasons, only a fraction of the incremental geostrophic wind as computed above is added to the model wind field (cf. Figure 3.10).

Firstly, in order to account for the limited validity of geostrophy in the mesoscale, all geostrophic increments are reduced by a general factor. Operationally, it decreases linearly in pressure from 0.5 at 400 hPa to 0.3 at 1000 hPa. Such a general reduction is also motivated by the fact that the assimilation process of the observational (pressure) information is not ideal. Dynamic or other effects often tend to moderately counteract the nudging. This reduces the rate at which the pressure increments are diminished, and the total geostrophic wind increments integrated over few hours could become larger than the geostrophic wind related to the initial pressure observation increments.

Secondly, in order to account for the limited validity of geostrophy in the planetary boundary layer, another weight is applied which increases linearly in pressure from zero at the ground to one at about 50 hPa above the ground.

Thirdly, in order to account for the limited validity of the geostrophic approximation at low latitudes, another weight is applied for latitudes $\phi < 45$ degrees which is set equal to $\sin(2\phi)$. This function approximates the latitude dependency of the validity of the geostrophic approximation that is obtained by the NMC method for estimation of the climatological errors (in the global model GME). This third weighting prevents the geostrophic wind increment from growing to infinity as one approaches the equator.

Next, a general upper safety limit is imposed to the size of the increments in such a way that the maximum increments can add up to 20 m/s within an hour.

Finally, the geostrophic wind correction is set to zero in the lateral boundary relaxation zone (at the outermost 10 grid rows of the model domain), because the horizontal gradients of the pressure analysis increments are strongly influenced by the boundary relaxation. Towards the inner part of the domain, the weight is gradually increased from zero to one within 4 grid points.

3.8.3 Geostrophic Near-Surface Pressure Increments

Basic Idea

The concept of adding (limited) explicit balancing between the wind and mass field analysis increments by use of a diagnostic relationship can also be applied in the opposite direction compared to the previous section. Here, a wind analysis increment field at the lowest model level as obtained from the direct nudging of wind observations is used as input. The aim is to compute a pressure increment field at the lowest model level which is in geostrophic balance with the wind increment field.

This type of balancing is particularly appealing for the assimilation of 10-m wind observations. The correlation of the 10-m wind (forecast) errors with upper-air wind errors is often rather limited, so that the scale of the vertical weight function should not be very large (see Section 3.7.1). As a result, the perturbation induced by a direct nudging of 10-m wind data is rather shallow, and for shallow imbalances, the wind field tends to adjust to the mass field rather than vice versa (according to the geostrophic adjustment theory). Therefore, the information on the wind field will not be retained well without explicit perturbation of the mass field. The introduction of balanced mass field perturbations significantly enhances the assimilation of 10-m wind data by nudging.

Two options have been implemented for the observation input used for the wind analysis increment field:

1. scatterometer 10-m wind data only
2. scatterometer and other 10-m wind observations, e.g. from land surface stations, ships, and buoys.

The wind increments for this balancing step are computed using only the isotropic horizontal weight function (see Section 3.7.4) without any non-isotropic corrections (see Section 3.7.3. This ensures that the wind increment field is smooth even in the presence of steep orography. A noisy wind field with large small-scale gradients would lead to a noisy 'balanced' pressure increment field and eventually result in gravity wave oscillations. Note that geostrophy is not a good approximation for very small-scale flow features.

Derivation of a Poisson Equation for Geopotential Increments

There are different methods to compute pressure increments which are in geostrophic balance with a prescribed wind (increment) field. Following Holton (1992), the relation of nondivergent winds to the pressure field is given by the well-known general balance equation:

$$\nabla^2 \left(\Phi + \frac{\vec{v} \cdot \vec{v}}{2} \right) = -\nabla \left[\vec{k} \times \vec{v}(\zeta + f) \right] \tag{3.99}$$

where ζ denotes the vertical component of the vorticity, Φ the geopotential, $\vec{k}$ the unit vector in vertical direction, $\vec{v}$ the horizontal wind vector, and f the coriolis parameter. 3.99 can be written as:

$$\nabla^2 \Phi = f \cdot \zeta - \beta \cdot u + 2 \cdot J(u, v) \tag{3.100}$$

with the Jacobian

$$J(u, v) = \frac{\partial u}{\partial x} \cdot \frac{\partial v}{\partial y} - \frac{\partial u}{\partial y} \cdot \frac{\partial v}{\partial x} \quad , \quad \beta = \frac{\partial f}{\partial y} \tag{3.101}$$

Equation 3.100 is valid for a nondivergent windfield. But even over the ocean surface, the flow will not be completely free of divergence. One way to obtain a solution would be to decompose the fow field into a nondivergent and a divergent part and discard the divergent part. Since both decomposing the flow field and solving equation 3.100 is done by iterations, much computation time would be consumed in order to yield a converged solution.

Thus in the current implementation, the simplest approximation of the balance equation is deployed by saving only the two largest terms. Equation 3.100 reduces then to the geostrophic relationship, which attains the form of a Poisson equation given that geopotential is the unknown quantity:

$$\nabla^2 \Phi = f \cdot \zeta \tag{3.102}$$

with ζ the relative vorticity:

$$\zeta = rot\, \vec{v}_H = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \tag{3.103}$$

Note that the wind increment field and vorticity are equal to zero at the lateral boundary of the total model domain. Hence the balancing geopotential increment field is zero there too.

The approximation 3.102 states, that the wind field is in geostrophic balance rather than in gradient wind balance. The neglection of terms in equation 3.100 results in an unbalanced

inertial force. This can cause errors in computing the appropriate pressure field and can result in some gravity wave oscillation as the flow attempts to adjust back to the gradient wind balance. The errors are prevalent in regions where the flow is significantly curved. They result in an underestimation of the pressure gradient in regions of strong cyclonic curvature and an overestimation of the pressure gradient force in regions of strong anticyclonic curvature. For most of the synoptic situations however, the errors are small (at synoptic and meso-α scales). The right hand of equation 3.102 is known and can be computed at each grid point from the incremental wind field using expression 3.103.

Solving the Poisson Equation

Assuming $\Delta x = \Delta y = \Delta s$, equation 3.102 becomes:

$$\Phi_{i,j} = \frac{1}{4} \cdot \left[\Phi_{i+1,j} + \Phi_{i-1,j} + \Phi_{i,j+1} + \Phi_{i,j-1} - \Delta s^2 \cdot f_{i,j} \cdot \zeta_{i,j} \right] \tag{3.104}$$

The coriolis parameter $f_{i,j}$ is interpolated to the mass point of the grid. In fact in the current version, the coriolis parameter is included in Φ so that:

$$\Phi \rightarrow \frac{\Phi}{f} \tag{3.105}$$

$$\Phi_{i,j}^{n+1} = \frac{1}{4} \cdot \left[\left(\Phi_{i-1,j}^{n+1} + \Phi_{i,j-1}^{n+1} \right) + \left(\Phi_{i+1,j}^{n} + \Phi_{i,j+1}^{n} \right) - \Delta s^2 \cdot \zeta_{i,j} \right] \tag{3.106}$$

Thus the solution for the unknown Φ of equation 3.106 has to be multiplied with f in order to recover the real geopotential Φ. Equation 3.106 can be solved for Φ by a single step iteration procedure; i.e. in each iteration, $\Phi_{i,j}$ is updated successively at each grid point in order of increasing indices i, j. Thus, to update $\Phi_{i,j}$ in iteration $n+1$, the values of iteration $n+1$ for Φ can be used at grid points $(i-1,j)$, $(i,j-1)$, whereas for grid points $(i+1,j)$, $(i,j+1)$, the values of iteration n have to be used. In practice the speed of convergence of the iteration can be significantly enhanced by two actions:

1. The iterative corrections of $\Phi_{i,j}$ are multiplied with a relaxation factor $\omega \neq 1$ such that

$$\Delta\Phi_{i,j}^{n+1} = \Phi_{i,j}^{n+1} - \Phi_{i,j}^{n} \tag{3.107}$$

$$\Phi_{i,j}^{*\,n+1} = \Phi_{i,j}^{n} + \omega \cdot \Delta\Phi_{i,j}^{n+1} \tag{3.108}$$

$\omega > 1$ means overrelaxation, $\omega < 1$ underrelaxation. Combining equations 3.107 and 3.108 leads to:

$$\Phi_{i,j}^{*\,n+1} = (1-\omega) \cdot \Phi_{i,j}^{n} + \omega \cdot \Phi_{i,j}^{n+1} \tag{3.109}$$

Substituting $\Phi_{i,j}^{n+1}$ on the right hand of 3.109 by expression 3.106 yields:

$$\Phi_{i,j}^{*\,n+1} = (1-\omega) \cdot \Phi_{i,j}^{n} + \frac{\omega}{4} \cdot \left[\left(\Phi_{i,j-1}^{n+1} + \Phi_{i-1,j}^{n+1} \right) + \left(\Phi_{i+1,j}^{n} + \Phi_{i,j+1}^{n} \right) - \Delta s^2 \cdot \zeta_{i,j} \right] \tag{3.110}$$

In the next iteration, $\Phi^{*\,n+1}$ is set to Φ^{n}. Equation 3.110 is the successive overrelaxation (SOR) method applied to equation 3.106; $\omega > 1$ will accelerate the convergence of the iterations. The convergence of 3.110 benefits from the first two terms in the bracket, since these Φ-values are already computed during the same sweep through the domain.

2. Unfortunately it turns out that for large grids as used in typical operational configurations of the COSMO model, the speed of convergence is still unsatisfying. The reason is that the values communicate right next to their neighbours but information propagates only very slowly over long distances in terms of number of grid points. As a result, the SOR method which is suitable for small grids, is completely unsuitable for large fine mesh grids due to slow convergence in this case.

   Potentially, a way out of this could be to interpolate the problem from the large fine mesh to an auxilliary coarse grid and solve the problem there. Generally this is not a viable solution, because a coarse grid cannot capture the structure of a rough function on the fine grid. In the current application however, the input function given by the wind increment field is always rather smooth. This is because the scale of the isotropic horizontal weight function used to derive that wind increment field is larger than 75 km (and the effective scale of the weight function is not much reduced in areas with high data density. Moreover, scatterometer wind data are available at or thinned to a mean distance of about 50 km , and the data density of the in-situ observations is usually even smaller). Consequently, a coarse grid with a mesh width of about 75 km will capture most of the information of the wind increment field.

   Based on these considerations, the following approach is adopted. First, the wind increment field is interpolated to an auxilliary coarse grid, on which the Poisson equation is solved using equation 3.110. Then the coarse-grid solution is interpolated back onto the fine-mesh COSMO grid and used as a first guess for solving equation 3.110 on the fine grid.

   The coarse grid is defined by pre-specifying its number of grid points n_s^c in each direction by

$$n_s^c - 1 \quad = \quad nint((n_s - 1) \cdot d_s \cdot \beta) \tag{3.111}$$

   where n_s is the number of grid points and d_s the mesh width in degrees of the original fine mesh, and s stands for the longitudinal and latitudinal directions x resp. y . Lower and upper limits on n_s^c are imposed by $(n_s - 1)/20 \leq (n_s^c - 1) \leq (n_s - 1)/5$.

   The parameter β determines the resolution of the coarse grid. Currently, $\beta = 1.5$ is specified. This results in a mesh width of $\Delta s_c \approx a \cdot 2\pi/360 \,/\, 1.5 \approx 74\,\text{km}$, where a is the earth radius. This is a suitable value given the scale of the horizontal correlation function used to derive the wind increment field. The scale of this function may be adapted to the mean distance of the surface-level wind observations which may change in the future. In this case, β can easily be adapted to obtain a suitable mesh width Δs_c .

As the computation is executed on a rotated latitude-longitude grid, equation 3.110 is written in spherical coordinates. At the same time, the restriction of $\Delta x = \Delta y$ is dropped. The Laplacian of Φ then reads:

$$\nabla^2\Phi \quad = \quad \frac{1}{(a \cdot \cos\varphi)^2} \cdot \frac{\partial^2\Phi}{\partial\lambda^2} \; + \; \frac{1}{a^2 \cdot \cos\varphi} \cdot \frac{\partial}{\partial\varphi}\left(\cos\varphi \cdot \frac{\partial\Phi}{\partial\varphi}\right) \tag{3.112}$$

and the vorticity

$$\zeta \quad = \quad \frac{1}{a \cdot \cos\varphi} \cdot \left(\frac{\partial v}{\partial\lambda} \; - \; \frac{\partial}{\partial\varphi}\left(\cos\varphi \cdot u\right)\right) \tag{3.113}$$

The discretized form is as follows (the subscript 'c' for the coarse grid is dropped hereafter for convenience):

$$\left(\nabla^2\Phi\right)_{i,j} = e_1 \cdot (\Phi_{i+1,j} - 2\,\Phi_{i,j} + \Phi_{i-1,j}) + e_2 \cdot (\Phi_{i,j+1} - \Phi_{i,j}) - e_3 \cdot (\Phi_{i,j} - \Phi_{i,j-1}) \tag{3.114}$$

$$\zeta_{i,j} = \frac{v_{i+1,j} - v_{i-1,j}}{2 \cdot a \cdot \cos\varphi_j \cdot \Delta\lambda} - \frac{\cos\varphi_{j+1} \cdot u_{i,j+1} - \cos\varphi_{j-1} \cdot u_{i,j-1}}{\cos\varphi_j \cdot 2 \cdot a \cdot \Delta\varphi} \tag{3.115}$$

and after rearranging the terms and recalling equation 3.110

$$\Phi^{n+1}_{i,j} = (1-\omega) \cdot \Phi^n_{i,j} + \omega \cdot e_4 \cdot \Big[e_1 \cdot \left(\Phi^n_{i+1,j} + \Phi^{n+1}_{i-1,j}\right) + e_2 \cdot \Phi^n_{i,j+1} + e_3 \cdot \Phi^{n+1}_{i,j-1} - \zeta_{i,j} \Big] \tag{3.116}$$

with

$$e_1 = \frac{1}{(a \cdot \cos\varphi_j \cdot \Delta\lambda)^2} \quad , \quad e_2 = \frac{\cos\varphi_{j+1/2}}{a^2 \cdot \cos\varphi_j \cdot \Delta\varphi^2} \quad , \quad e_3 = \frac{\cos\varphi_{j-1/2}}{a^2 \cdot \cos\varphi_j \cdot \Delta\varphi^2}$$
$$e_4 = (2 \cdot e_1 + e_2 + e_3)^{-1} \tag{3.117}$$

The procedure to solve equation 3.116 is as follows:

1. Interpolate the components of the wind increments bi-linearly onto the coarse grid.
2. Solve equation 3.116 iteratively on the coarse grid (replacing $\Delta\lambda$, $\Delta\varphi$ by $\Delta\lambda_c$, $\Delta\varphi_c$) with $\Phi = 0$ as initial value for the first iteration (out of a fixed number of 1640 iterations). In order to avoid a second decomposition of the model domain for parallel processing, this is executed on a single processor. The calculation is not very expensive since the iterations are linear operations and the coarse grid has typically 1 - 2 orders of magnitude less grid points than the original one.
3. Interpolate (bi-linearly) the resulting Φ-field back onto the fine model grid.
4. Solve equation 3.116 iteratively on the fine grid using the interpolated Φ-field as initial value for the first iteration. This is done in parallel mode and for a fixed value of 50 iterations. Since the domain is decomposed here, $\Phi^{n+1}_{i-1,j}$ and $\Phi^{n+1}_{i,j-1}$ have to be replaced by $\Phi^n_{i-1,j}$ and $\Phi^n_{i,j-1}$ in order to ensure reproducibility independent from the domain decomposition. Thus, the convergence cannot benefit from the Φ-values of the same sweep. But this disadvantage is more than compensated by parallel processing.
5. Multiply the final result of Φ with the coriolis parameter f in order to recover the real geopotential increment field.

Further Processing of the Geopotential Increments

Since the non-isotropic part of the lateral weights, which reflect mainly the orography, has not been included in the intial wind analysis increments, the resulting geostrophic geopotential increments should be regarded as being representative for the mean sea level. (More precisely, they are representative for (a variable combination of) the height levels of the wind observations, but this detail aspect is neglected here.) They have to be converted to

pressure increments and then extrapolated to the model orography level. For the conversion into pressure, the required mean temperature is set equal to the average of the model temperature at the lowest model level and the temperature extrapolated to the mean sea level. The temperature extrapolation uses a constant climatological lapse rate of $0.0065\,\mathrm{K/m}$. The pressure increments valid at the mean sea level pressure are then extrapolated to the lowest model level by applying the same correction factor that is used in the lateral spreading of the surface pressure observation increments (see Section 3.7.5). Finally, the near-surface pressure increments (i.e. increments at the lowest model level) are scaled by a constant factor less or equal to one, which reflects the limited validity of the geostrophic approximation.

The final near-surface pressure increments derived in the current explicit balancing step are then further balanced with respect to the mass field. This is done in the same way as for the normal near-surface pressure analysis increments obtained by the nudging of surface pressure observations. I.e. at first, a temperature correction is computed according to Section 3.8.1. Then upper-air pressure increments balancing hydrostatically the near-surface pressure and upper-air temperature increments are derived as described in the following Section 3.8.4. However, the current near-surface pressure increments are of course excluded from the input to the geostrophic wind correction.

3.8.4 Hydrostatic Upper-Air Pressure Correction

Basic Idea

By nudging the types of conventional observations as used currently, analysis increments are computed for wind, temperature, and humidity at all model levels, and for pressure at the lowest model level. Without any modification to pressure at upper levels, the total value of the right side of the vertical momentum equation would be changed in general by adding these increments to the model fields. In other words, the nudging terms would be a direct source of vertical wind. Yet, the scales to be analyzed by the nudging of pressure, wind, temperature, and humidity data are typically larger than $100\,\mathrm{km}$ in the horizontal according to the lateral weight functions applied. On such scales, vertical wind velocities are known to be relatively small, and they should not be significantly enhanced by the nudging. Furthermore, from the set of observations used, and without vertical wind observations being available for assimilation, there is usually no indication whether the vertical wind should be increased or decreased.

Consequently, direct sources of vertical wind should be avoided, and the total analysis increments should be balanced hydrostatically. This is achieved by a hydrostatic upper-air pressure correction which is determined in the following way. From the pressure increments at the lowest model level and the temperature and humidity increments at all levels, upper-air pressure increments above the lowest level are computed by integrating the hydrostatic equation from the lowest model level upwards. This correction is the only feature directly related to fact that the model is non-hydrostatic. It is stressed that although the total analysis increments are balanced hydrostatically, this does not change the non-hydrostatic properties of the full model fields.

It should be mentioned that these considerations may not apply if other types of observations are used. For instance, observed significant small-scale precipitation exceeding the simulated precipitation may indicate that condensation as well as updraft winds in a convective cell are underestimated in the model. The so-called latent heat nudging aims to adjust the effects

of the condensation which would be required to produce the observed precipitation rates. In fact, it adjusts the latent heating and thus the model temperature (and optionally the humidity in addition). In the present example, the temperature would be increased. Without hydrostatic balancing, this would be a source of increased upward motion. In this case, such an effect is desired since it would tend to support the underestimated convection processes. Note that the horizontal scales considered here are far smaller than the ones discussed above.

Mathematical Formulation

The formulation of the hydrostatic equation that is deployed here must be exactly consistent with the model formulation. Hence, it must be derived from the vertical momentum equation by neglecting sub-grid scale processes and by setting the total derivative of vertical wind to zero. In finite differences, the resulting equation is defined on model half level and reads

$$\frac{1}{\sqrt{\gamma}}\,\delta_\zeta\, p' \;=\; \overline{\frac{T_0}{T^n p_0}\,p'}^{\,\zeta} \;-\; \overline{\frac{T-T_0}{T^n}}^{\,\zeta} \;-\; \overline{\left(\frac{R_v}{R_d}-1\right) q^v - q^l - q^f}^{\,\zeta} \tag{3.118}$$

where ζ denotes the computational vertical coordinate ($\zeta_k = k$, where k is the number of the k^{th} main model level), and $\sqrt{\gamma}$ the variation of reference pressure with ζ (i.e. $\sqrt{\gamma} = \partial p_0 \,/\, \partial \zeta$). $\overline{\quad}^{\,\zeta}$ symbolizes the vertical averaging operator. With the use of this operator, of the vertical differencing operator, and of the factor $\sqrt{\gamma}$ at half levels as defined for the discretized form of the model equations (see Documentation Part I on Dynamics and Numerics), Eq. (3.118) takes the form

$$\begin{aligned}\frac{p'_{k+1}-p'_k}{\frac{1}{2}\left(\sqrt{\gamma_{k+1}}+\sqrt{\gamma_k}\right)} \;=\; & \frac{\sqrt{\gamma_k}}{\sqrt{\gamma_{k+1}}+\sqrt{\gamma_k}}\left\{\frac{T_{0_{k+1}}}{T^n_{k+1}p_{0_{k+1}}}p'_{k+1} \;-\; \frac{T_{k+1}\;-\;T_{0_{k+1}}}{T^n_{k+1}} \;-\; q^{virt}_{k+1}\right\} \\ & +\; \frac{\sqrt{\gamma_{k+1}}}{\sqrt{\gamma_{k+1}}+\sqrt{\gamma_k}}\left\{\frac{T_{0_k}}{T^n_k p_{0_k}}p'_k \;-\; \frac{T_k\;-\;T_{0_k}}{T^n_k} \;-\; q^{virt}_k\right\}\end{aligned} \tag{3.119}$$

where $q^{virt} \;=\; (R_v/R_d - 1)\cdot q^v - q^l - q^f$.

For increments instead of full model fields, it follows that

$$\begin{aligned}2\left(\Delta p'_{k+1}-\Delta p'_k\right) \;=\; & \sqrt{\gamma_k}\,\frac{T_{0_{k+1}}}{T^n_{k+1}p_{0_{k+1}}}\cdot\Delta p'_{k+1} \;-\; \sqrt{\gamma_k}\,\frac{1}{T^n_{k+1}}\cdot\Delta T_{k+1} \;-\; \sqrt{\gamma_k}\cdot\Delta q^{virt}_{k+1} \\ & +\; \sqrt{\gamma_{k+1}}\,\frac{T_{0_k}}{T^n_k p_{0_k}}\cdot\Delta p'_k \;-\; \sqrt{\gamma_{k+1}}\,\frac{1}{T^n_k}\cdot\Delta T_k \;-\; \sqrt{\gamma_{k+1}}\cdot\Delta q^{virt}_k\end{aligned} \tag{3.120}$$

Solving for $\Delta p'_k$ finally yields

$$\begin{aligned}\Delta p'_k \;=\; & \left(2+\sqrt{\gamma_{k+1}}\,\frac{T_{0_k}}{T^n_k p_{0_k}}\right)^{-1}\cdot\;\left\{\left(2-\sqrt{\gamma_k}\,\frac{T_{0_{k+1}}}{T^n_{k+1}p_{0_{k+1}}}\right)\cdot\Delta p'_{k+1}\right. \\ & +\;\sqrt{\gamma_k}\,\frac{1}{T^n_{k+1}}\cdot\Delta T_{k+1} \;+\; \sqrt{\gamma_k}\cdot\Delta q^{virt}_{k+1} \\ & \left.+\;\sqrt{\gamma_{k+1}}\,\frac{1}{T^n_k}\cdot\Delta T_k \;+\; \sqrt{\gamma_{k+1}}\cdot\Delta q^{virt}_k\right\}\end{aligned} \tag{3.121}$$

where $\Delta q^{virt} \;=\; (R_v/R_d - 1)\cdot\Delta q^v - \Delta q^l - \Delta q^f$. Since the analysis increments ΔT and Δq^{virt} are known at all levels and $\Delta p'_{k+1}$ is known at the lowest model level, $\Delta p'_k$ can be computed from the second but lowest level upwards.

3.9 Determination of the Analysis Increments

This section describes the order of the various steps that are required to derive the final analysis increments, and presents the nudging equations for the different variables that are analyzed. Here, the analyzed variables are denoted to be those model variables, for which a nudging term (or explicit balancing term) is added to the original dynamic model equation. Currently, this set of variables consists of temperature, specific humidity, and the horizontal wind components at all model levels, and pressure (deviation from the base state) at the lowest model level. Pressure analysis increments further above are derived by means of the hydrostatic upper-air pressure correction. For the remaining atmospheric model variables, i.e. the vertical wind component, cloud water and ice contents, and turbulent kinetic energy, no analysis increments are computed. Thus, apart from the fact that cloud water and ice contents can be modified in the presence of condensation or evaporation within the nudging, these variables are not influenced by the nudging except indirectly due their interaction with the analyzed variables in the model dynamics and physics. This interaction causes some transfer of observational information to these remaining variables, and it is in this sense that the complete nudging process provides an analysis of all atmospheric model fields.

After the observation processing and the determination of all the observation increments including the quality control, pressure analysis increments at the lowest model level are computed first. The corresponding equation can be derived straightforwardly from Eq. (3.10) by replacing ψ by p_s :

$$\Delta_A\, p_s \;=\; \frac{2\Delta t\; G_{p_s}\; \mu_{p_s}}{1 + 2\Delta t\; G_{p_s}\; \mu_{p_s}} \cdot \frac{\sum_k (w_k^{\,2} \cdot [p_{s_k}^{obs} - p_{s_F}^{n+1}(\mathbf{x}_k)])}{\sum_{k'} w_{k'}^{\,2}} \tag{3.122}$$

where Δ_A denotes analysis increments, and $\mu_{p_s} = f_{lbc} \cdot \sum_k w_k^2 \,/\, \sum_{k'} w_{k'}$. f_{lbc} is an additional weight which is equal to 1 in the inner model domain but reduces the nudging weight near the lateral boundaries ($f_{lbc} = 1 - \mu_b \,/\, \mu_b^{max}$, where μ_b is the boundary relaxation parameter as described in Part I of the COSMO documentation, and μ_b^{max} is the maximum value of μ_b given at the outermost grid row of the domain). f_{lbc} is used in the nudging equations for all the analyzed variables in an analogous way. Before the pressure analysis increments given by Eq. (3.122) are added to the model fields, the temperature increments related to the hydrostatic temperature correction are computed (Section 3.8.1).

All the other increments are subsequently determined one model level after another from the lowest level upwards. After the spreading of the upper-air and surface-level observation increments and the computation of the sums of weights and weighted increments at the grid points, the first step at each model level is the nudging of temperature. To derive the nudging equation, the updating of temperature is considered a step-by-step process where the first step is the temperature change imposed by the temperature correction, $\Delta_{p_s} T$. The temperature T_c after this is given by

$$T_c^{n+1} \;=\; T_F^{n+1} \;+\; \Delta_{p_s} T \tag{3.123}$$

The second step is a normal nudging step given by

$$T^{n+1} \;=\; T_c^{n+1} \;+\; 2\Delta t\; G_T \cdot \sum_k W_k \cdot [T_k^{obs} - T^{n+1}] \tag{3.124}$$

in its implicit formulation. Combining these two steps by using Eq. (3.123) in Eq. (3.124)

and solving for T^{n+1} yields

$$T^{n+1} = T_F^{n+1} + \frac{2\Delta t\ G_T\ \mu_T}{1+2\Delta t\, G_T \mu_T} \cdot \frac{\sum_k (w_k^2 \cdot [T_k^{obs} - T_F^{n+1}])}{\sum_{k'} w_{k'}^2} + \frac{1}{1+2\Delta t\, G_T \mu_T} \cdot \Delta_{p_s} T \quad (3.125)$$

As an optional supplementary step related to temperature nudging, specific humidity can be adjusted so that relative humidity is not directly affected by temperature nudging. This adjustment of humidity may also be chosen to be confined to certain environs of convectively precipitating grid points only. However, it is not applied at all to the current operational version, firstly because preserving relative humidity may result in overestimation of convective precipitation, and secondly because preserving specific humidity tends to enhance – and usually improve – the strong vertical gradient of relative humidity in cases where a low-level inversion associated with low stratus is enhanced by nudging of temperature data.

The next step consists of the nudging of humidity observations. With respect to upper-air data, at least, an adjustment of the simulated relative humidity to the observed relative humidity is expected to offer a better representation of the initial cloudiness e.g. of low stratus than an adjustment to the observed specific humidity. This implies that the observation increments should be expressed as differences between observed and simulated relative humidity. That can be done even though the model variable is specific humidity and Eq. (3.10) indicates that the observation increments must usually be expressed in terms of the model variables for the nudging. The ordinary form of the nudging equation for specific humidity $q_v \doteq q$ reads

$$q^{n+1} = q_F^{n+1} + \frac{2\Delta t\ G_q\ \mu_q}{1+2\Delta t\ G_q\ \mu_q} \cdot \left[\frac{\sum_k (w_k^2 \cdot [q_k^{obs} - q_F^{n+1}(\mathbf{x}_k)])}{\sum_{k'} w_{k'}^2} \right] \quad (3.126)$$

With $\tilde{W}_k \doteq w_k^2 \ / \ \sum_{k'} w_{k'}^2$, the fractional term in the square brackets can be written as

$$\begin{aligned} \sum_k \tilde{W}_k \cdot [q_k^{obs} - q_F^{n+1}(\mathbf{x}_k)] &= \sum_k \tilde{W}_k \cdot [(q_F^{n+1}(\mathbf{x}) + q_k^{obs} - q_F^{n+1}(\mathbf{x}_k)) - q_F^{n+1}(\mathbf{x})] \\ &\doteq \sum_k \tilde{W}_k \cdot [q_k^{pso} - q_F^{n+1}(\mathbf{x})] \end{aligned} \quad (3.127)$$

Thus, the observation increments determined at the observation locations can be replaced by increments between pseudo observations q_k^{pso} defined at the target grid point and the model value at that point. The pseudo observations are equal to that model value plus the original observation increments. Specific humidity can be expressed as a function f_q of relative humidity U, temperature T, and pressure p, where

$$f_q(U,T,p) = \frac{(R_d/R_v)\,(p_{sw}^v(T)/p)\cdot U}{1 - (1 - R_d/R_v)\,(p_{sw}^v(T)/p)\cdot U} \doteq \frac{a\ U}{1 - b\ U} = q \quad (3.128)$$

p_{sw}^v is the saturation vapour pressure over water, R_d and R_v are the gas constants for dry air respectively water vapour. Hence, the pseudo observations of specific humidity can be expressed as a function of model temperature and pressure and of relative humidity pseudo observations U_k^{pso} which are given by $U_k^{pso} = U_F^{n+1}(\mathbf{x}) + \Delta_O U_k$ where $\Delta_O U_k$ are the relative humidity observation increments at the observation locations. Taking into account that $\sum_k \tilde{W}_k = 1$, the right side of Eq. (3.127) can then be written as

$$\begin{aligned} \sum_k \tilde{W}_k \cdot [f_q(U_k^{pso}) - q_F^{n+1}(\mathbf{x})] &= \sum_k \tilde{W}_k \cdot \frac{a\left(U_F^{n+1}(\mathbf{x}) + \Delta_O U_k\right)}{1 - b\left(U_F^{n+1}(\mathbf{x}) + \Delta_O U_k\right)} - q_F^{n+1}(\mathbf{x}) \\ &\cong \sum_k \tilde{W}_k \cdot \frac{a\left(U_F^{n+1}(\mathbf{x}) + \Delta_O U_k\right)}{1 - b\left(U_F^{n+1}(\mathbf{x}) + \sum_{k'} \tilde{W}_{k'} \Delta_O U_{k'}\right)} - q_F^{n+1}(\mathbf{x}) \end{aligned} \quad (3.129)$$

Since b is of the order of 0.01 or less, the error related to the approximation which is applied to the denominator in Eq. (3.129) for convenience is typically of the order of 0.001 or less. Note that Eq. (3.129) is exact for grid points which are influenced by only one observation or by several observations with the same value for the observation increments. Plugging Eqs. (3.128) and (3.129) into Eq. (3.126) yields the final form of the nudging equation for humidity:

$$q^{n+1} = q_F^{n+1} + \frac{2\Delta t\ G_q\,\mu_q}{1+2\Delta t\,G_q\mu_q}\cdot\left[\frac{\frac{R_d}{R_v}\,\frac{p_{sw}^v}{p}\cdot\left(U_F^{n+1}+\frac{\sum_k(w_k^2\cdot[U_k^{obs}-U_F^{n+1}(\mathbf{x}_k)])}{\sum_{k'}w_{k'}^2}\right)}{1-\left(1-\frac{R_d}{R_v}\right)\frac{p_{sw}^v}{p}\cdot\left(U_F^{n+1}+\frac{\sum_l(w_l^2\cdot[U_l^{obs}-U_F^{n+1}(\mathbf{x}_l)])}{\sum_{l'}w_{l'}^2}\right)} - q_F^{n+1}\right] \quad (3.130)$$

As a result of the temperature and humidity change from the above steps, subsaturation in the presence of cloud water and supersaturation can occur. Hence, a saturation adjustment as described in Part II of the COSMO documentation is performed by means of evaporation and condensation to finally render 100 % relative humidity in areas with cloud water.

Next, pressure increments above the lowest model level are computed by means of the hydrostatic upper-air pressure correction (Section 3.8.4) and added to the model fields. This has also an impact on saturation vapour pressure, however since it is very small, the saturation adjustment is not performed here again (but it is redone after the boundary relaxation which follows the nudging step).

At last, the analysis increments for horizontal wind are determined. Analogously to the temperature correction, the inclusion of the geostrophic wind correction $(\Delta u_g\,,\ \Delta v_g)$ (Section 3.8.2) is regarded as part of a step-by-step process. Apart from the interpolation of the net weights μ and the net observation increments from the center to the faces of the Arakawa C grid boxes, the two nudging equations for the wind components take the same form as that for temperature, i.e.

$$u^{n+1} = u_F^{n+1} + \frac{2\Delta t\ G_u\,\overline{\mu_u}^{\lambda}}{1+2\Delta t\,G_u\overline{\mu_u}^{\lambda}}\cdot\overline{\left(\frac{\sum_k(w_k^2\cdot[u_k^{obs}-u_F^{n+1}])}{\sum_{k'}w_{k'}^2}\right)}^{\lambda} + \frac{1}{1+2\Delta t\,G_u\overline{\mu_u}^{\lambda}}\cdot\Delta u_g \quad (3.131)$$

$$v^{n+1} = v_F^{n+1} + \frac{2\Delta t\ G_v\,\overline{\mu_v}^{\phi}}{1+2\Delta t\,G_v\overline{\mu_v}^{\phi}}\cdot\overline{\left(\frac{\sum_k(w_k^2\cdot[v_k^{obs}-v_F^{n+1}])}{\sum_{k'}w_{k'}^2}\right)}^{\phi} + \frac{1}{1+2\Delta t\,G_v\overline{\mu_v}^{\phi}}\cdot\Delta v_g \quad (3.132)$$

It is noted that in fact, these computations for wind are done for all model levels only after the previous steps for the nudging of the mass field variables have been completed for all levels (provided that the geostrophic wind correction is computed and / or the program runs in parallel mode).

Last but not least, it is mentioned that with the current use of observations and choice for the temporal weight functions, the analysis increments usually change only slowly, gradually and often almost linearly in time. Hence, it is not necessary to recompute these increments at each timestep. Instead, as an approximation, they can be determined once for the mean time of a period of several timesteps and be (re-)used at each timestep within the period. These periods are regular except when the reading of new observations in the observation processing (once per hour) makes an additional re-computation of the increments meaningful. At timesteps, when the analysis increments are reused, all the computations as described above can be omitted except for the simple addition of the analysis increments to the model fields, the saturation adjustment, and the hydrostatic upper-air pressure correction (which ensures an exact hydrostatic balance of the total analysis increments). In this way, a large fraction of the computing time required for the nudging can be saved.

Section 4

Variational Soil Moisture Analysis

4.1 Overview

As noted in section 3.3.2, 2–m temperature is not assimilated by nudging since it potentially disturbes the stability in the boundary layer. Rather, 2–m temperature observations are assimilated by the variational soil moisture analysis, as presented in the following. Soil water content influences screen–level values for temperature and relative humidity during clear–sky days. An inaccurate specification can result in forecast errors up to several degrees centigrade. Since direct measurements of soil moisture contents are rarely available, an indirect determination is necessary. This is done by a variational method: The optimal soil moisture contents minimize a cost functional that expresses the differences between model derived and observed screen–level temperature and humidity[1] (cf. Hess (2001), J.-F. (1991), Callies et al. (1998), Rhodin et al. (1999), Bouyssel et al. (1999)).

Since, basically, the soil moisture fields are adapted so that the model screen–level forecasts approximate the observed values, the retrieved soil moisture fields depend essentially on the used forecast model, especially on the applied soil and boundary layer parameterizations. Errors of the model forecasts are reflected in errors of the retrieved values. However, the resulting forecasts are improved and the computed heat and moisture fluxes at the bottom of the atmosphere model have a good chance to be improved as well.

Since the soil–atmosphere coupling is strongest with high radiative impact, observations close to noon are assimilated. The diurnal variation of 2–m temperature and 2–m relative humidity have to be provided by the physical parameterizations.

The soil moisture analysis is applied once per day in order to provide improved soil moisture fields to be used by the forecast that starts at the following day. Because the soil–atmosphere coupling is not always strong enough to derive sufficient information on the soil moisture contents to compute them in a reliable way, a Kalman filter cycled analysis is applied that incorporates a background state along with background error estimates. In case of low radiative impact, the retrieved moisture fields remain close to the background. High impact, on the other hand, results in improved moisture fields and reduced background error estimates.

[1] Currently only 2–m temperature observations are assimilated; 2–m relative humidity observations can be included in the analysis in a straight forward way as soon as the model forecast values are in comparative quality.

The variational Kalman–filter analysis scheme that is documented in this section requires one additional forecast run for each analyzed soil moisture layer, but does not depend on a specific soil model nor on a certain boundary layer model and is applicable for general numerical models and technical environments.

4.2 Variational Analysis

The variational analysis scheme derives improved soil moisture contents by minimizing a cost functional. The minimization problem is of high dimension in general; the moisture contents of each horizontal grid column of every soil layer has to be retrieved. However, since the screen–level values for temperature and relative humidity are mainly vertically coupled to the soil moisture contents of the same grid column (at least in case of clear–sky conditions with strong soil moisture–atmosphere coupling), a horizontal decoupling is assumed that reduces the high–dimensional minimization problem to a large series (one for each horizontal grid point) of low–dimensional (the number of analyzed soil layers) minimizations[2]. In this way the computational requirements are essentially reduced.

In the following the formulation of the variational soil moisture analysis scheme is given for an arbitrary horizontal grid point and for the assimilation of 2–m temperature observations only.

Let η and η^b denote vectors of dimension n^{soil} containing the moisture contents of the analyzed soil layers and their background states, respectively. The vectors T^o and $T(\eta)$ of dimension n^{obs} contain analyses (based on synoptic observations) and model forecasts, respectively, of 2–m temperature for specified observation times. The cost functional $\mathcal{J}$ to be minimized at each analysis step (i. e. daily) reads

$$\mathcal{J}(\eta) = \mathcal{J}^o(\eta) + \mathcal{J}^b(\eta) \tag{4.1}$$

with the observation term

$$\mathcal{J}^o(\eta) = \frac{1}{2}\Big(T^o - T(\eta)\Big)^T \mathbf{R}^{-1} \Big(T^o - T(\eta)\Big) \tag{4.2}$$

and the background term

$$\mathcal{J}^b(\eta) = \frac{1}{2}(\eta - \eta^b)^T \mathbf{B}^{-1} (\eta - \eta^b) \quad , \tag{4.3}$$

with $ADP \leq \eta_j \leq PV\,, \quad j=1,\ldots,n^{soil}$. The components of η (indicated with lower indices) are limited by air dryness point (ADP) and pore volume (PV) of the actual soil type.

Matrix $\mathbf{R} \in \mathbb{R}^{n^{obs}\times n^{obs}}$ denotes the observation error covariance matrix and $\mathbf{B} \in \mathbb{R}^{n^{soil}\times n^{soil}}$ the background error covariance matrix. Both matrices $\mathbf{R}$ and $\mathbf{B}$ are symmetric and positive definite for physical reasons. Matrix $\mathbf{R}$ is assumed constant and diagonally dominant. At the start of the cycled soil moisture analysis scheme matrix $\mathbf{B}$ is initialized with estimated error variances and error covariances of the first guess soil moisture fields that are used as initial background η^b . Further on, the background values η^b and the background error covariance matrix $\mathbf{B}$ are provided within the cycled Kalman filter analysis scheme, see Section 4.2.2 .

[2] Experiments have been carried out that confirm the assumption of horizontal decoupling for grid sizes of 14 km and 7 km.

Minimization of $\mathcal{J}$ results in the analyzed soil moistures η^a :

$$\mathcal{J}(\eta^a) \leq \mathcal{J}(\eta) \quad \text{for all} \quad \eta \neq \eta^a \quad . \tag{4.4}$$

The minimization is performed under the assumption that the linearization given in Section 4.2.1 is valid; a unique minimum is guaranteed.

4.2.1 Minimization

Although the soil moisture – 2–m temperature dependency is nonlinear in general, linearization around the background state provides good approximations as long as the retrieved values are not too far from the background state.

Linearization of the model 2–m temperature $T(\eta)$ around η^b gives

$$T(\eta) \doteq T(\eta^b) + \mathbf{\Gamma}\,(\eta - \eta^b) \quad , \tag{4.5}$$

where the Jacobian $\mathbf{\Gamma} \in \mathbb{R}^{n^{obs} \times n^{soil}}$ is approximated by one–sided finite differences,

$$\Gamma_{i,j} := \min\left(\frac{T_i(\eta^j) - T_i(\eta^b)}{\eta^j_j - \eta^b_j} \;,\; 0 \right) \quad , \tag{4.6}$$

with $i = 1, \ldots, n^{obs}$ and $j = 1, \ldots, n^{soil}$. The approximation (4.6) requires the routine forecast based on the background moisture contents η^b and n^{soil} additional forecast runs with varied soil moisture contents η^j . The partial derivatives are known to be negative for physical reasons.[3] The components of vector η^j are computed by

$$\eta^j_k = \begin{cases} \eta^j_j & \text{for} \quad k = j \\ \eta^b_j & \qquad\quad k \neq j \end{cases} \quad , \quad k = 1, \ldots, n^{soil} \quad , \tag{4.7}$$

where the varied soil moisture content η^j_j is altered depending on air dryness point (ADP) and field capacity (FC) of the soil model in order to reduce the influence of the soil type of the actual horizontal grid point. The size of the alteration $\Delta\eta$ is given by

$$\Delta\eta = (FC - ADP)\,\Delta\epsilon \quad , \tag{4.8}$$

the fraction $\Delta\epsilon \in]0, \frac{1}{2}[$ is a tuning parameter. The direction of the variation is chosen according to the forecast error as long as the background η^b is not too close to the limits ADP and FC, i. e. :

$$\eta^j_j = \begin{cases} \min(\eta^b_j, FC) + \Delta\eta & \text{for} \begin{cases} \sum_{i=1}^{n^{obs}} T_i(\eta^b) > \sum_{i=1}^{n^{obs}} T^o_i \text{ and } \eta^b_j + \Delta\eta < FC \\ \text{or} \\ \eta^b_j - \Delta\eta \leq ADP \end{cases} \\ \min(\eta^b_j, FC) - \Delta\eta & \text{for} \begin{cases} \sum_{i=1}^{n^{obs}} T_i(\eta^b) \leq \sum_{i=1}^{n^{obs}} T^o_i \text{ and } \eta^b_j - \Delta\eta > ADP \\ \text{or} \\ \eta^b_j + \Delta\eta \geq FC \end{cases} \end{cases} \tag{4.9}$$

[3] Slightly varying cloud covers in the different forecast runs can result in erroneous positive finite differences, that are eliminated by the minimization.

Although soil moisture influences evapotranspiration and 2–m temperature only between ADP and FC (as defined by the applied soil model of COSMO[4]), higher moisture values up to PV are allowed to reduce the instant impact on the soil model and to provide more realistic soil moisture values (e. g. in case of heavy precipitation).

In order to provide reasonable analysis increments for background values η^b that exceed FC, it is necessary to derive non–vanishing soil moisture – 2–m temperature dependencies. The minimization of η_j^b by FC in Equation (4.9) assures that the varied soil moisture contents reside in the sensitive range between ADP and FC even if η_j^b exceeds FC.

Using the linearization (4.5) the gradient of the cost function can be analytically expressed as

$$\nabla \mathcal{J}(\eta^b) = -\mathbf{\Gamma}^T \mathbf{R}^{-1} \left(T^o - T(\eta^b) - \mathbf{\Gamma} (\eta - \eta^b) \right) + \mathbf{B}^{-1} (\eta - \eta^b) \quad . \tag{4.10}$$

For the low–dimensional minimization problem it is highly efficient to solve

$$\nabla \mathcal{J}(\eta^a) \stackrel{!}{=} 0 \tag{4.11}$$

directly[5]. Little calculus gives the minimum η^a of the cost function as

$$\eta^a = \eta^b + \left(\mathbf{\Gamma}^T \mathbf{R}^{-1} \mathbf{\Gamma} + \mathbf{B}^{-1} \right)^{-1} \mathbf{\Gamma}^T \mathbf{R}^{-1} \left(T^o - T(\eta^b) \right) \quad . \tag{4.12}$$

Equation (4.12) is the formula that is implemented to actually compute η^a.

Worth to mention that the applied minimization by linearization and direct solution is no degradation in accuracy of the retrieved soil moisture fields.

The minimization of the cost functional $\mathcal{J}$ is performed under the constraint that the analyzed values η^a are in the range

$$ADP \leq \eta_j^a \leq PV \quad , \quad j = 1, \ldots, n^{soil} \quad . \tag{4.13}$$

If the global minimum of the cost functional resides outside this valid range, the minimum at the boundaries is computed to result in the analyzed values. The algorithm that computes the minimum considering the boundary side constraints is rather technical and will not be reported here[6].

4.2.2 Kalman–filter Cycling

The soil moisture analysis is performed daily for 0 UTC. For the start of the cycled soil moisture analysis scheme the background error covariance matrix $\mathbf{B}$ is initialized with estimated error variances and covariances $\mathbf{B}^0$ of first guess moisture fields that are used as initial background η^b.

The background state $(\eta^b)^{t+1}$ and the background error covariance matrix $(\mathbf{B})^{t+1}$ for the following day are provided in a Kalman–filter cycled analysis (the valid times of the variables η^a, η^b, $\mathbf{A}$, and $\mathbf{B}$ are indicated from now on by upper indices outside brackets). An

[4] Similar boundaries should exist for other soil models.

[5] Because of the linearization a unique minimum is guaranteed.

[6] A simpler approach would be to compute the global minimum and restrict the moisture values in both soil layers independently to the valid range. Experiments showed only slight differences in the analyses.

increase of confidence in the retrieved soil moisture values due to the assimilated screen–level observations as well as a decrease due to the model error of the soil model are taken into account.

The background soil moisture contents for the following day $(\eta^b)^{t+1}$ are computed as

$$(\eta^b)^{t+1} = (\eta^a)^t + \left(M_t^{t+1}((\eta^b)^t) - (\eta^b)^t\right) \quad , \tag{4.14}$$

where $M_t^{t+1}((\eta^b)^t)$ are the 24 h model values that result from the routine forecast that is started at 0 UTC with the background fields $(\eta^b)^t$. Changes in soil moisture contents by precipitation and evapotranspiration during the 24 hours are taken into account in this way without the requirement of another additional forecast run.

The confidence in the retrieved values $(\eta^a)^t$ is given by the analysis error covariance matrix $\mathbf{A} \in \mathbb{R}^{n^{soil} \times n^{soil}}$,

$$(\mathbf{A})^t = \left(\nabla^2 \mathcal{J}\right)^{-1} = \left(\mathbf{\Gamma}^T \mathbf{R}^{-1} \mathbf{\Gamma} + ((\mathbf{B})^t)^{-1}\right)^{-1} \quad , \tag{4.15}$$

which is the inverse of the Hessian of $\mathcal{J}$ (e. g. Tarantola (1987)). If there is little dependence of the moisture contents on the 2–m temperatures ($\mathbf{\Gamma} \approx 0$), then $(\mathbf{A})^t$ almost equals the background error covariance matrix $(\mathbf{B})^t$. The larger the dependence is, the smaller the estimated analysis errors become.

An auxiliary new background error covariance matrix $(\widetilde{\mathbf{B}})^{t+1}$ is computed by

$$(\widetilde{\mathbf{B}})^{t+1} = \mathbf{M}\,(\mathbf{A})^t\,\mathbf{M}^T + \mathbf{Q} \quad , \tag{4.16}$$

where matrix $\mathbf{M} \in \mathbb{R}^{n^{soil} \times n^{soil}}$ is an estimation of the tangent linear of the forecast operator M_t^{t+1}. Matrix $\mathbf{Q} \in \mathbb{R}^{n^{soil} \times n^{soil}}$ expresses the assumed error of M_t^{t+1}. This additive term reduces the sensitivity of the background to past observations and is important to keep the retrieved moisture contents variable in long–term cycled analyses.

In case of weak soil–atmosphere coupling ($\mathbf{\Gamma} \approx 0$) for a sequence of days, the background error covariance matrices $(\mathbf{B})^{t+1}, (\mathbf{B})^{t+2}, \ldots$ increase linearly with $\mathbf{Q}$, which reflects the reduced confidence in the retrieved moisture fields and larger variations become possible in subsequent analyses. The error variances and error covariances of $\mathbf{Q}$ are the main tuning parameters of the cycled assimilation scheme.

The background error covariance matrix $(\mathbf{B})^{t+1}$ is limited by $\mathbf{B}^{max}$ in order to prevent the background errors from unlimited growth that could affect the stability of the soil moisture analysis in case of longer periods with no soil moisture impacts (e. g. snow). Strictly speaking: the variances (diagonal elements of $(\widetilde{\mathbf{B}})^{t+1}$) are limited by $\mathbf{B}^{max}$ and the covariances (non–diagonal elements) are adjusted in order to result in the same correlation coefficients, i. e. the coefficients of $(\mathbf{B})^{t+1}$ are computed as

$$b_{i,j} = \begin{cases} \min(\tilde{b}_{i,j}, b_{i,j}^{max}) & \text{for} \quad i = j \\ \tilde{b}_{i,j} \sqrt{\frac{b_{i,i}\, b_{j,j}}{\tilde{b}_{i,i}\, \tilde{b}_{j,j}}} & \text{for} \quad i \neq j \end{cases} \quad , \quad i,j = 1, \ldots, n^{soil} \quad .$$

4.2.3 Free Parameters

The tuning parameters of the variational soil moisture analysis scheme are the matrices $\mathbf{R}$, $\mathbf{Q}$, $\mathbf{B}^0$, $\mathbf{B}^{max}$, and $\mathbf{M}$ as well as the fraction $\Delta\epsilon$. (They are controlled by namelist parameters of the model `SMA`.)

Section 5

Snow Analysis

5.1 Overview

Knowledge of the distribution of snow is necessary for the determination of surface albedo and surface fluxes in weather prediction models. The presence of a snow cover reduces dramatically the absorption of short-wave radiance at the ground and hence strongly influences the local near-surface temperature. Moreover, the distribution of snow cover can also have a significant large-scale impact as it may affect the characteristics of air masses.

Therefore, snow depth is analyzed 2-dimensionally in a separate procedure and transformed into the model variable snow water content in order to be used in COSMO. Typically, this is done at a frequency of once every 6 hours during the data assimilation cycle. Note that in the context of short-range numerical weather prediction, it is less important to determine exactly the snow depth than the horizontal extent of the snow-covered areas.

The object of the snow depth analysis is to find a proper value for each model grid point which is representative of the grid box and the height of the model orography. This height is a mean value of the real orography within the grid box.

5.2 Input Data

There are three sources of information that are used in the snow depth analysis:

1. surface-level synoptic observations (SYNOP);
2. model 'forecast' values of snow water content transformed into snow depth from the COSMO nudging run for data-pour areas where the total influence (weight) of the SYNOP observations is below a given threshold;
2. monthly snow depth climatology (from ECMWF) for defining permanently ice-covered glacial areas.

Depending on the information content, different data are extracted from SYNOP reports:

- Total snow depth. If this is not reported then also:
- 6-hourly precipitation sum. This is converted into a snow depth increment provided that the 2-m temperature T_{2m} (observed or, if it is not observed, model-derived) is

below $0°\,\mathrm{C}$ and the present or past weather observations ww indicate snowfall. If 6-hourly precipitation is also missing, then also:

- Present and past weather observations. In case of reported snowfall, a snow depth increment is calculated using an empirical function of ww. Other weather observations result in zero snow depth increments.

Quality Control

At first, observed snow depth is subject to a plausibility check. It is rejected if it exceeds an acceptance limit $d_{sn}^{al} = 1.5\,[m] \cdot (1 + z_{ob}/800\,[m])$ which depends on station height z_{ob}.

Then, a first guess quality control check is performed. Here, the previous snow depth analysis $d_{sn}^{a(t-1)}$ is regarded as a first guess for truth, and the observation is rejected if it deviates from that guess by more than a threshold value d_{sn}^{thr} given by

$$d_{sn}^{thr} \;=\; 0.8\,[m] \cdot \left(1 + \frac{z_{ob}}{2000\,[m]}\right) \cdot \max\left(0, \min\left(1, \frac{287.16\,[K] - T_{2m}}{10\,[K]}\right)\right) \qquad (5.1)$$

5.3 Analysis Method

The analysis method used for snow depth is based on a simple weighted averaging of observed values and does not use a background field except in data-pour regions. The individual weight w_k of an observation k at a target grid point depends both on the horizontal distance Δr and the vertical distance Δz between these two points,

$$w_k \;=\; \max\left(\frac{s^2 - \Delta r^{\,2}}{s^2 + \Delta r^{\,2}}\,,\; 0\right) \;\cdot\; \max\left(\frac{z_s^2 - \Delta z^{\,2}}{z_s^2 + \Delta z^{\,2}}\,,\; 0\right) \qquad (5.2)$$

The weight is zero for observations which are further away from the target point than the radii of influence s and z_s. The horizontal radius s is set to 120 km in data-dense areas and to 200 km elsewhere, whilst the vertical radius of influence $z_s = 0.4 \cdot z + 180\,\mathrm{m}$ increases with increasing height z of the target point.

Using Eq. (5.2), the weighted averages of snow depth observations $\overline{d_{sn}}^{ob}$ and of snow depth increments $\overline{\Delta d_{sn}}^{ob}$ derived from precipitation and weather observations are calculated as well as the sum of weights $W^d = \sum w_k^d$ resp. $W^{\Delta d} = \sum w_k^{\Delta d}$. The total weights W^d and $W^{\Delta d}$ are measures for the local data density of the two types of observations.

As these data densities may vary strongly in space and time, the way to determine the final analyzed snow depth depends on them. If the sum of weights of snow depth data is greater than a prescribed value W_{suf} then the analyzed snow depth d_{sn}^a at that grid point is set simply to the weighted mean value of observed snow depth, i.e.

$$d_{sn}^a \;=\; \overline{d_{sn}}^{ob} \;=\; \sum_k w_k \cdot d_{sn}^k \;/\; W^d \qquad , \qquad W^d \;\geq\; W_{suf} \;\equiv\; 2 \qquad (5.3)$$

If this condition is not satisfied but the mean of the two data densities is greater than W_{suf}, i.e. if $\frac{1}{2} \cdot (W^d + W^{\Delta d}) \;\geq\; W_{suf} \;>\; W^d$, then (only) the weighted mean of snow depth increments $\overline{\Delta d_{sn}}^{ob}$ is additionally taken into account. The weighted mean increment

is added to the previous analysis $d_{sn}^{a(t-1)}$, and a possible snow melt Δd_{sn}^{melt} is subtracted. The result is combined with the mean weighted observed snow depth so that

$$d_{sn}^{a} = \frac{W^d}{W_{suf}} \cdot \overline{d_{sn}}^{ob} + \left(1 - \frac{W^d}{W_{suf}}\right) \cdot \left(d_{sn}^{a(t-1)} + \overline{\Delta d_{sn}}^{ob} - \Delta d_{sn}^{melt}\right) \tag{5.4}$$

Whenever the mean of the two data densities is also smaller than W_{suf}, the third data source, i.e. the 6-hour model forecast of snow water content converted into snow depth d_{sn}^{fc}, is also used. Then, the analyzed value is given by

$$d_{sn}^{a} = \frac{W^d}{W_{suf}} \cdot \overline{d_{sn}}^{ob} + \frac{W^{\Delta d}}{2\,W_{suf}}\left(1 - \frac{W^d}{W_{suf}}\right) \cdot \overline{d_{sn}}^{inc} + \left(1 - \frac{W^{\Delta d}}{2\,W_{suf}}\right)\left(1 - \frac{W^d}{W_{suf}}\right) \cdot d_{sn}^{fc} \tag{5.5}$$

where $\overline{d_{sn}}^{inc} = d_{sn}^{a(t-1)} + \overline{\Delta d_{sn}}^{ob} - \Delta d_{sn}^{melt}$.

At grid points where climatological data indicate permanent ice cover the snow depth is set to 100 m. In a final check, it is ensured that the analysis increment at any grid point does not exceed a limit that depends on height and temperature.

Section 6

Analysis of Sea Surface Temperature and Ice Cover

Overview

Over water, the sensible and latent heat fluxes at the surface strongly depend on the surface temperature. The correctness of its specification is a prerequisite for realistic simulations of cyclonic development, particularly in baroclinic regions. In addition, the location of the sea ice boundary is also very important since strong vertical heat fluxes can be induced when cold air masses are advected from ice-covered areas to the open water.

Therefore, sea surface temperature and the location of the sea ice boundary are analyzed 2-dimensionally in another separate procedure. Typically, this analysis is performed once a day at 0 UTC as part of the data assimilation.

The object of the sea surface temperature analysis for COSMO is to capture not only the slowly varying large-scale temperature patterns, but also smaller-scale phenomena such as cold anomalies due to upwelling currents or the relatively rapid warming in shallow coastal areas in periods of strong solar irradiation. However, diurnal or very small-scale variations should be filtered out because sea surface temperature is held constant during the forecast.

Analysis Method

Firstly, the extent of the sea ice cover is analyzed. For this purpose, external analyses are directly interpolated to the model grid. Namely for the Baltic Sea, a weekly analysis from the Federal Maritime and Hydrographic Agency of Germany (BSH) with a latitudinal–longitudinal resolution of 0.1 by $0.1\overline{6}$ degrees is used. For other areas, a daily global analysis (Grumbine (1996)) from the Ocean Modelling Branch of NCEP based on SSMI satellite data is available at a resolution of 0.5 by 0.5 degrees and could be used if required. Sea ice temperature is currently derived from ECMWF climatology, however it is planned to compute this quantity by means of a one-layer sea ice model incorporated in COSMO in the foreseeable future.

Sea surface temperature (SST) is analyzed by means of a correction scheme. The analysis is determined by adding weighted observation increments to a first guess or background field in the environs of the observations. As a first guess field, the interpolated SST analysis of GME is deployed for which the first guess field is given by a 0.5 by 0.5 degree SST analysis

of NCEP. The latter analysis does not only use in-situ observations but also bias-corrected satellite data (eynolds and Smith (1994)). Hence, the SST analysis of COSMO does also benefit from these satellite data.

The observational data set used for the SST analysis of COSMO comprises of all the ship and buoy data from within the previous 5 days. As a quality control, the data are checked against the first guess and against observations from the other stations in the vicinity. For the analysis at each grid point, the first guess value is corrected by a weighted mean of all the observation increments. The individual weights depend on the distance between analysis and observation time, on the observation type, and on the spatial distance of the observation from the target point according to Eq. (5.2) with $\Delta z = 0$ and $s = 200\,\mathrm{km}$ ($s = 430\,\mathrm{km}$ for the SST analysis of GME).

References

A., S. (1989). A GCM parameterization for the shortwave radiative properties of water clouds. *J. Atmos. Sci. 46*, 1419–1427.

Anthes, R. A. (1974). Data assimilation and initialization of hurricane prediction models. *J. Atmos. Sci. 31*, 702–719.

Barwell, B. R. and A. C. Lorenc (1985). A study of the impact of aircraft wind observations on a large scale analysis and numerical weather prediction system. *Quart. J. Roy. Meteor. Soc. 111*, 103–129.

Bell, R. S., A. C. Lorenc, B. Macphersona, R. Swinbank, and P. Andrews (1996). *The analysis correction data assimilation scheme*. UK Met Office.

Benjamin, S. (1989). An isentropic meso-α-scale analysis system and its sensitivity to aircraft and surface observations. *Mon. Wea. Rev. 117*, 1586–1603.

Benjamin, S. and N. Seaman (1985). A simple scheme for objective analysis in curved flow. *Mon. Wea. Rev. 113*, 1184–1198.

Bevis, M., S. Businger, S. Chiswell, T. A. Herring, R. A. Anthes, C. Rocken, and R. H. Ware (1994). GPS Meteorology: Mapping Zenith Wet Delays onto Precipitable Water. *J. Appl. Meteor. 33*(3), 379–386.

Bouyssel, F., V. Cassé, and J. Pailleux (1999). Variational surface analysis from screen level atmospheric parameters. *Tellus A 51*, 453–468.

Callies, U., A. Rhodin, and D. P. Eppel (1998). A case study on variational soil moisture analysis from atmospheric observations. *J. Hydr. 212-213*, 95–108.

Daley, R. (1985). The analysis of synoptic scale divergence by a statistical interpolation procedure. *Mon. Wea. Rev. 113*, 1066–1079.

Davies, H. and R. Turner (1977). Updating prediction models by dynamical relaxation: An examination of the technique. *Quart. J. R. Met. Soc. 103*, 225–245.

eynolds, R. W. and T. M. Smith (1994). Improved global sea surface temperature analysis. *J. Climate 7*, 929–948.

Grumbine, R. W. (1996). *Automated passive microwave sea ice concentration analysis at NCEP*. NCEP.

Hess, R. (2001). Assimilation of screen-level observations by variational soil moisture analysis. *Meteorol. Atmos. Phys. 77*, 155–166.

Holton, J. (1992). *An Introduction to Dynamic Meteorology (3rd edition)*. Academic Press, San Diego.

J.-F., M. (1991). Analysis of soil moisture from near-surface parameters. A feasibility study. *J. Appl. Meteor. 7*, 506–526.

Jacobsen, I. and E. Heise (1982). A new economic method for the computation of the surface temperature in numerical models. *Contr. Atmos. Phys. 55*, 128–141.

Leiterer, U., H. Dier, D. Nagel, T. Naebert, D. Althausen, and K. Franke (2000). Correction method for RS80 A-Humicap humidity profiles. Internal report, Deutscher Wetterdienst. available from Meteorological Observatory Lindenberg.

Lorenc, A. C., R. S. Bell, and B. Macpherson (1991). The Meteorological Office analysis correction data assimilation scheme. *Quart. J. Roy. Meteor. Soc. 117*, 59–89.

Louis, J.-F. (1979). A parametric model of vertical eddy fluxes in the atmosphere. *Bound. Layer Meteor. 17*, 187–202.

Lynch, P., D. Girard, and V. Ivanovici (1997). Improving the efficiency of a digital filtering scheme. *Mon. Wea. Rev. 125*, 1976–1982.

Miloshevich, L. M., H. Vömel, D. N. Whiteman, and T. Leblanc (2009, 06). Accuracy assessment and correction of Vaisala RS92 radiosonde water vapor measurements. *J. Geophys. Res. 114*.

Mironov, D. (2008). Parameterization of lakes in numerical weather prediction. Description of a lake model. COSMO Technical Report 11, Consortium for Small-Scale Modelling, 41 pp. doi:10.5676/DWD_pub/nwv/cosmo-tr_11 (available from http://www.cosmo-model.org).

Mironov, D. and B. Ritter (2004). Testing the new ice model for the global NWP system GME of the German Weather Service. Technical documentation, 1220, WMO, WMO, Geneve, Switzerland.

Okland, H. (1970). On the adjustment toward balance in primitive equation weather prediction models. *Mon. Wea. Rev. 98*, 271–279.

Rhodin, A., F. Kucharski, U. Callies, D. P. Eppel, and W. Wergen (1999). Variational analysis of effective soil moisture from screen-level atmospheric parameters: Application to a short-range weather forecast model. *Quart. J. Roy. Meteor. Soc. 125*, 2427–2448.

Ritter, B. and J.-F. Geleyn (1992). A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations. *Mon. Wea. Rev. 120*, 303–325.

Schär, C., D. Leuenberger, O. Fuhrer, D. Lüthi, and C. Girard (2002). A new terrain-following vertical coordinate formulation for atmospheric prediction models. *Mon. Wea. Rev. 130*, 2459–2480.

Schraff, C. (1996). Data assimilation and mesoscale weather prediction: A study with a forecast model for the Alpine region. Publication 56, Swiss Meteorological Institute, Zürich.

Schraff, C. (1997). Mesoscale data assimilation and prediction of low stratus in the Alpine region. *Meteorol. Atmos. Phys. 64*, 21–50.

Schraff, C., H. Reich, A. Schomburg, K. Stephan, A. Periáñez, and R. Potthast (2016). Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). *Q. J. R. Meteorol. Soc. 142*, 1453–1472.

Skamarock, W. and J. Klemp (1992). The stability of time-split numerical methods for the hydrostatic and nonhydrostatic elasic equations. *Mon. Wea. Rev. 120*, 2109–2127.

Stauffer, D. and J.-W. Bao (1993). Optimal determination of nudging coefficients using the adjoint equations. *Tellus 45A*, 358–369.

Stauffer, D. and N. Seaman (1990). Use of four-dimensional data assimilation in a limited-area mesoscale model. Part I: Experiments with synoptic-scale data. *Mon. Wea. Rev. 118*, 1250–1277.

Stauffer, D. and N. Seaman (1994). Multiscale four-dimensional data assimilation. *J. Appl. Meteor. 33*, 416–434.

Stauffer, D. R., N. L. Seaman, and F. S. Binkowski (1991). Use of four-dimensional data assimilation in a limited-area mesoscale model. Part II: Effects of data assimilation within the planetary boundary layer. *Mon. Wea. Rev. 119*, 734–754.

Stephan, K., S. Klink, and C. Schraff (2008). Assimilation of radar-derived rain rates into the convective-scale model COSMO-DE at DWD. *Quart. J. R. Meteor. Soc. 134*, 1315–1326.

Tarantola, A. (1987). *Inverse Problem Theory*. Elsevier Science.

Thomas, S., C. Girard, G. Doms, and U. Schättler (2000). Semi-Implicit Scheme for the DWD Lokal-Modell. *Meteorol. Atmos. Phys. 73*, 105–125.

Tiedtke, M. (1989). A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev. 117*, 1779–1799.

Wicker, L. and W. Skamarock (2002). Time-splitting methods for elastic models using forward time schemes. *Mon. Wea. Rev. 130*, 2088–2097.

Woodage, M. (1985). The preparation of data for the Meteorological Office 15-level forecast model. *Meteorol. Mag. 114*, 1–13.