3 Model System Overview

The limited-area model LM is designed as a flexible tool for numerical weather prediction on the meso-β and on the meso-γ scale as well as for various scientific applications using grid spacings from 50 km down to about 50 m. 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 is required to run a NWP system at a meteorological service. In the following sections, the components of the LM package - as available to the COSMO group - are briefly described.

3.1 Short Description of the LM

This section gives only a brief overview of the Lokal-Modell. For a comprehensive description, the reader is referred to the documentation of the LM package (see section 3.6). An overview is given by Steppeler et al. (2003).

3.1.1 Dynamics and Numerics

The regional model LM is based on the primitive hydro-thermodynamical equations describing compressible nonhydrostatic flow in a moist atmosphere without any scale approximations. A basic state is subtracted from the equations to reduce numerical errors associated with the calculation of the pressure gradient force in case of sloping coordinate surfaces. The basic state represents a time-independent dry atmosphere at rest which is prescribed to be horizontally homogeneous, vertically stratified and in hydrostatic balance. The basic equations are written in advection form and the continuity equation is replaced by a prognostic equation for the perturbation pressure, i.e. the deviation of pressure from the reference state.

The model equations are formulated with respect to a rotated lat/lon-grid with coordinates (λ, φ). The rotated coordinate system results from the geographical (λ_g, φ_g) coordinates by tilting the north pole. In the vertical, we use a generalized terrain-following height coordinate ζ, where any unique function of geometrical height can be used for transformation. Since ζ does not depend on time, the $(\lambda, \varphi, \zeta)$-system represents a non-deformable coordinate system, where surfaces of constant ζ are fixed in space - in contrast to the pressure based coordinate system of most hydrostatic models, where the surfaces of constant vertical coordinate move in space with changing surface pressure. By default, a hybrid sigma-type (formulated with respect to the base-state pressure) vertical coordinate is used.

The model equations are solved numerically using the traditional Eulerian finite difference method. In this technique, spatial differential operators are simply replaced by suitable finite difference operators and the time integration is by discrete stepping using a fixed timestep Δt. The model variables are staggered on an Arakawa-C/Lorenz grid with scalars (temperature, pressure and humidity variables) defined at the centre of a grid box and the normal velocity components defined on the corresponding box faces. For a given grid spacing, this staggering allows for a more accurate representation of some differential operators than in the A-grid, where all variables are defined at the same point. In general, we use second order centered finite difference operators for horizontal and vertical differencing.

Because the governing nonhydrostatic equations describe a compressible model atmosphere, meteorologically unimportant sound waves are also part of the solution. As acoustic waves are very fast, their presence severely limits the time step of explicit time integration schemes. In order to improve the numerical efficiency, the prognostic equations are separated into terms

Table 1: LM Model Formulation: Dynamics and Numerics

Model Equations:	Basic hydro-thermodynamical equations for the atmosphere: – advection form, – non-hydrostatic, fully compressible, no scale approximations, – subtraction of horizontally homogeneous basic state at rest.
Prognostic Variables:	Horizontal and vertical Cartesian wind components, temperature, pressure perturbation, specific humidity, cloud water content. Options for additional prognostic variables: – turbulent kinetic energy, cloud-ice, rain, snow and graupel content.
Diagnostic Variables:	Total air density, precipitation fluxes of rain and snow.
Coordinate System:	Rotated geographical (lat/lon) coordinate system horizontally; generalized terrain-following height-coordinate vertically. Built-in options for the vertical coordinate are: – hybrid reference pressure based σ-type coordinate (default), – hybrid version of the Gal-Chen coordinate, – hybrid version of the SLEVE coordinate (Schär et al., 2002).
Grid Structure:	Arakawa C-grid, Lorenz vertical grid staggering.
Spatial Discretization:	Second order horizontal and vertical differencing.
Time Integration:	Leapfrog HE-VI (horizontally explicit, vertically implicit) time-split integration scheme by default; includes extensions proposed by Skamarock and Klemp (1992). Additional options for: – a two time-level split-explicit scheme (2nd order Runge-Kutta scheme (Gassmann, 2002), – a three time-level 3-d semi-implicit scheme (Thomas et al., 2000), – a two time level 3rd-order Runge-Kutta scheme (regular or TVD) with various options for high-order spatial discretization (Förstner and Doms, 2004).
Numerical Smoothing:	4th order linear horizontal diffusion with option for a monotonic version including an orographic limiter (Doms, 2001); Rayleigh-damping in upper layers; 3-d divergence damping and off-centering in split steps.
Lateral Boundaries:	1-way nesting using the lateral boundary formulation according to Davies (1976). Options for: – boundary data defined on lateral frames only, – periodic boundary conditions.

which are directly related to acoustic and gravity wave modes and into terms which refer to comparatively slowly varying modes of motion. This mode-splitting can formally be written in the symbolic form

$$\frac{\partial \psi}{\partial t} = s_\psi + f_\psi \,, \tag{1}$$

where ψ denotes a prognostic model variable, s_ψ the forcing terms due to the slow modes and f_ψ the source terms related to the fast acoustic and gravity wave modes. f_ψ is made up of the pressure gradient terms in the momentum equations, the temperature and pressure contributions to the buoyancy term in the equation for the vertical velocity, and the divergence term in the pressure and the temperature equation. The subset of equations containing the f_ψ-terms is then integrated with a special numerical scheme.

The default time integration method used in LM is a variant of the Klemp and Wilhelmson (1978) scheme including extensions proposed by Skamarock and Klemp (1992). It is based on a Leapfrog integration for the slow modes from time level $n-1$ to time level $n+1$ using an integration interval of $2\Delta t$. The slow mode tendencies are evaluated at time level n for horizontal advection and at time level $n-1$ for most physical forcings. Tendencies from vertical advection and diffusion are calculated by a quasi-implicit scheme. The integration step is then subdivided into a number N_s of small time steps $\Delta\tau_s$ according to $2\Delta t = N_s\Delta\tau$ and the prognostic equations (1) are stepped forward according to

$$\psi^{\nu+1} = \psi^{\nu} + f_{\psi}^{\nu}\Delta\tau + s_{\psi}^{n}\Delta\tau\,. \tag{2}$$

In the integration of (2), sound waves are treated explicitly for horizontal directions using the forward-backward method while implicitly for the vertical direction (HE-VI scheme). Thus, the small time step $\Delta\tau$ is limited by the CFL stability criterion for horizontal but not for vertical sound wave propagation. This makes the HE-VI scheme numerically very efficient for large grid aspect ratios, i.e. $\Delta x/\Delta z \gg 1$, which are typically used in meso-β and meso-γ applications. An additional 3-D divergence damping as well as a slight time off-centering in the vertical implicit formulation is applied to damp acoustic modes. On the big time step, the Asselin time filter and a 4th-order horizontal diffusion are used for numerical smoothing.

Three alternative time integration schemes have also been implemented for optional use: a two time-level time-split method based on a modified 2nd-order Runge-Kutta integration (Gassmann, 2002), a three-timelevel Leapfrog-based Eulerian 3-D semi-implicit scheme according to Thomas et al. (2000) and recently a new two time-level scheme based on 3rd-order Runge-Kutta integration with total variation diminishing (TVD) option (Förstner and Doms, 2004). The latter scheme is intended to be used for high-resolution applications of LM in near future. Table 1 summarizes the dynamical and numerical key features of the LM.

3.1.2 Initial and Boundary Conditions

For operational applications and real data simulations, LM is driven by the global model GME of DWD using the traditional boundary relaxation technique (see Section 3.3). Information on the GME as well as on recent changes to the global model are summarized in the *Quarterly Report of the Operational NWP-Models of the Deutscher Wetterdienst*. This report series is available online at the DWD web-site (*www.dwd.de*). Optionally, initial and boundary data may also be provided from the IFS global model at ECMWF. In this context, a new option for using boundary data which are defined on lateral frames only (by default, the boundary conditions are defined on the full 3-d model domain) has been introduced.

A four-dimensional data assimilation cycle based on a nudging analysis scheme (see Section 3.2) can be installed for operational NWP with the LM at COSMO meteorological services. In this case, the initial conditions come from the continuous LM assimilation stream and only boundary data have to be provided by GME forecasts. However, an operational NWP-system can also be set-up without a data assimilation cycle by relying on pure dynamical adaption of large-scale initial fields. In this case, the initial conditions come from interpolated (and initialized) GME analyses. To reduce noise generation and spin-up effects resulting from non-balanced interpolated data, a diabatic digital filtering initialization (DFI) scheme (Lynch et al., 1997) has been implemented. By default, the DFI initialization consists of a 1-h adiabatic backward integration followed by a 1-h diabatic forward integration of the model.

For various research applications as well as for model testing and evaluation, the LM provides a capability to handle idealized cases using user-defined artificial initial and boundary

data. For these types of application, periodic lateral boundary conditions can be specified optionally. Additionally, a 2-dimensional model configuration can be used.

3.1.3 Parameterization of Physical Processes

A variety of subgrid-scale physical processes is taken into account by parameterization schemes.

Table 2: LM Model Formulation: Physical Parameterizations

Grid-scale Clouds and Precipitation:	Cloud water condensation/evaporation by saturation adjustment. Precipitation formation by a bulk parameterization including water vapour, cloud water, rain and snow (scheme HYDOR), where rain and snow are treated diagnostically by assuming column equilibrium (default). Further options are: – the LM cloud ice scheme HYDCI (Doms, 2002), – a warm rain scheme following Kessler (1969), – prognostic treatment of rain and snow (Gassmann, 2002; Baldauf and Schulz, 2004), – a scheme including graupel content as prognostic variable.
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.
Moist Convection:	Mass-flux convection scheme (Tiedtke, 1989) with closure based on moisture convergence (default). Further Options: – a modified closure based on CAPE within the Tiedtke scheme. – the Kain-Fritsch convection scheme.
Radiation:	δ-two stream radiation scheme after Ritter and Geleyn (1992) for short and longwave fluxes; full cloud-radiation feedback.
Vertical Diffusion:	Diagnostic K-closure at hierarchy level 2 by default. Optional: – a new level 2.5 scheme with prognostic treatment of turbulent kinetic energy; effects of subgrid-scale condensation and evaporation are included and the impact from subgrid-scale thermal circulations is taken into account.
Surface Layer:	Constant flux layer parameterization based on the Louis (1979) scheme (default). Optional: – a new surface scheme including a laminar-turbulent roughness sublayer.
Soil Processes:	Soil model after Jacobsen and Heise (1982) with 2 soil moisture layers and Penman-Monteith transpiration; snow and interception storage are included. Climate values changing monthly (but fixed during forecast) in third layer. Optional: – a new multi-layer soil model including freezing of soil water (Schrodin and Heise, 2001).

Initially, the physics package of LM has been adapted from the former operational hydrostatic models EM/DM. Meanwhile, a number of additional schemes have been developed and implemented for optional use: a new scheme for vertical diffusion based on prognostic turbulent kinetic energy, a new diagnostic scheme for surface layer transports, a new grid-scale cloud and precipitation scheme including cloud ice, rain and snow water content and graupel content as (optional) prognostic variables, a new multi-layer soil model, and the

Kain-Fritsch scheme for deep moist convection. Work on a new lake model, a revised version of the surface layer scheme, a 3D turbulence scheme and a new scheme for shallow convection is in progress.

Table 2 gives a short overview on the parameterization schemes used by default and on additional options implemented so far.

3.1.4 External Parameters

The parameterization of physical processes and also the adiabatic model part require some parameters which are not derived by data assimilation or by interpolation from a driving model. These so-called external parameters are defined in additional data sets. The LM requires the following external parameters: mean topographical height, roughness length, soil type, vegetation cover, land fraction, root depth and leaf area index. The sources for these data are indicated below.

- Mean orography:
  Derived from the GTOPO30 data set (30"x30") from USGS. Also evaluated is the GLOBE dataset from the National Geophysical Data Center (NGDC) with a resolution of 30"x30".
- Prevailing soil type:
  Derived from the DSM data set (5'x5') of FAO.
- Land fraction, vegetation cover, root depth and leaf area index:
  Derived from the CORINE data set of ETC/LC and the GLCC (Global Land Cover Characterization) from USGS.
- Roughness length:
  Derived from the GTOPO30 and CORINE datasets.

External parameters for LM can be derived by a preprocessor program for any domain on the globe at any required spatial resolution. However, this is very time consuming because of the size of the high-resolution global data sets. Within the COSMO group, we thus have prepared some predefined data sets with external parameters on three different domains. These domains are shown in Fig. 1.

Domain 1 covers Europe and surrounding countries; data sets for this domain are available at 28 km, 21 km, 14 km and 7 km grid spacing. The smaller Domain 2 covers Germany and surrounding countries; the corresponding data set gives the external parameters at 7 km resolution. Domain 2 is only used at DWD. Finally, Domain 3 covers central and southern parts of Europe. For this domain, the external parameters are given at 2.8 km resolution. The LM can very easily be positioned anywhere within these domains.

Details on the location of the three domains are shown in Table 3, where longitude (λ) and latitude (ϕ) of the rotated coordinates and those of the geographical lat-lon grid (λ_g, ϕ_g) are given in degree. The resolution and the corresponding file names (these are required by the interpolation programs to generate initial and boundary data from a host model) are indicated in Table 4. The specifications refer to a rotated lat-lon grid of LM with the north-pole at geographical latitude 32.5° N and longitude 170.0° W.

Two new data sets were provided recently for nearly the same area as Domain 1 but with a different rotated pole. In order to keep the size of the grid boxes as equal as possible,

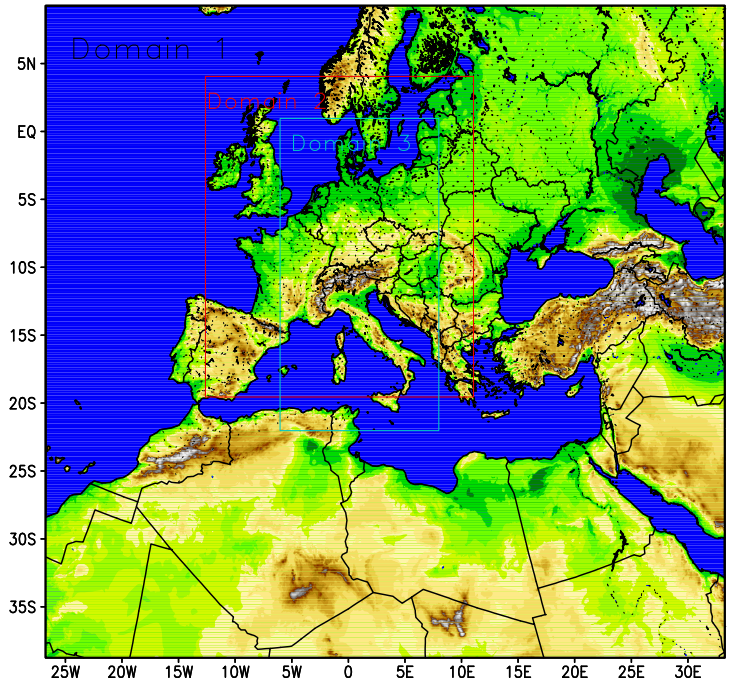

Figure 1: Domains of external parameter datasets used by COSMO partners

Table 3: Location of domains in rotated and in geographical coordinates

Name	Domain corners	λ	φ	λ_g	φ_g
Domain 1	upper left	- 26.75	+ 9.25	- 42.74	+ 56.07
	upper right	+ 33.25	+ 9.25	+ 70.36	+ 51.49
	lower left	- 26.75	- 38.75	- 11.26	+ 14.54
	lower right:	+ 33.25	- 38.75	+ 35.96	+ 12.34
Domain 2	upper left	- 12.625	+ 4.125	- 15.25	+ 59.26
	upper right	+ 11.125	+ 4.125	+ 32.48	+ 59.77
	lower left	- 12.625	- 19.50	- 4.87	+ 36.62
	lower right:	+ 11.125	- 19.50	+ 23.15	+ 36.92
Domain 3	upper left	- 6.00	+ 1.00	- 1.37	+ 58.00
	upper right	+ 8.00	+ 1.00	+ 25.06	+ 57.61
	lower left	- 6.00	- 22.00	+ 3.19	+ 35.20
	lower right:	+ 8.00	- 22.00	+ 19.06	+ 34.97

the equator was put to the center of the domain. The north-pole of this rotated grid is at geographical latitude 40.0° N and longitude 170.0° W. The two resolutions are 7 km (Domain 5) and 2.8 km (Domain 0), resp. These data sets are used by DWD for testing the LME (LM over whole Europe) and the LMK (LM Kürzestfrist: very high resolution LM for nowcasting purposes). The domains are shown in Fig. 2

The specifications of the new datasets are given in Table 5 and Table 6.

Table 4: Grid spacing $\Delta\lambda$ ($= \Delta\varphi$) in degrees, approximate resolution Δs in m, number of grid points and file name of the datasets for the domains

Name	$\Delta\lambda, \Delta\varphi$ (°)	Δs (m)	no. of grid points	File name
Domain 1	0.2500	28000	241 × 193	lm_d1_28000_241x193.g1
	0.1875	21000	321 × 257	lm_d1_21000_321x257.g1
	0.1250	14000	481 × 385	lm_d1_14000_481x385.g1
	0.0625	07000	961 × 769	lm_d1_07000_961x769.g1
Domain 2	0.0625	07000	381 × 379	lm_d2_07000_381x379.g1
Domain 3	0.0250	02800	561 × 921	lm_d3_02800_561x921.g1

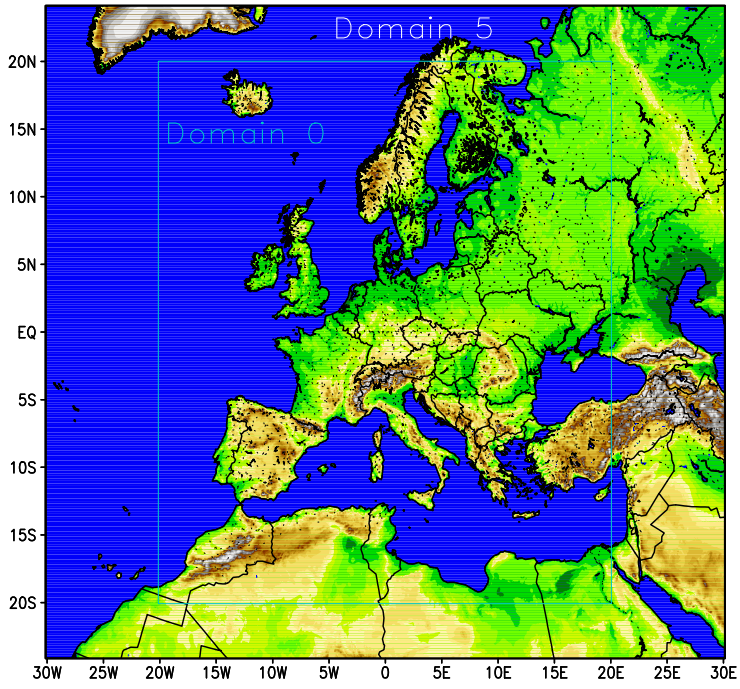

Figure 2: Domains of external parameter datasets with different rotated pole

Table 5: Location of new domains in rotated and in geographical coordinates

Name	Domain corners	λ	φ	λ_g	φ_g
Domain 5	upper left	- 30.125	+ 24.125	- 57.01	+ 60.16
	upper right	+ 30.125	+ 24.125	+ 77.01	+ 60.16
	lower left	- 30.125	- 24.125	- 19.17	+ 20.00
	lower right:	+ 30.125	- 24.125	+ 39.17	+ 20.00
Domain 0	upper left	- 20.05	+ 20.05	- 36.60	+ 63.69
	upper right	+ 20.05	+ 20.05	+ 56.60	+ 63.69
	lower left	- 20.05	- 20.05	- 11.21	+ 27.11
	lower right:	+ 20.05	- 20.05	+ 31.21	+ 27.11

Table 6: Grid spacing $\Delta\lambda$ (= $\Delta\varphi$) in degrees, approximate resolution Δs in m, number of grid points and file name of the datasets for the domains

Name	$\Delta\lambda, \Delta\varphi$ (°)	Δs (m)	no. of grid points	File name
Domain 5	0.0625	07000	965 × 773	lm_d5_07000_965x773.g1
Domain 0	0.025	02800	1605 × 1605	lm_d2_07000_1605x1605.g1

3.1.5 Coding and Parallelization

In order to meet the computational requirements of the model, the program has been coded in Standard Fortran 90 and parallelized using the MPI library for message passing on distributed memory machines. Thus it is portable and can run on any parallel machine providing MPI. Besides this, the model can still be executed on conventional single-processor computers without MPI.

The parallelization strategy is the two dimensional domain decomposition which is well suited for grid point models using finite differences (see Fig. 3). Each processor gets an appropriate part of the data to solve the model equations on its own subdomain. This subdomain is surrounded by halo grid-lines which belong to the neighboring processors. The number of halo grid-lines is soft-coded to be able to choose different values, when necessary. At present, we use two grid-lines for the three time-level leapfrog scheme and three grid-lines for the two time-level Runge-Kutta scheme that uses higher order advection schemes. During the integration step each processor updates the values of its local subdomain; grid points belonging to the halo are exchanged using explicit message passing. The number of processors in longitudinal and latitudinal direction can be specified by the user to fit optimal to the hardware architecture (vector, scalar, cache-size, etc.).

Table 7 shows the timings of the LM for a one-hour full-physics simulation on a 325×325×35 grid points domain using 7 km grid spacing and a 40 sec time step for various numbers of processors on an IBM SP3. Starting with 15 processors - which are required to fit the model into core memory - the number of processors has been doubled up to 240. The speedup behaviour of the LM shows an almost ideal scaling (half the CPU-time for twice the number

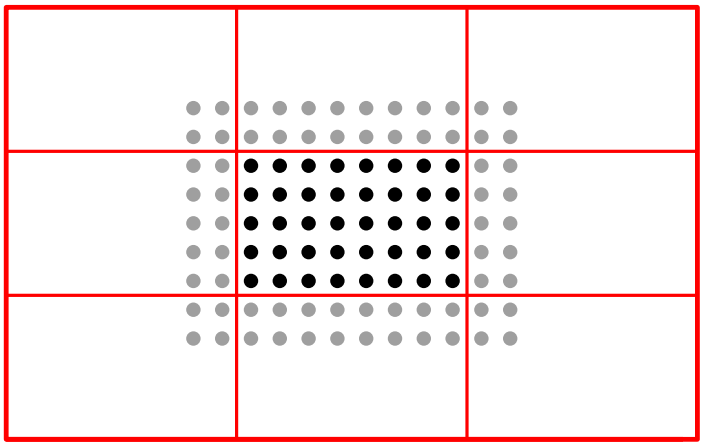

Figure 3: 2-D domain decomposition with a 2 gridline halo

of processors) up to 120 processors. With 240 processors, the execution time becomes larger than expected from ideal scaling. Here, the subdomain treated by a processor is so small that the ratio between physics-dynamics calculations – which scale superlinear – and the time for communication (data exchange and wait times due to load imbalances) and I/O – which both scale sub-linear – becomes disadvantageous.

Table 7: Timings for a 1-h LM forecast for various numbers of processors on an IMB-SP3

Number of Processors	15	30	60	120	240
Total CP-Time	571.53	284.04	136.34	67.61	38.10
Dynamics, Physics, Diagnostics	541.13	261.39	118.25	51.59	22.96
Communication	22.05	16.69	12.81	11.25	9.23
Input and Output	8.35	5.96	5.28	4.77	5.91

3.2 Data Assimilation

The requirements for the data assimilation system for the operational LM 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 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 driving model, and the main purpose of the assimilation scheme is to analyze the meso scales.

By design, 3-dimensional analysis methods 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 at significant computational costs. Moreover, the geostrophic approximation, a key ingredient of some of these schemes as used e.g. for the GME, is of limited validity on the meso scale. Therefore, 4-dimensional methods offer potential advantages since they include the model dynamics in the assimilation process directly. However, the 4-dimensional variational (4DVAR) method is too expensive for operational application of the LM considering the small amount of time available to produce the analyses and forecasts.

As a result, a scheme based on the observation nudging technique has been developed to

define the atmospheric fields. It is based on an experimental nudging analysis scheme which had been developed for DM and the Swiss model version SM (Schraff, 1996; 1997) and which compared favorably with the operational OI-analysis of the DM in various case studies. The new LM-scheme, however, has been adapted to the nonhydrostatic modelling framework and runs on distributed memory machines using domain decomposition. To compute the analysis increments locally for the grid points of each sub-domain, the observational information of the total domain is previously distributed to the sub-domains.

In addition to this, for some of the surface and soil fields, a set of 2-dimensional intermittent analysis schemes is applied. This comprises the snow analysis, the sea surface temperature (SST) analysis, and the variational soil moisture analysis scheme.

3.2.1 Nudging-Based Assimilation Scheme

Nudging or Newtonian relaxation consists of relaxing the model's prognostic variables towards prescribed values within a predetermined time window (see e.g. Davies and Turner (1977), Stauffer and Seaman (1990)). In the present scheme, nudging is performed towards direct observations which is more appropriate for 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 a prognostic variable $\psi(\mathbf{x}, t)$ is given by

$$\frac{\partial}{\partial t}\psi(\mathbf{x},t) \quad = \quad F(\boldsymbol{\psi},\mathbf{x},t) \quad + \quad G_\psi \cdot \sum_{k_{(obs)}} W_k \cdot [\psi_k - \psi(\mathbf{x}_k,t)] \tag{3}$$

F denotes the model dynamics and physical parameterizations, ψ_k the value of the k^{th} observation influencing the grid point $\mathbf{x}$ at time t, $\mathbf{x}_k$ the observation location, G_ψ the constant so-called nudging coefficient and W_k an observation-dependent weight which usually varies between 0 and 1. Neglecting the dynamics and physics and assuming a single observation with a constant weight W_k equal 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.

In practical applications, the nudging term usually remains smaller than the largest term of the dynamics so that the dynamic balance of the model is not strongly disturbed. The coupling between the mass and wind field innovations is primarily induced 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, and an initialization step is not required.

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, or the temporal (w_t) difference between the observation and the target grid point. If an increasing number of observations influence the grid point the total nudging weight should be limited to avoid the nudging term to become dominant over the dynamics. This is achieved by complementing the individual weight w_k by a relative weight (Benjamin and Seaman, 1985):

$$W_k \quad = \quad \frac{w_k}{\sum_j w_j} \cdot w_k \tag{4}$$

$$w_k \quad = \quad w_t \cdot w_{xy} \cdot w_z \cdot \epsilon_k \tag{5}$$

Currently, only conventional observations are used operationaly, namely from TEMP and PILOT (temperature and wind, including the significant levels; humidity up to 300 hPa;

geopotential only to derive one pressure increment at the lowest model level), AIRCRAFT (all data), WIND PROFILER and SYNOP, SHIP and DRIBU reports (station pressure; wind for stations below 100 m above msl; humidity; 2-m temperature is used only for the soil moisture analysis). Note that given a cut-off time of 2.5 hours, observations from up to about 2 hours after the actual analysis time can still be assimilated in the first hours of the operational forecast runs. As a quality control, the observed values are compared with the model fields of the assimilating run itself. For multi-level temperature data, a hydrostatic height and thickness check is included, and a spatial consistency check is performed for the station pressure data.

Equation (3) indicates that in principle the scheme consists of two main steps, i.e. the determination of the observation increments and the computation of the weights. With respect to the vertical interpolation required for the first step, the vertical scale of multi-level temperature and wind observations is adjusted to the vertical model resolution by averaging the observed profile over the thickness of model layers. As a result, the simulated thickness between two pressure levels is automatically relaxed towards the observed thickness when nudging temperature data. In contrast, humidity data are interpolated without averaging in order to capture thin layers of clouds as well as possible. Note that the increments are determined as differences in relative humidity which implies that relative rather than specific humidity is relaxed towards the observed humidity. In this sense, the analyzed quantities are horizontal wind, potential temperature, relative humidity, and pressure at the lowest model level.

Related to the second step, incomplete profiles and single-level increments are vertically extended and provided with vertical weights w_z according to a Gaussian (approx.) in log pressure (correlation scale is $1/\sqrt{3}$ for upper-air wind and 0.2 for upper-air temperature and humidity, and the cut-off is 850 m for surface-level wind resp. the lowest model layer for surface-level humidity). Thereafter, upper-air increments are spread laterally along horizontal surfaces since spreading along the terrain-following model levels as usually applied in nudging-type schemes has disadvantages near steep orography particularly in cases with low stratus (Schraff, 1997). In contrast, surface-level increments are spread along the model levels to limit the influence to the area close to the ground. The spreading includes the computation of the horizontal weights w_{xy} using the function $(1 + \Delta r/s) \cdot e^{-\Delta r/s}$ for the scalar quantities (Δr being the horizontal distance between observation and target grid point). The wind correlations are split into a longitudinal and transverse part, and this allows to specify the degree of divergence (γ) of the resulting wind analysis increment field (Lorenc et al., 1991). Both the correlation scales s and the non-divergence factor γ increase with height and with distance to the observation time and vary between about 60 km and 160 km resp. 0.4 and 0.7 . The function used for the temporal weights w_t is 1 at the observation time and decreases linearly to zero at 3 hours (for radiosonde data) resp. 1.5 hours (for other data) before and 1 resp. 0.5 hours after the observation time. Hourly or more frequent data are linearly interpolated in time.

In the current scheme, the resulting analysis increment fields are partly balanced explicitly in a third major step before being added to the model fields. Three types of balancing are applied. First, a hydrostatic upper-air temperature correction balances the pressure analysis increments at the lowest model layer. It is nearly constant within the lowest 1500 m (therefore hardly modifies the stability within the boundary layer) and decreases rapidly further above such that the geopotential above 400 hPa is not directly modified by the surface pressure nudging (for hydrostatic conditions). This significantly reduces the vertical extent of the mass field disturbance imposed by the pressure nudging and results in a better adjustment of the wind field and a greatly improved assimilation of the pressure data. Secondly, a geostrophic wind correction partly balances the wind field with respect to the mass field

increments imposed by the surface pressure nudging including the temperature correction. Finally, an upper-air pressure correction balances the total analysis increments of the mass field hydrostatically. This is the only feature directly related to the fact that the model is non-hydrostatic. Note that it does not change the non-hydrostatic properties of the full model fields. The correction prevents the introduction of direct sources of vertical wind for which there is no direct control without vertical wind observations being available to be assimilated. This is important since the vertical velocity is still small on the scales to be analyzed with the current scheme and observations (in contrast e.g. to a latent heat nudging scheme). Table 8 summarizes the main features of the LM nudging scheme.

Table 8: Data assimilation for LM

Method	Nudging towards observations
Implementation	continuous cycle of 3-hour assimilation runs
Realization	identical analysis increments used during 6 advection time steps
Balance	1. hydrostatic temperature increments (up to 400 hPa) balancing 'near-surface' pressure analysis increments 2. geostrophic wind increments balancing 'near-surface' pressure analysis increments 3. upper-air pressure increments balancing total analysis increments hydrostatically
Nudging coefficient	$6 \cdot 10^{-4} s^{-1}$ for all analyzed variables except pressure $1.2 \cdot 10^{-3} s^{-1}$ for 'near-surface' pressure
Analyzed variables	horizontal wind vector, potential temperature, relative humidity 'near-surface' pressure (i.e. at the lowest model level)
Spatial analysis	Data are analyzed vertically first, and then spread laterally along horizontal surfaces. vertical weighting: approximately Gaussian in log(p) horizontal weighting: isotropic as function of distance
Temporal weighting	1.0 at observation time, decreasing linearly to 0.0 at 3 hours (upper air) resp. 1.5 hours (surface-level data) before and 1.0 resp. 0.5 hours after observation time; linear temporal interpolation of frequent data.
Observations	SYNOP, SHIP, DRIBU: - station pressure, wind (stations below 100 m above msl) - humidity TEMP, PILOT: - wind, temperature: all standard levels, significant levels up to 150 hPa - humidity: all levels up to 300 hPa - geopotential used for one 'near-surface' pressure increment AIRCRAFT: - all wind and temperature data WIND PROFILER: - all wind data (except from blacklisted stations)
Quality control	Comparison with the model fields from assimilation run itself

3.2.2 Sea Surface Temperature Analysis

Since the latent and sensible heat fluxes over water depend crucially on the surface temperature, a sea surface temperature (SST) analysis is performed once per day (00 UTC). The global SST analysis for GME is deployed as first guess, which incorporates satellite data indirectly by making use of a global SST analysis from NCEP. All the ship and buoy observations from the previous 4 days are then used in a correction scheme based on Cressman-type weighting. For the sea-ice cover in the Baltic Sea, an external analysis (from the Bundesamt für Seeschifffahrt und Hydrographie) is used.

3.2.3 Snow Depth Analysis

The occurrence of a snow cover strongly influences the radiative absorption and reflection properties of the land surface and therefore the screen-level temperature. The snow water content is a prognostic quantity of the model, and is analyzed once every 6 hours. The method is based on a simple weighted averaging of SYNOP snow depth observations. The weighting depends both on the horizontal and vertical distances to the target grid points. In areas, where the density of these data is not sufficient, an average of snow depth increments derived from SYNOP precipitation, temperature, and weather reports as well as the model prediction are also included.

3.2.4 Soil Moisture Analysis

In land areas without snow, the soil water content influences significantly the screen-level temperature (and humidity) on clear-sky days. An inadequate specification of soil moisture can lead to forecast temperature errors of several degrees. The variational analysis scheme (Hess, 2001) derives improved moisture contents once per day by minimizing a cost functional J which depends on the deviations of the forecast temperature $T(\eta)$ from the observed (resp. analyzed) temperature T^o and of the soil moisture η from a given background state η^b:

$$J(\eta) = \frac{1}{2}\Big(T^o - T(\eta)\Big)^T \mathbf{R}^{-1}\Big(T^o - T(\eta)\Big) + \frac{1}{2}\Big(\eta - \eta^b\Big)^T \mathbf{B}^{-1}\Big(\eta - \eta^b\Big) \qquad (6)$$

The observation error covariance $\mathbf{R}$ and background error covariance $\mathbf{B}$ reflect the trust in the observations resp. the background. To solve the minimization problem, two assumptions are made. Firstly, since the 2-m temperature mainly depends on the soil moisture at the same location, the problem can be decoupled horizontally, and a low-dimensional (equal to the number of analyzed soil layers) minimization can be performed for each grid point individually. Secondly, (moderate) changes of soil moisture are assumed to lead to linear changes in temperature. This allows to derive the linear relationships Γ by means of one additional forecast run per analyzed soil layer where each of these forecasts has slightly different initial soil moisture values for the respective layer. The minimum of J can then be found by solving $\nabla J(\eta) = 0$ directly without using the adjoint method.

In the current implementation, two additional 15-hour forecasts are required to analyze two (sets of) soil layers for 00 UTC of the previous day by comparing forecast and observed temperature at 12 and 15 UTC. The analysis increments are then added to the soil moisture of the 24 h free forecast valid for 00 UTC of the current day. The resulting soil moisture is used both as initial state for the operational LM forecast of the current day and as background state for the next soil moisture analysis. This background state η^b is important in order to reduce the daily variation of the soil moisture contents and to stabilize the minimization in

cases of weak soil-atmosphere coupling (i.e. cloudy situations). Together with η^b (see above), the background error covariance $\mathbf{B}$ for the following day is provided in a Kalman-filter cycled analysis:

$$(\mathbf{B})^{next} \quad = \quad \mathbf{A} \; + \; \mathbf{Q} \quad , \quad where \;\; \mathbf{A} = (\nabla^2 J)^{-1} = \left(\Gamma^T \mathbf{R}^{-1} \Gamma + \mathbf{B}^{-1} \right)^{-1} \tag{7}$$

This takes into account both an increase of confidence in the retrieved soil moisture values due to the utilized screen-level observations (as part of the analysis error covariance $\mathbf{A}$) and a decrease of confidence due to the model error $\mathbf{Q}$ of the soil model. While $\mathbf{A}$ can be computed explicitly, $\mathbf{Q}$ is the main tuning parameter of the scheme. It influences the relative weight given to the past and the present observations and has an impact on the temporal variability of the soil moisture. The scheme has been successfully tested in various case studies and it is operated at DWD since March 2000.

3.3 Boundary Conditions from Driving Models

The LM can be nested in the global model GME of DWD (Majewski, 1998; Majewski et al., 2002), the ECMWF global spectral model IFS and also in itself. The lateral boundaries are treated by the Davies (1976) relaxation technique, where the internal model solution is nudged against an externally specified solution within a narrow boundary zone by adding a relaxation forcing term to the equations.

The external solution is obtained by interpolation from the driving host model at discrete time intervals. The interpolated fields are hydrostatically balanced, i.e. a hydrostatic pressure is prescribed for the nonhydrostatic pressure variable in LM at the lateral boundaries. Within these specified time intervals, the boundary data are interpolated linearly in time (which is done inside the model). Normally the boundary update interval is chosen to be one hour for meso-β scale applications of the LM. The boundary values (and initial values, if no data assimilation suite is operated) are obtained by a preprocessing program from the host model.

The different interpolation tools (GME2LM, IFS2LM, LM2LM) have now been combined to a single program INT2LM, which takes the data from GME, IFS or LM as input and interpolates to a specified LM grid.

A User Guide of the INT2LM preprocessor program is available at the COSMO Web-site.

3.4 Postprocessing

Postprocessing includes all applications that use the direct model output of LM runs. In general, there is a wide range of such applications at each meteorological service, ranging from simple graphical display of weather charts or meteograms for single grid points, or statistical correction of near surface weather elements by Kalman filtering, to more complex derived products supplying information on environment and health, transportation, agriculture and media presentation. Most of these postprocessing tools are very specific to the computer platform, data base system and visualization software of each service and thus cannot be shared within the COSMO group. There is, however, a number of postprocessing programs available within COSMO.

(a) Graphics

Work on two common plotting packages has been completed. The first has been developed at MeteoSwiss and uses Metview with an interface to the GRIB1 LM output data; the other one has been developed at ARPA-SIM and is based on the public domain VIS5D packages; a special routine converts the GRIB1 binary format to the VIS5D data format.

(b) Models

A *Lagrangian Particle Dispersion Model* (LPDM) may be used operationally in case of radioactive accidental releases to predict long-range transport, dispersion, and wet and dry deposition of radioactive material. The calculation of about $10^5 - 10^6$ trajectories of tracer particles is based on wind fields from LM (at hourly intervals) and superimposed turbulent fluctuations (TKE, Monte Carlo method). Radioactive decay and convective mixing are included. The concentration is calculated by counting the particle masses in arbitrary grids.

A *Trajectory Model* may provide guidance on transport routes. The meteorological input is derived from LM at hourly intervals.

An integral part of the NWP system at DWD is a *Wave Prediction Suite* comprising two models, namely the global model GSM (global sea state model), and a local one (LSM) which covers the Baltic Sea, the North Sea and the Adriatic Sea with a high-resolution mesh. GSM and LSM have been developed by the research institute GKSS in Geesthacht (Germany).

(c) Interpretation

An objective weather interpretation scheme (developed at DWD) derives the forecasted 'weather', i.e. the WMO weather code, based on LM output fields. Pressure, temperature, dew point temperature, liquid water content, cloud cover, precipitation and wind speed values are used as input parameters to define the present weather.

3.5 Data Flow of the LM Package

The various components of the LM package and the corresponding data flow are illustrated in Figure 4. In case of a set-up without data assimilation (right part of Figure 4), the interpolation program INT2LM provides initial and boundary conditions for the LM forecast runs (LM-FCT) from the corresponding driving models. This step involves the data set of the external parameters (see Section 3.1.4).

With a system set-up using the LM nudging analysis (left part of Figure 4), the INT2LM provides boundary conditions (LM-BC) from the GME assimilation cycle for the LM runs in nudging analysis mode (LM-NUD) within the assimilation stream. The LM-NUD runs start from a given LM analysis (LM-ANA) to generate an analysis for the next analysis time. The forecasts then start from these LM-ANA initial data using boundary conditions from the GME forecast.

To run the LM in nudging mode, a preprocessor program is required which provides the observational data in a special data file format (AOF). The LM analysis file may be modified by incremental analyses of sea surface temperature, snow depth and soil moisture (see Section 3.2). All these programs use GTS and non-GTS observational data, which are archived in a local data base system. The interface to these data is usually not portable as it depends on the data base system of each meteorological centre.

The LM runs in forecast mode generate direct model output, which includes also fields from the LM internal postprocessing (see Appendix B). These data are then subject to various visualization tools, external postprocessing and other applications such as follow-up models at COSMO Met Services.

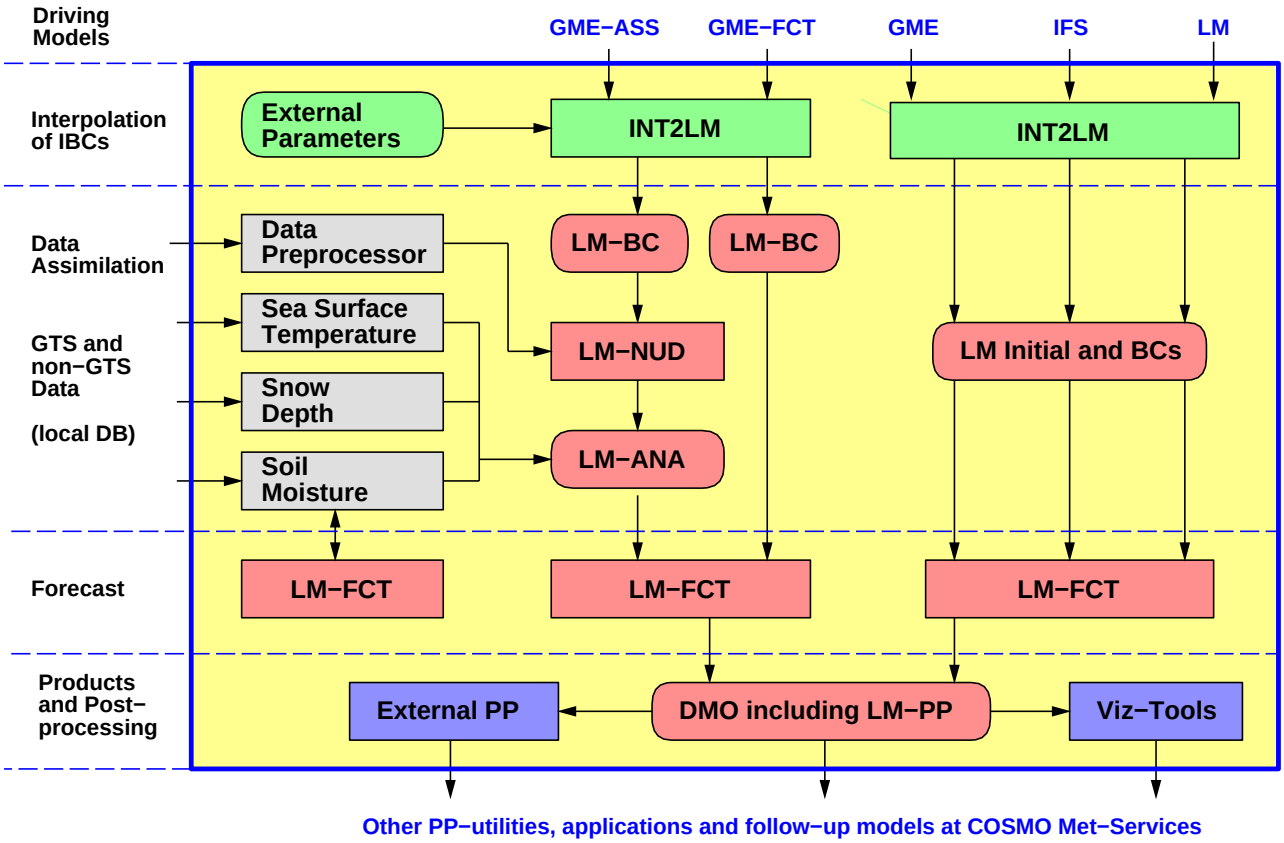

Figure 4: Process and data flowchart of the LM package for a set-up using data assimilation (left part) and a set-up without data assimilation (right part). Rectangular boxes indicate components of the package (programs), rounded boxes indicate data files generated by the components.

3.6 Documentation

The new release of the LM documentation covers the following parts:

A Description of the Nonhydrostatic Regional Model LM

Part I: *Dynamics and Numerics*
Part II: *Physical Parameterization*
Part III: *Data Assimilation*
Part IV: *Implementation Documentation*
Part V: *Preprocessing*
Part VI: *Postprocessing*
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 including soil moisture analysis. The scientific documentation is independent of (i.e. does not refer to) the code itself.

Part IV describes 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.

The generation of initial and boundary conditions from coarse grid driving models is described in Part V. Available internal and external postprocessing utilities are described in Part VI.

Finally, the User's Guide (Part VII) 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.

Available at the COSMO web-site *www.cosmo-model.org* during 2004 are Parts I, II, III, V and VII. Part IV and VI will follow as soon as possible.