4. Four Dimensional Data Assimilation (FDDA)

The concept of combining current and past data in an explicit dynamical model such that the model's prognostic equations provide time continuity and dynamic coupling among the various fields has become known as four-dimensional data assimilation (FDDA). Current interest in the use of FDDA in mesoscale models, for either model initialization (dynamic initialization) or for use of the model as an analysis/research tool (dynamic analysis), is a logical extension of the traditional link between objective analysis methods and dynamic relationships.

Currently, two major types of FDDA are used operationally and in research. The first is an intermittent process of initializing an explicit prediction model, using the subsequent forecast (typically 3-12 h) as a first guess in a static three-dimensional objective analysis step, and then repeating the process for another forecast cycle. The second is a continuous (every model time step) dynamical assimilation where forcing functions are added to the governing model equations to gradually "nudge" the model state toward the observations. This continuous nudging type of FDDA is used in the PSU/NCAR modeling system. Nudging was first developed and tested at Penn State by Kistler (1974), under Prof. J. Hovermale, and by Anthes (1974), and Hoke and Anthes (1976). See Stauffer and Seaman (1990) for an historical overview of the technique.

Nudging or Newtonian relaxation is a relatively simple but very flexible technique: the data used for nudging can be of any type, measured or derived, analyzed to a grid for assimilation into the model or inserted as individual observations. Gridded analyses of the observations that are assimilated can be obtained by successive correction, variational, or statistical optimal interpolation (OI) techniques, and the weights used when nudging to individual observations can be simple Cressman-type (distance-weighted) functions or more complicated statistical functions based on OI. It can be shown that successive corrections, OI, and variational approaches are all fundamentally related to the "idealized analysis" and thus to each other. In fact, nudging is basically a successive-correction technique which uses a numerical model to include the time dimension.

The method of Newtonian relaxation or nudging relaxes the model state toward the observed state by adding, to one or more of the prognostic equations, artificial tendency terms based on the difference between the two states. The model solution can be nudged toward either gridded analyses or individual observations during a period of time surrounding the observations. These two techniques, hereafter referred to as "analysis nudging" and "obs nudging", respectively, can be used individually or simultaneously on any MM5 model grid. For guidance in selecting which nudging technique(s) to use for your particular application, as well as suggested parameter specifications, see Stauffer and Seaman (1990), Stauffer et al. (1991) and Stauffer and Seaman (1993).

4.1 Analysis Nudging

The analysis-nudging term for a given variable is proportional to the difference between the model simulation and an analysis of observations calculated at every grid point. The general form for the predictive equation of variable $\alpha(\mathbf{x},t)$ is written in flux form as

$$\frac{\partial p^*\alpha}{\partial t} = F(\alpha,\mathbf{x},t) + G_\alpha \cdot W_\alpha(\mathbf{x},t)\cdot \epsilon_\alpha(\mathbf{x})\cdot p^*(\hat{\alpha}_0-\alpha) + G_{p^*}\cdot W_{p^*}\cdot \epsilon_{p^*}(\mathbf{x})\cdot \alpha(\hat{p}_0^*-p^*) \quad (4.1.1)$$

All of the model's physical forcing terms (advection, Coriolis effects, etc.) are represented by F, where α are the model's dependent variables, $\mathbf{x}$ are the independent spatial variables, and t is time. The second and third terms on the right of (4.1.1) are similar in form and represent the nudging terms for α and p*, respectively. Due to the flux form of the predictive equation, the third term appears in (4.1.1) when nudging pressure in the continuity equation of the hydrostatic version of MM5. (Note that this term is zero in the nonhydrostatic version of MM5 because p^* is computed from the hydrostatic reference state and is constant in time.)

With $G_{p^*}=0$, or in the nonhydrostatic version of MM5, (4.1.1) simplifies to

$$\frac{\partial p^*\alpha}{\partial t} = F(\alpha,\mathbf{x},t) + G_\alpha \cdot W_\alpha \cdot \epsilon_\alpha(\mathbf{x})\cdot p^*(\hat{\alpha}_0-\alpha) \quad (4.1.2)$$

The nudging factor G_α determines the magnitude of the term relative to all the other model processes in F. Its spatial and temporal variation is determined mostly by the four-dimensional weighting function, W, which specifies the horizontal, vertical and time weighting applied to the analysis, where $W = w_{xy}w_\sigma w_t$. The analysis quality factor, ϵ, which ranges between 0 and 1, is based on the quality and distribution of the data used to produce the gridded analysis. The estimate of the observation for α analyzed to the grid is $\hat{\alpha}_0$.

The nudging factor G_α is defined based on scaling arguments. Because the nudging contribution is artificial, it must not be a dominant term in the governing equations and should be scaled by the slowest physical adjustment process in the model (inertial effects). Thus the G_α is usually defined to be similar in magnitude to the Coriolis parameter, and it must also satisfy the numerical stability criterion, $G_\alpha \leq \frac{1}{\Delta t}$. Typical values of G_α are $10^{-4}s^{-1}$ to $10^{-3}s^{-1}$ for meteorological systems, where values of $G_\alpha = 3 \times 10^{-4}s^{-1}$ to $6 \times 10^{-4}s^{-1}$ are usually "large enough". A value of G_α which is too large will force the model state too strongly toward the observations. This is undesirable because (a) the ability of the model equations to resolve mass-momentum imbalances will be decreased; and (b) the ability of the model to generate its own mesoscale meteorological structures (e.g. fronts, squall lines) will be impaired by heavy insertion of the observed analyses. Such problems arise because the analyses may not resolve these mesoscale structures or may be contaminated by observational and analysis errors. On the other hand, if G_α is too small, the observations will have minimal effect on the evolution of the model state, allowing phase and amplitude errors to grow.

For simplicity, if we drop the physical forcing terms F from (4.1.2), and assume $W(\mathbf{x}, t) = 1$, $\frac{\partial p^*}{\partial t} = 0$ and the observational analysis is perfect and time invariant, then

$$\frac{\partial \alpha}{\partial t} = G_\alpha(\hat{\alpha}_0 - \alpha) \tag{4.1.3}$$

which has the solution

$$\alpha = \hat{\alpha}_0 + (\alpha_i - \hat{\alpha}_0)e^{-G_\alpha t} \tag{4.1.4}$$

where α_i is the initial value of α at the start of the nudging period. Therefore, the model state approaches the observed state exponentially with an e-folding time of $T_G = \frac{1}{G_\alpha}$, which is about 0.93 h for $G_\alpha = 3 \times 10^{-4}s^{-1}$. This implies that very high frequency fluctuations in the data, as might be available from wind profilers or Doppler radars (say, every 5 min), would not be retained well unless G_α were much greater; but then the nudging term may not be small compared to some terms of F.

Nudging the vorticity is a alternative method whereby the model's divergent wind is allowed to freely respond in the model's geostrophic adjustment process. Equation 4.1.3 can be modified for u and v such that

$$\frac{\partial u}{\partial t} = G_u \cdot \frac{\partial}{\partial y}(\hat{\zeta}_0 - \zeta) \tag{4.1.5}$$

$$\frac{\partial v}{\partial t} = G_v \cdot \frac{\partial}{\partial x}(\hat{\zeta}_0 - \zeta), \tag{4.1.6}$$

where ζ is the model vorticity and $\hat{\zeta}_0$ is the analyzed observed vorticity. Letting a constant $G = G_u = G_v$ and forming the vorticity equation from (4.1.5) and (4.1.6) we get

$$\frac{\partial \zeta}{\partial t} = G\nabla^2(\hat{\zeta}_0 - \zeta) \tag{4.1.7}$$

Therefore, the model vorticity is diffused toward the observed vorticity. However, the Laplacian in (4.1.7) introduces a scale dependence when nudging vorticity. The model vorticity will be diffused more strongly to small-scale features in the observed vorticity analysis. Thus, when nudging toward large-scale vorticity fields, small-scale features in the model vorticity will be selectively damped. After all factors are considered, it is generally advisable to nudge the u- and v-components of the winds directly whenever possible (see Stauffer and Seaman, 1990).

Although this analysis-nudging technique has been traditionally used to assimilate 3-D analyses based on rawinsonde observations, it can also be used to assimilate 2-D surface analyses within the model PBL (Stauffer et al., 1991).

While the twice daily rawinsonde observations are spaced at about 400 km and number approximately 100 over the U.S., the spatial distribution of the surface data, available at 3-h intervals from the NCAR data archives, is considerably more dense. Depending on the time of day, there are roughly 500- 1200 surface data sites with an average spacing of about 100 km. The greater horizontal and temporal resolution of conventional surface data is especially attractive for mesoscale data assimilation. The surface-level analyses are assimilated in the same manner as the 3-D analyses, except for the vertical extent of their influence. These surface analyses, which are assumed to be representative at 10 m above ground level (AGL), should represent only those scales resolved by the model grid onto which they are assimilated. This is necessary to avoid difficulties related to the insertion of small-scale divergence patterns (near $2\Delta x$) which might interact adversely with the model's parameterizations (e.g., the moisture-convergence criterion used to determine the existence and intensity of Anthes-Kuo convection).

Effective assimilation of single-level data depends on the equivalent depth over which the data are inserted into the model. Beneficial effects on numerical forecasts can be achieved by distributing single-level data throughout several model layers. This approach requires that the data be applied in a consistent manner such that they are assured to be representative of those layers. For example, the homogenizing effect of vertical mixing during free convective conditions allows us to assume that surface-layer wind and mixing ratio (q) observations can be applied throughout the model PBL according to a conceptual model of boundary-layer structure. Such an idealized conceptual model is given by Garratt et al. (1982) and is based on typical moderate to large instabilities observed at Wangara and Minnesota (Fig. 4.1). However, in this same situation the frequent presence of a shallow superadiabatic layer near the surface makes surface temperature or potential temperature data poorly representative of the boundary layer as a whole, and similarity relationships describing the potential temperature profile become inaccurate. The same is true during nocturnal inversion conditions. These and other factors make assimilation of single-level surface temperature observations unattractive for defining the temperature of the mixed layer above the surface layer. For example, nudging towards an inaccurately diagnosed mixed-layer temperature can cause serious errors in the simulated PBL depth or even lead to a sudden spurious collapse of the unstable PBL structure. This can result from assimilating a surface temperature observation which is cooler than the model-simulated value by only a few tenths of one degree. In general, surface temperature data should not be directly assimilated into the model (see Stauffer et al., 1991).

Figure 4.1 shows the unstable lower troposphere comprised of three distinct layers: a surface layer extending to height h_s, a well-mixed layer from h_s to h_1 and a transition layer extending from h_1 to h_2. With the x-axis defined parallel to the mean wind, Fig. 4.1 suggests that the v component of the wind is zero and there is thus no directional shear through the lower two layers. The surface wind speed analysis, discussed above and

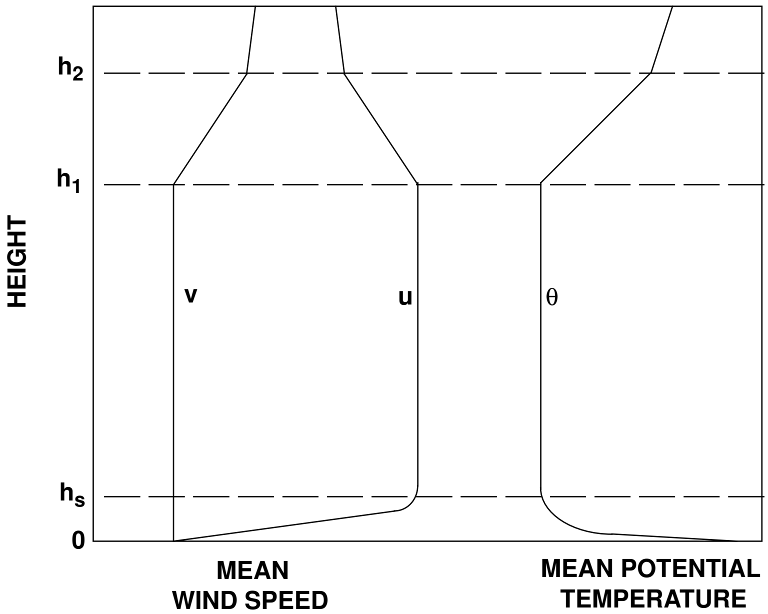

Figure 4.1 Schematic representation of mean wind speed and potential temperature profiles in an idealized conceptual model of the unstable atmospheric boundary layer.

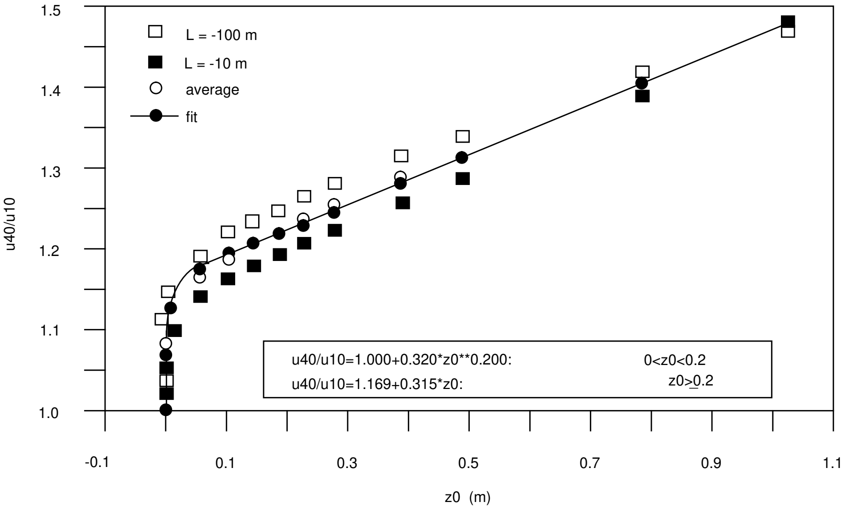

Figure 4.2 Schematic representing the relationship used to adapt the 10-m wind to the 40-m level as a function of roughness length (z_0) and Monin length (L).

assumed to apply at 10 m AGL, can be modified via similarity theory to apply at h_s. This modified wind for h_s is representative of the mean mixed layer up to h_1. These assumptions allow the surface-layer wind information to be applied throughout the model's multilayer PBL rather than at a single level. Similarly, the surface-layer mixing ratio (not shown) is also assumed to be representative of the entire mixed layer, and may be applied over several model layers during free-convective conditions.

The surface-layer wind analyses, however, must be adapted for the depth of the surface layer, h_s, which is assumed to be the height of the lowest model layer under unstable conditions (40 m AGL for $\sigma = 0.995$). Similarity relationships that assume a logarithmic profile of wind with height are used to adjust the 10-m surface wind analysis to the lowest model-layer height. Figure 4.2 shows that this wind adjustment is more strongly dependent on roughness (z_0) than on stability as measured by the Monin length (L); therefore, a "best-fit" relationship is determined for average stability conditions for roughness lengths ranging from 0.0 to 1.0 m. Thus, during free-convective conditions, the mean wind assimilated throughout the model PBL is based on the surface wind analysis modified to account for the model surface-layer height and the grid-box roughness. The surface analysis of mixing ratio is assumed to be representative for the lowest model layer and throughout the mixed layer.

During unstable conditions, the Blackadar scheme may develop a PBL which extends through several model layers. The "nudging correction" to the wind field at the lowest model layer is applied throughout the model PBL to simulate the conceptual model. Because this idealization is also closely reproduced by the PBL scheme (without nudging), the nudging strategy within the PBL is compatible with the boundary-layer physics. The observed surface mixing ratio applied throughout the model mixed layer is adjusted if necessary to remove any supersaturation with respect to the current model-simulated temperature and moisture conditions for a given level and grid point. This prevents the moisture assimilation term from becoming an artificial source of precipitation. During stable conditions, on the other hand, the height of the PBL is defined to be that of the lowest model layer since the boundary layer is largely decoupled from the free troposphere above.

Therefore, the 3-hourly surface-analysis nudging is also given by (4.1.2), but the vertical extent of the nudging is controlled by the model- simulated PBL depth, with $\hat{\alpha}_0$ for wind and moisture adjusted as previously discussed above. The analysis confidence factor, ϵ, for the 3-h surface analyses, is functionally dependent on the spatial distribution of the surface observations used to produce the analysis. Over land it varies from unity at grid boxes within one-half the prescribed radius of influence of a surface observation to 0.2 for grid boxes outside the prescribed radius.

The vertical weighting factor, w_σ, is defined as

$$w_\sigma = w_\sigma^R + w_\sigma^S \leq 1. \tag{4.1.8}$$

where w_σ^R and w_σ^S represent w_σ for assimilation of 3-D rawinsonde and 2-D surface data, respectively, and w_σ^S depends on the model-simulated PBL depth. The surface data are assimilated with full strength ($w_\sigma^S = 1.0$) within the layers defining the PBL and with reduced strength ($w_\sigma^S = 0.9$) one layer above (in the transition layer). The vertical weighting function used to assimilate 3-D rawinsonde data is defined such that $w_\sigma^R = 0.0$ in the PBL, 0.1 in the transition layer and 1.0 aloft. During stable conditions, therefore, the surface data are applied with full strength only in the lowest model layer and with reduced strength one layer above. Both types of analysis nudging generally assimilate temporally interpolated gridded analyses; that is, $\hat{\alpha}_0$ in (4.1.2) is interpolated in time, for example, from either 12-h 3-D analyses or 3-h 2-D surface analyses. Therefore, w_t is usually set to unity, except when decreasing the nudging at the end of a dynamic-initialization period.

4.2 Observational Nudging

This alternative scheme does not require gridded analyses of observations throughout the case study period, and may be better suited for situations with high-frequency asynoptic data (e.g., profilers), especially on the subalpha scales. Its form is similar to (4.1.2) and it uses only those observations which fall within a predetermined time window that is centered about each model time step. The set of differences between the model and the observed state is computed at the observation locations, and analyzed back to the grid in a region surrounding the observations. The tendency for $\alpha(\mathbf{x}, t)$ with $G_{p^*} = 0$ is given by

$$\frac{\partial p^* \alpha}{\partial t} = F(\alpha, \mathbf{x}, t) + G_\alpha \cdot p^* \frac{\sum_{i=1}^{N} W_i^2(\mathbf{x}, t) \cdot \gamma_i \cdot (\alpha_o - \hat{\alpha})_i}{\sum_{i=1}^{N} W_i(\mathbf{x}, t)}, \tag{4.2.1}$$

where F and G_α are as defined earlier, subscript i denotes the i th observation of a total of N which are within a preset radius of a given grid point, α_o is the locally observed value of α, and $\hat{\alpha}$ is the model's prognostic variable interpolated to the observation location in three dimensions. The observational quality factor,γ , which ranges from 0 to 1, accounts for characteristic errors in measurement systems and representativeness. The four-dimensional weighting function accounts for the spatial and temporal separation of the ith observation from a given grid point at a given time step.

The four-dimensional weighting function for each observation i in (4.2.1) is rewritten as

$$W(\mathbf{x}, t) = w_{xy} \cdot w_\sigma \cdot w_t \tag{4.2.2}$$

The horizontal weighting function, w_{xy}, is a Cressman-type spatial weighting function defined by

$$w_{xy} = \frac{R^2 - D^2}{R^2 + D^2} \qquad 0 \le D \le R \tag{4.2.2}$$

and

$$w_{xy} = 0 \qquad D > R, \tag{4.2.3}$$

where R is the radius of influence and D is the distance from the i th observation location to the grid point. The vertical weighting function, w_σ is also a distance-weighted function defined by

$$w_\sigma = 1 - \frac{|\sigma_{obs} - \sigma|}{R_\sigma} \qquad |\sigma_{obs} - \sigma| \le R_\sigma \tag{4.2.4a}$$

$$w_\sigma = 0 \qquad |\sigma_{obs} - \sigma| > R_\sigma, \tag{4.2.4b}$$

where R_σ is the vertical radius of influence and σ_{obs} is the vertical position of the i th observation. The temporal weighting function is given by

$$w_t = 1 \qquad |t - t_0| < \tau/2 \tag{4.2.5}$$

$$w_t = \frac{\tau - |t - t_0|}{\tau/2} \qquad \tau/2 \le |t - t_0| \le \tau \tag{4.2.6}$$

$$w_t = 0 \qquad\qquad |t - t_0| > \tau, \tag{4.2.7}$$

where t is the model-relative time, t_0 is the model-relative time of the ith observation, and τ is the half-period of a predetermined time window over which an observation will influence the model simulation.

For economy, the multi-level observations (soundings) used for obs nudging are usually vertically interpolated to the model sigma surfaces at each observation location prior to each simulation. Although the vertical component of the weighting function, w_σ (4.2.4), is also a distance-weighted function, the vertical radius of influence, R_σ , can be defined to be small (less than the spacing of the model layers) so that each observation above the model surface layer influences only one sigma layer at a given location.

Figure 4.3 illustrates schematically the horizontal and temporal components of W used for nudging to observations. The horizontal weighting function, w_{xy}, is the Cressman function given by (4.2.2) and (4.2.3). As shown in the top of the figure, the horizontal radius of influence varies linearly in the vertical with pressure, from R_s at the surface to the preset value R' at a pressure level p' representing the free atmosphere, where terrain influences are assumed to be small. At pressures less than or equal to this user-defined value, defined by default as 500 mb, the horizontal radius of influence is defined by default as twice the value used in the surface layer, R_s. For example, if R_s = 100 km, R' = 200 km. For certain situations, such as with upward propagating mountain-induced gravity waves, the assumption of negligible terrain influence within the troposphere is invalid and should be avoided.

As shown in the top of Fig. 4.3, the corrections computed at a given observation site and vertical level above the surface layer (lowest model layer) are spread laterally along a constant pressure level and thus across several sigma layers in regions of sloping terrain. That is, for any given grid point within the horizontal radius of influence, the obs-nudging correction in the horizontal direction is applied to the sigma layer which has a pressure value closest to that of the observation.

Observations within the model surface layer are spread along constant sigma surfaces, but with a modified Cressman function (dashed contours in the middle of Fig. 4.3) which reduces the influence of an observation as a function of the surface pressure (terrain).

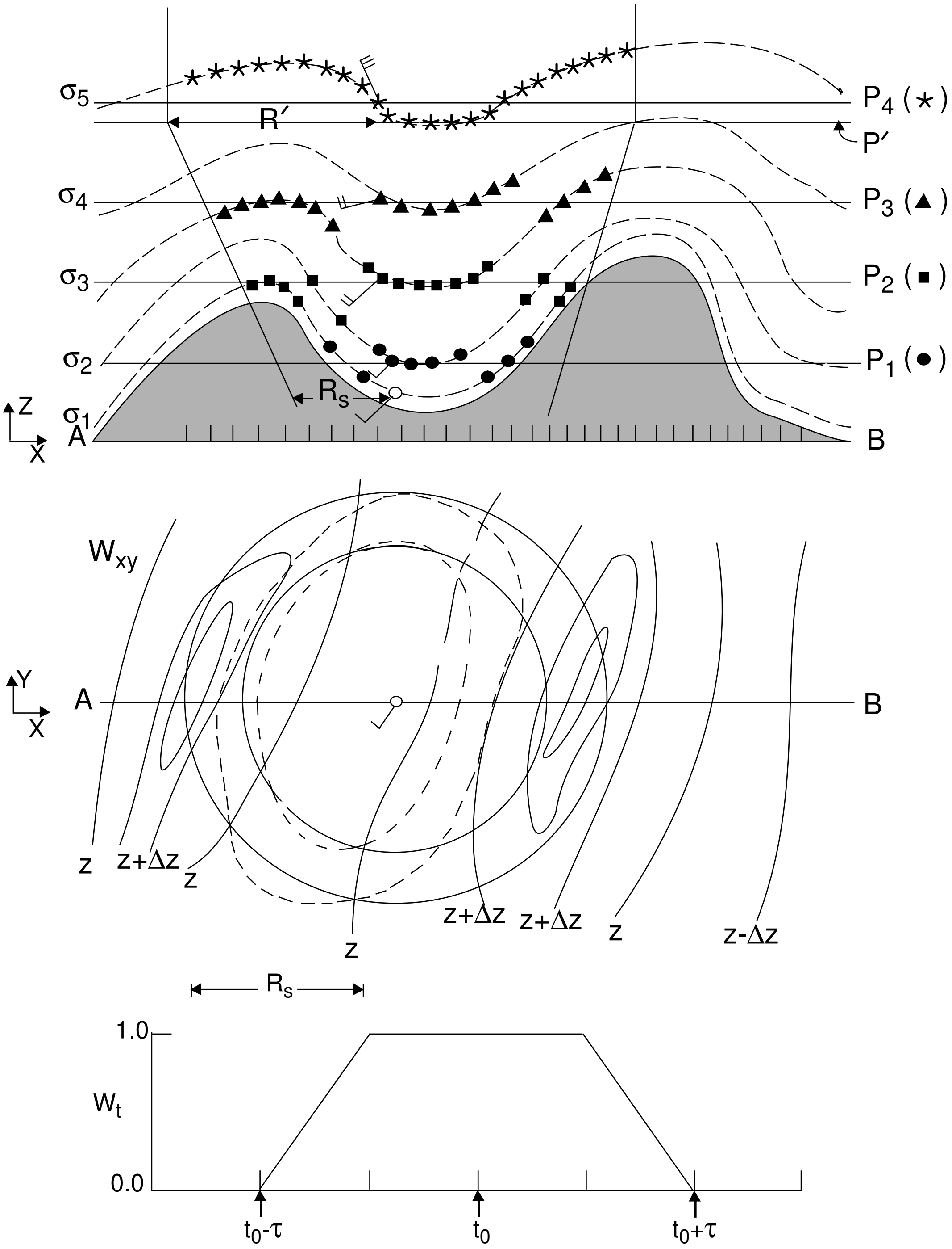

Figure 4.3: Schematic showing the horizontal weighting function, w_{xy}, and the temporal weighting function, w_t, used for obs nudging.

Thus spreading the influence of surface-layer observations along the lowest sigma ensures that the FDDA forcing near the surface in uneven terrain is continuous, and not like a pebble skipping across a pond. For observations in the surface layer, the distance factor D in (4.2.2) is replaced with D_m,

$$D_m = D + R_S C_m^{-1} |p_{s0} - p_s|, \tag{4.2.8}$$

where D is as defined above, C_m is a constant, and p_{s0} and p_s are the surface pressures at the observation location and the grid point, respectively. For example, C_m is typically defined as 75 mb, and R_s is the surface-layer value for the horizontal radius of influence. As the difference in surface pressure between the observation location and the grid point approaches C_m, the second term in (4.2.8) approaches R, and w_{xy} tends to zero faster for a given D. Therefore, the effect of assimilating surface-layer observations in the valley (mountains) on grid point locations in the mountains (valley) will be much reduced. This minimizes the possibility that observations in complex terrain will influence the model solution in areas where they may not be representative. Also, the vertical weighting factor, w_σ, for these surface-layer observations is defined so that the vertical influence of the surface-layer observations decreases linearly through the lowest 3 or so model layers (about 250 m AGL). As mentioned earlier, single-level data are retained better by the model if assimilated through several vertical layers.

The temporal weighting function, w_t (4.2.5-4.2.7), shown in the bottom of Fig. 4.3, is nonzero during a preset time window centered about the observation time, t_0. It determines the time period over which the ith observation can influence the model simulation via (4.2.1). In general, this time window can also be defined as a function of the pressure level of the observation similar to the effect of the horizontal radius of influence, R, in (4.2.2). Thus the final correction to the model solution via obs nudging reflects a weighted average of all observations during the preset time window about the current time step and within some three-dimensional neighborhood of each grid point.