Contract No:

This document was prepared in conjunction with work accomplished under Contract No. DE-AC09-08SR22470 with the U.S. Department of Energy.

Disclaimer:

The Application of an Evolutionary Algorithm to the Optimization of a Mesoscale Meteorological Model

Dr. Lance O'Steen
Savannah River National Laboratory
Aiken, SC 29801

and

Dr. David Werth
Savannah River National Laboratory
Aiken, SC 29801

February 15, 2008

Submitted to Journal of Applied Meteorology

Corresponding author: David Werth, Savannah River Site, Savannah River National Laboratory, Building 773A, Room. A-1012, Aiken, SC, david.werth@srnl.doe.gov, 803-725-3717

Abstract

We show that a simple evolutionary algorithm can optimize a set of mesoscale atmospheric model parameters with respect to agreement between the mesoscale simulation and a limited set of synthetic observations. This is illustrated using the Regional Atmospheric Modeling System (RAMS). A set of 23 RAMS parameters is optimized by minimizing a cost function based on the root mean square (rms) error between the RAMS simulation and synthetic data (observations derived from a separate RAMS simulation). We find that the optimization can be efficient with relatively modest computer resources, thus operational implementation is possible. The optimization efficiency, however, is found to depend strongly on the procedure used to perturb the 'child' parameters relative to their 'parents' within the evolutionary algorithm. In addition, the meteorological variables included in the rms error and their weighting are found to be an important factor with respect to finding the global optimum.

1. **Introduction**

Improvement in mesoscale atmospheric model simulations is often thought to be primarily a matter of finer spatial resolution. This is probably true, at least until all the surface features having a significant impact on the meteorological scales of interest have been resolved. At this point, further improvement in model performance must come from other aspects of the simulation. Even if a specific surface feature is well-resolved, its impact on the atmosphere must be adequately described and this typically involves a parameterization. Recent atmospheric simulations of the Salt Lake Valley by Zhong and Fast (2003) demonstrate that there is a limit to the improvement one can obtain by simply decreasing the grid size of a numerical model. They conclude that "relatively large forecast errors can still exist even with sufficiently fine spatial resolution", and that further improvements in forecasts will require better model parameterizations. While this might not be surprising, it does leave the mesoscale modeler with the task of determining which parameterizations to use for a specific problem and what values to use for individual model parameters.

Mesoscale atmospheric models typically have a large number of user-input and "hard-coded" parameters that exert considerable influence on the simulated fluxes of mass, momentum and energy. Depending on one's viewpoint, this is either a blessing or a curse. In either case, decisions have to be made based on previous experience, intuition or expert advice. The majority of model parameters control surface processes and thus the greatest impact of these parameters is on the fluxes of water, energy and momentum between the earth's surface and the surface-layer of the atmosphere. In addition, fluxes *within* the boundary layer are influenced by parameters that control turbulence and

radiative transfer in the atmosphere. The effect of unresolved topography on air flow within the boundary layer is also determined by model parameters. Even non-physics based parameters such as the size and location of grids and how the mesocsale model is forced by large-scale data (nudging) can be crucial. The accuracy of a given numerical simulation is often a matter of a judicious choice of many model parameters. In fact, as most modelers know all too well, generating a simulation that simply runs without a catastrophic failure can depend on the choice of model parameters.

Improvement in mesoscale simulations by parameter adjustment is often attempted in a subjective manner by observing the agreement with observations while changing a few of the user-input parameters. However, it is unlikely that the resulting simulation is actually optimized relative to the observations. A more systematic, objective and robust procedure is needed that can optimize an extremely complex model with a large number of adjustable parameters that might be highly correlated.

Evolutionary Programming (EP) methods provide such an optimization procedure. We use 'evolutionary' here instead of 'genetic' since only mutation operators are employed in the computational algorithm (genetic methods typically allow mating processes). These methods are iterative in nature, starting with an initial guess for the parameters to be adjusted, and perturbing them in a prescribed manner until agreement between the simulation and observations is maximized. The objective or cost function to be minimized in the EP algorithm is also user-defined. EP methods are well-suited for optimizing large numbers of parameters with extremely complex cost function surfaces (Schwefel, 1995). In addition, evolutionary algorithms are conceptually simple and highly parallel, which is extremely important for computationally-expensive problems

such as atmospheric simulation. Evolutionary methods have been successfully applied in such diverse areas as image processing, artificial intelligence, signal processing, biotechnology and game theory (Bäck et al, 1997). In atmospheric science, evolutionary methods are currently being explored for optimizing transport model parameters (source terms and meteorology) based on agreement with concentration observations (Haupt, 2005).

This paper represents an initial attempt to establish the efficacy of one such method for optimizing the Regional Atmospheric Modeling System (Pielke et al., 1992) (RAMS). As will be shown, model performance can be greatly improved if a sufficiently robust set of parameters are adjusted in a mutation/selection process that uses root mean square (rms) error between simulation and observations to guide the process.

2. **RAMS**

RAMS is a three-dimensional, finite difference model that solves the Reynolds-averaged primitive equations of geophysical fluid dynamics on an Arakawa C-grid using a leapfrog scheme in time. The RAMS model (Pielke, 1992) is primarily used to simulate weather on regional scales. A two-way nested grid system is used so that the inner grids can affect and be affected by the outer grids. The model uses a terrain-following vertical coordinate system with variable grid spacing in the vertical For the simulations here, we use 40 levels in the vertical that vary in spacing from 30m at the surface to 1000m at the top, reaching an upper limit at 15610m. The model boundary conditions are from the National Center for Environmental Prediction (NCEP) aviation (AVN) model, and are

updated at 6 hour increments. The RAMS grids are 'relaxed' towards the imposed boundary conditions using a Newtonian nudging scheme.

The Mellor/Yamada scheme (Mellor and Yamada, 1974) is used for both vertical and horizontal turbulent transport in the boundary layer for all grids, and the LEAF-2 surface exchange scheme is used to parameterize the exchange of energy, moisture and momentum at the surface. The moisture complexity level is set to 2 on all grids. This allows water vapor to be condensed to cloud water if oversaturation occurs, although in this simulation clouds do not occur. For this reason, we use the simple Maher-Pielke shortwave and longwave radiation schemes in which clouds are neglected. The Kuo convective scheme is used for the outer 2 grids (see below for grid information).

All of these schemes require the use of preset parameters (for example, the nudging time in the Newtonian relaxation scheme). These are the parameters that will be perturbed by the evolutionary programming algorithm to create a new generation, as described in the next section.

3. **Experiment Description**

In practice a set of actual field observations would serve as the basis for the parameter optimization. In this case, however, the global minimum in the cost function and true optimum parameter set (that is, the best possible score and the parameters needed to obtain it) would not be known and our ability to evaluate the optimization method would be limited to what could be inferred from the decline in the cost function. Even if a significant reduction in the cost function is achieved, this does not guarantee that the global minimum has been found, nor do we know how close the optimized

parameters are to those corresponding to the true global optimum. In this study, we would like to maximize our ability to evaluate the methodology by using synthetic observations generated by a RAMS simulation with a specified set of parameters $\{P\}_T$. These parameters and the corresponding simulation will be referred to as the "target". In reality, a significant reduction in the cost function is certainly desirable, regardless of whether it represents the global minimum. If the simulation results are in better agreement with observations in a region and time period that the user is interested in, then it is likely that the model will perform better for whatever use the simulation was intended.

The RAMS simulation grid system is shown in Fig. 1. All simulations use 4 grids: 45km (grid 1), 15km (grid 2), 5km (grid 3), and 1.7km (grid 4), and are run from 1200 UTC 25 October 2000 to 1800 UTC 26 October 2000. This choice of grids and spatial resolution corresponds closely to those used by Zhong and Fast (2003) for their simulation studies supporting the Vertical Transport and Mixing Experiment (Doran, et. al., 2002). The time period corresponds to the 10^{th} field campaign of the VTMX program. This domain contains a wide range of topography and surface features, making it a challenging simulation for optimization since more of the model parameters are likely to have a significant impact on the simulation.

Vertical profiles at the Salt Lake City (SLC) airport were extracted from the target simulation at 6-hour intervals and used as the observational data by which the various RAMS simulations were scored. The choice of observational data (from the 'target' simulation) requires some discussion. With synthetic data, we know in advance that the model can perfectly reproduce the target simulation since it was used to generate the

target simulation in the first place (i.e., the model is 'perfect'). Optimization based on different sets of synthetic observations (number, location, and variable type) drawn from the target simulation will have cost functions with different response surfaces, although each of the cost functions will approach 0 as $\{P\} \rightarrow \{P\}_T$. Thus, the desired optimal solution exists for any set of synthetic observations. However, for small numbers of observations, it is likely that there will be multiple sets of parameters producing scores close to 0. How many of these "degenerate" parameter sets exist and how close they are to $\{P\}_T$ depends on the response surface. Thus, it is likely that the choice of observation data set will impact the optimization. Exactly how parameter convergence depends on the observational data is certainly a large study in itself. It does seem reasonable to assume that errors between the optimized simulation and the target simulation will increase with distance from the observational data. For this initial study a single sounding seemed like a reasonable starting point (as, in reality, it is often all the data that is available). In addition, the SLC airport sounding was one of the observational data sets used during the VTMX field campaigns.

For the evolutionary algorithm used, a set of 23 model parameters to be optimized $\{EP\}$ was selected and initial values for these parameters set. The simulation with these initial parameter values are herein referred to as the 'base' case simulation. These initial 'parent' parameters were then perturbed about their initial values based on a Gaussian distribution of variates with user-defined *initial* variances for each parameter. N sets of perturbed parameters were generated $\{EP\}_{n=1,N}$ where N is the number of available processors. We used 32 AMD Opteron processors with a single simulation run on each processor (N=32). Each simulation was run for 30 hours, although only the final 24

hours are utilized for parameter optimization. Each set of 32 simulations represents a 'generation' in the evolutionary process. After all the simulations were finished, the numerical results were compared to the synthetic Salt Lake Airport sounding during the final 24 hours of the simulations. An rms error was calculated for each simulation with a user-specified relative weight assigned to the different variables. Options exist also for spatial or vertical weighting of errors. Vertical weights are used primarily to limit the impact of errors above the boundary layer. In this simulation, vertical weights are equal below 1500m AGL and 0 above this level. Since there can be a large disparity in the number of surface observations and sounding data, an additional weight can be included control the impact of these two sources of error, but this was not used in the current optimization.

The weighting of meteorological observables must also be considered. This weighting should be assigned based on the end use of the simulation. If one is interested in regional transport, then upper air wind data over the transport region would probably be more heavily weighted than surface observations. In addition, dewpoint data would not be weighted as much as wind speed, direction and temperature. If one is interested in obtaining the best temperature estimate at a given location, however, then temperature data from nearby stations would be heavily weighted relative to all other meteorological observations. However, as we shall see, deleting a variable that is deemed irrelevant to the end-use of the simulation can seriously impact the optimization. For the first experiment in this study, wind speed, wind direction and temperature for the SLC sounding were given approximately equal weighting. The dewpoint was assigned only

10% of the temperature weight. For the other experiments, all 4 observational variables were given equal weighting.

After evaluating the total weighted rms error for each simulation, the parameter set producing the best score is selected and the entire process repeated with this parameter set serving as the new parent for the next generation. The variance of the offspring parameters will not be a constant, but will generally be reduced for each generation in a manner defined by the user. Typical factors controlling the offspring variance are generation number, best child score and child success rate. This reduction in variance is important. If it is too rapid, the likelihood of finding the global minimum in the cost function is reduced. On the other hand, if it is too slow the search becomes inefficient. It is also important to ensure that the optimization involves a "wide" search. In the procedure used here, it is assumed that the 32 initial perturbations will at least start the search with a fairly broad distribution of parameter values. The search breadth could be further improved by selecting more than one 'successful' child (i.e., scoring better that its parent) at each generation.

The 23 RAMS parameters that were optimized are given in Table 1 (EP1 to EP23) along with their target values. Most of them affect the model surface fluxes, since these are poorly known and therefore subject to a wide range of plausible variability. Several other variables, however, relate to uncertainties in the turbulence, radiation and the data used to initialize and control the boundary conditions for the model:

- EP1-4 are the topographical enhancement factors (one value for each grid) that specify how the topography initialization accounts for unresolved variance in the topography at the spatial resolution of the simulation, given

that the simulations used here employed the envelope orography procedure for creating the topography.

- EP5 is a standard RAMS parameter for calculating a surface roughness height based on sub-grid topography. EP6 is the minimum vertical velocity necessary for convection, and EP7 is the bare soil roughness.
- The albedos in Table 1 (EP8, EP9) are for dry soil and saturated soil. Albedos for intermediate soil moisture are linearly interpolated between these two values.
- In this optimization procedure, we adjust only the vegetation fraction for the 3 classes with the largest population on the 2 innermost grids: evergreen needle-leaf tree (EP10), short grass (EP11) and semi-desert (EP12).
- Soil moisture over the simulation domain was taken from the NCEP Global forecasting System (GFS) gridded data (formerly known as the Aviation model) at 1^{o} resolution and 2 vertical soil levels. This data was then interpolated to 11 soil levels in the RAMS simulation. The two soil moisture parameters in Table 1 (EP13, EP14) were used to scale all the soil moisture data before applying this data to the RAMS grids. This is very crude, but soil moisture is a high-impact parameter and is poorly known, thus it is an ideal candidate for optimization.
- EP parameters 15 and 16 are scale factors for the vertical length scales calculated for both convective and nocturnal turbulent diffusion. These are applied in the Mellor-Yamada turbulence parameterization that uses a

TKE-scaled length and a length scale due to Andre (1978) for stable conditions.

- The Pielke radiation scheme (which does not include the effect of clouds on radiative transfer) was used for these simulations since clouds were not a factor. A simple scaling parameter is used for the downwelling longwave radiation (EP17) to control the net radiation flux at the surface.
- EP18 is the cloud scale factor.
- EP19,20, and 21 are standard input parameters to RAMS that control the "nudging" (i.e., the Newtonian relaxation timescale in seconds) of the lateral, interior and top grid points to the results from a global simulation within which the RAMS simulation is nested.
- EP22 and EP23 are used to make small, horizontal adjustments to the interpolation of the global model data to the RAMS outer grid. This is used to account for small errors in the location of fronts or small scale features in the global data that are lead to significant inconsistencies with observations.

4. **Results and Discussion**

a. Optimization Methods – Cost function Behavior and Efficiency

1) Experiments with Variance based on Score

We begin with two optimization experiments in which the parameter perturbation variance is based on the value of the cost function or score. In the first (EXP1), the parameter variance factor σ was calculated as

$$s = e^{-\alpha(1-F)}, \qquad (1)$$

where F is the ratio of the best score of the current generation and the base (initial) score and α is a user defined constant. As the score improves, F will decrease and the variance of the progeny will decrease according to Eq. 1. This factor is used to scale all parameters being optimized at each generation. For the first experiment, α was initially set to 5, but manually adjusted to control the optimization procedure during the evolutionary process if it appeared that the best simulation 'score' at each generation (32 simulations) was changing too slowly or not at all (i.e., fallen into a local minimum). Of course, interactive control is not a particularly desirable feature for optimization and one goal of this study was to determine how to best control the parameter variances based on the best score history. The best score results for the first optimization attempt are shown in Figure 2. Note that the scores have been normalized with respect to the initial parent score (default parameters), so that all start at 1 and will ideally work their way to 0.

Initially the normalized score fell rapidly to about 0.52 after 30 generations. At this point, the decline abated and it appeared that the parameter variance reduction (following Eq. 1) was possibly too large. Suspecting that the optimization had become trapped in a local minimum, the variance was increased at generation 39 and again at generation 55 in an attempt to force the optimization out of the assumed local minimum and accelerate the convergence to the global minimum of 0. This resulted in another rapid decrease in the score to about 0.32 at generation 61. At this point the decline in the score slowed substantially, decreasing from 0.32 to 0.18 over the next 80 generations. Attempts to accelerate this decline in score by changing the variance had little impact (note the numerous changes in variance during this time).

At this point examination of the EP values parameters indicated that the near surface soil moisture (EP12) was far too high and consequently the diurnal variation of the surface temperature was too small and, more important, the surface dew-point temperatures were far too high. These results are shown in Figure 3, and reveal that the excessive soil moisture was being maintained despite selection pressure that should have favored a drier surface (with a larger diurnal temperature cycle). Although the error in the surface temperature was relatively large, this was having a small impact on the overall score. The dewpoint error should have led to a stronger selection of hydrological factors, but the in the weighting of the errors the dewpoint was only weighted at 10% of temperature error. With this in mind, the dewpoint error contribution to the cost function was increased by a factor of 4. This was done at generation 141. The result (Fig. 2) was an immediate increase in the score from 0.18 to 0.29 due to the additional dewpoint error term. This score was quite stable for 8 generations. Then the variance was again artificially increased. This produced an immediate but small increase in the score that was followed by a dramatic decrease that continued to the end of the optimization (again helped along by artificial manipulation of the variance). The improvement in near surface soil moisture (EP12) can be seen in Figure 4. Presumably, the same score reduction could have eventually been accomplished by increasing the temperature error weighting, but including a measured variable that is directly affected by a parameter is a better approach than overweighting a variable that is only indirectly affected by the parameter. The near-surface dewpoint is dominated by the soil moisture in this problem and thus it is likely that including the dewpoint resulted in a more rapid decline in the score than would have occurred with a larger weight for the temperature error. Over the

full 231 generations (32 RAMS simulations/generation), the score decreased from 1 to about 0.04.

A second optimization experiment was run (EXP2) with the dewpoint error factored in strongly from the start, and α initially set to 2 in the variance factor (Eq. 1). The results in Fig. 5 suggest that maintaining a larger variance in the parameter perturbations leads to a more rapid drop in the normalized score. However, similar to the 1st experiment, this decline dramatically slows at a score of about 0.25, remaining almost constant between generations 30 to 54. At this point, increasing the parameter variances by 50% proved to be sufficient impetus for a further score reduction to 0.04. While the final score is very similar to that in the 1st experiment, the efficiency of the optimization is better. Even allowing for some subjectivity in the manual application of the parameter variance factor, i.e. number of constant score generations before a variance change is applied, it is still clear that the 2nd experiment is significantly more efficient than the 1st experiment (the score is better after the same number of generations).

2) Experiment with Variance based on Child Success Rate

In this experiment (EXP3) the variance of the child parameters were perturbed, not according to the score of the best child simulation, but by the number of "successful" children per generation or success rate. "Successful" children are defined here as those simulations which generate a score lower than the previous generation's best score (their parent). For this optimization, the EP algorithm maintains a record of the number of successful children from the previous three generations (for a total of 96 children). If that number falls below a preset value (in this case, 3), the variance is reduced by 10% from

the current value. The reasoning behind this approach is that very low success rates (over several generations) imply that the current parent resides in a minimal region of the response surface with respect to the parameter perturbation variance. Since the search was initiated with a large variance, we assume that the parent is in a global (or at least, a deep) minimum. Therefore the variance should be reduced so that solutions far beyond that point will not be explored. This systematically localizes the search, forcing the score deeper into the minimum.

This approach proved to be very effective as can be seen in Figure 6 where, for comparison, the score results for EXP1 and EXP2 are shown again. The normalized score for EXP3 falls very rapidly, reaching 0.2 in only 15 generations. Note that in the first 2 experiments this benchmark was not achieved until generations 141 and 57, respectively. The detailed behavior of the normalized scores is interesting. All exhibit a rapid drop initially to about 0.5 and then a precipitous fall to the 0.2 – 0.25 level. This is followed by a slow, steady decline to the final score of approximately 0.04. The difference is in the number of generations required to achieve these score levels, in particular the 0.5 and 0.2 - 0.25 levels. Early in the optimizations, before any manual variance adjustments in EXP1 and EXP2, there are significant differences in optimization efficiency. The optimizations that maintain larger parameter perturbation variances exhibit faster declines in score with generation number. Recall that EXP1 and EXP2 used equation 1 for variance reduction, basing it on the score. Variance reduction based on child success rate is clearly the best method of those examined here. Even if an α exists that will match the efficiency of the success rate method, it is not clear how to choose it. The success rate method depends on the choice of a cut-off success rate and an

averaging period, but reasonable values for both of these are fairly easy to select. In EXP3, the variance scale factor holds steady at 1.0 until generation 18. At this point, the number of successful children from the previous 3 generations has fallen below 3, indicating that the optimization is closing in on a minimum and the search needs to become more localized. The variances of the child parameter perturbations are therefore reduced. This process is repeated numerous times, yielding a sequence of 10% reductions in the variance at various intervals. The process reaches an asymptote of 0.04 at about generation 85, much sooner than either of the previous experiments.

b. Optimization Results - Error Analysis and Convergence

1) Errors at SLC

Since the cost function or score all the experiments was reduced by about 96%, the synthetic observations at SLC should be closely reproduced by the optimized simulations. We now examine the initial and optimal SLC sounding with respect to the target for EXP3.

Fig. 7 shows that the initial temperature profile is too cool throughout the column relative to the target. After the 117th generation is run, the model profile matches the target profile very closely. The same is true for the wind direction – the model and target profiles differ by as much as 8º below 1500m and over 10º above this level. After 117 generations are run, the two differ by little more than 1º, matching through several wind shifts as the height increases from the surface to 3km aloft. The progression of these variables with each generation can be seen in Fig. 8, in which the values of temperature at 412m and wind direction at 1115m vary with each succeeding generation. We see that

both variables relax fairly quickly to within .2º (for temperature) and 1º (for wind direction) of their respective target values after about 30 generations, and in the following generations make progressively smaller oscillations as they move toward their targets.

2) Errors for Entire Simulation Domain

The synthetic Salt Lake City sounding was the only data included in the cost function. How will the rest of the model domain respond? Surface plots of temperature, dewpoint and wind vectors for grid 1 in Fig. 9 reveal the effect that optimizing the model with one vertical profile at SLC has on the entire domain. We see that the errors in temperature, dewpoint and wind are all significantly reduced after optimization

In the initial run, the temperature field (Fig. 9a) shows cool errors in California and a warm bias over most of Utah. The warm temperature bias appears to be a grid 2 effect and is probably caused by incorrect initial values of some of the grid-dependent parameters. Optimization removes this grid 2 bias completely. The cold bias still exists in California over the central Sierra Nevada but its magnitude has been reduced. A few new but small errors have been introduced in northern California.

A similar response to optimization occurs with the dewpoint (Fig. 9b). An initial dry bias exists in the west, particularly in Arizona and Nevada. As with temperature, this is greatly reduced (though not eliminated) after optimization. The wind errors (Fig. 9c) are large for the initial run, with the most serious problems along the west coast extending into southern California. After optimization, the errors are much smaller throughout the domain, and are mostly limited to Nevada and southern California.

It is certainly reasonable to expect that simulated grid points closer to the observations will produce better agreement with the target values. Since points on the inner grids are progressively closer to SLC, the inner grids should more closely match the target simulation. This can be seen in Table 2, in which the RMSE values for the 4 grids are shown after the final simulation has run. We see that the optimization at SLC improves grid 1 performance by 40-45% in RMSE value, while grids 3 and 4 experience ~90% improvement. Interestingly, the 3rd grid exhibits slightly more improvement than grid 4. The points on grid 3 are still close enough to SLC to be strongly influenced by that point and are thus 'forced' to produce a better forecast than the outer 2 grids.

3) Evolution of Model Parameters

As discussed in section 2, the fact that the cost function or score (based on a single vertical profile) approaches 0 does not guarantee that the optimized parameters converge to their target values. That this is so can be seen in the parameter evolution plots presented in Figure 10. In these figures, the optimized parameters are normalized by their target values (if those values are non-zero) and plotted versus generation number. Thus the target parameter values are either 0 or 1. Note that while the score does not monotonically decrease with generation number, a faster approach to 0 or 1 implies a more efficient optimization. These plots also reveal sensitivity of the model to different parameters since changes in parameters to which the model is very sensitive will result in large changes in the model solution, and hence, the score. This will encourage the rapid adjustment of these parameters as the algorithm strives to improve the score as quickly as

possible. Conversely, parameters that have little effect on the model score will feel little selection pressure and be allowed to drift more or less randomly.

Figure 10a-d shows the parameter evolution for the grid-dependent topographic enhancement factors (TEF) employed in RAMS using the "envelope orography" option. The TEF for grid 1 does not appear to converging to its target value of 1 for any of the tests. In fact the test with the most efficient optimization and best final score (EXP3) appears to be optimized with a TEF for grid 1 of about 2.7. For grid 2 (Fig. 10b) the TEF seems to be converging to 1, but the convergence is slow and certainly not complete for EXP1 or EXP3. The TEF convergence for grids 3 (Fig. 10c) and 4 (Fig. 10d) is much faster and complete. The TEF for grid 5 was added as a check (there was no grid 5) and as expected it wanders aimlessly in all three tests (not shown here). These results suggest that the order of TEF importance for optimization based on a single sounding at SLC (grid 4) is from grid 4 (most important) to grid 1 (least important). These results are intuitively satisfying, as model grid points for the inner grids are progressively closer to SLC, and they are influenced more strongly by the EP process that uses SLC as the metric.

Similar results for vegetation are shown in Figure 10e-g. Here the vegetation fraction of the 3 most dominant vegetation types are shown as a function of generation for all three tests. The results suggest that vegetation fraction 2 (short grass, Fig. 10f) is the most important for optimization with convergence to 1 appearing to occur in all three tests. Vegetation fraction 1 (evergreen trees, Fig. 10e) seems to be converging to different values for each test but all 3 values are fairly close to 1. The evolution of

vegetation fraction 3 (semi-desert Fig. 10g) seems to be converging to 1.4 for EXP1 and 3, and to 2.0 for EXP2.

Certainly the most significant parameter in the optimization was the scale factor applied to the downwelling long-wave radiation used in the Pielke radiation parameterization. These results are shown in Figure 10h. In all 3 tests, this parameter found its target value of 1 quickly and exhibited little variability about that value. This is not too surprising since low level temperatures and flows under stable conditions are dominated by the net radiation flux at the surface. Results for all the model parameters used in the optimization are given in Table 3.

5. Principal Component Analysis

The optimized parameters are not completely independent of one another – some can have similar effects on the model variables and might therefore undergo similar adjustment by the evolutionary algorithm. For example, surface temperature is affected by both soil wetness and vegetation fraction. If the soil is too cool, soil wetness and vegetation fraction may both be adjusted downward by the evolutionary algorithm to compensate.

Given that the 23 variables are certainly correlated to some extent, it is possible to do a principal component (PC) analysis to identify the dominant patterns as the system evolves in time. By reducing the 23-parameter system to a subset of eigenvectors, we could define a reduced parameter space and perhaps apply the algorithm more efficiently in future runs.

By treating each of the 23 parameters as a time series (actually a generation series), we apply a PC analysis to the data to create a series of eigenvectors, each of which represents a set of values for each of the 23 parameters. Since the values of the different parameters vary over different orders of magnitude, each time series is normalized by its standard deviation before applying the PC analysis. The first 3 eigenvectors are shown in Fig. 11a, and reveal how certain variables tend to vary together (or oppositely) within a single eigenvector. For example, in the first eigenvector, the 2 roughness factors vary together, and the 3 vegetation fractions vary with the upper-level soil moisture. The respective coefficients of the first 4 eigenvectors are shown in Fig. 11b, and show how the first 4 eigenmodes have large deviations for the first half of the simulation but then settle towards final values in the latter half.

Fig. 12a shows the cumulative variance explained by the different eigenmodes, and it reveals that 95% of the variance is explained by the first 10 eigenvectors. At any value of g, therefore, the system can be well represented as a linear combination of these 10 eigenvectors. In Fig. 12b and 12c, we see both the original time series of the topographic factor and the interior nudging time, as well as the reconstructed series using only the first 5 eigenvalues, and they tend to match very closely.

This analysis can allow for the reduction in the number of adjusted parameters – instead of manipulating 23 parameters, the process can instead manipulate 10 eigenvectors. By reducing the dimensionality of the parameter space, we can sample a larger fraction of that subspace with the same computer power, allowing for a more rapid progress to the target. We are currently testing this methodology and will report on the results in a future publication.

6. Conclusions

A simple evolutionary algorithm is capable of optimizing a large set of model parameters for a meteorological simulation and based on agreement with a single vertical profile of synthetic observational data (wind speed, wind direction, temperature and dewpoint) with no measurement uncertainty. For the optimization studies performed here, the cost function (based on a weighted RSM error) was reduced by 96-99%. For an evolutionary method based on the child success rate, this high degree of optimization was achieved in under 100 generations using a 32 CPU cluster. As expected, the agreement between the optimal simulation and the synthetic grid data decreased with distance from the 'observations', but not in a simple linear fashion. The results presented here also suggest that all possible data types be included in the optimization if they are deemed reliable. Even though a particular data type (e.g. dewpoint) may are be considered less important with respect to the application at hand, it can still be critical to finding the global optimum. Finally, a PC analysis of the parameter optimization indicates that an evolutionary method based on a much smaller set of parameters can provide an extremely efficient approach to model optimization. In addition, to developing an algorithm for variance reduction we are also pursuing efficiency gains through the use of principal component analysis.

ACKNOWLEDGEMENTS

This research was supported by the National Nuclear Security Administration of the Department of Energy. This research used resources of the National Center for Computational Sciences at Oak Ridge National Laboratory, which is supported by the Office of Science of the Department of Energy under Contract DE-AC05-00OR22725. We would also like to acknowledge Dr. Susan Haupt for her suggestion to use synthetic data in our initial optimization studies.

REFERENCES

Andre, J.C., Ge De Moor, P. Lacarrere, G. Therry, and R. Du Vachat, 1978: Modeling the 24-hour evolution of the mean and turbulent structures of the planetary boundary layer. *J. Atmos. Sci*., **35**, 1861-1883

Bäck, T., U. Hammel, and H.P. Schwefel, 1997: Evolutionary computation: comments on the history and current state, *IEEE Trans. Evolutionary Computation*, **1**, pp. 3-17

Doran, J.C., J.D. Fast, and J. Horel, 2002: The VTMX 2000 campaign, *Bull. Amer. Meteor. Soc*. **83**, 537-551

Fogel, D., 2000: Evolutionary Computation, Spie Press, Bellingham, Washington, USA

Haput, S. E., 205: A demonstration of coupled receptor/dispersion modeling with a genetic algorithm, *Atmo. Env*. **39**, 7181-7189

Kalnay, M. Kanamitsu, and W.E. Baker, 1990: Global numerical weather prediction at the National Meteorological Center. *Bull. Amer. Meteor. Soc*., **71**, 1410-1428.

Mellor G. L. and T. Yamada, 1974: A hierarchy of closure models for planetary boundary layers, *J. Atmos. Sci.,*. **31**, 1791-1806

Pielke R. A., W. R. Cotton, R. L. Walko, C. J. Tremback, W. A. Lyons, L. D. Grasso, M. E. Nicholls, M. D. Moran, D. A. Wesley, T. J. Lee, and J. H. Copeland, 1992: A comprehensive meteorological modeling system--RAMS. *Meteor. Atmos. Phys.*, 49, 69-91.

Schwefel, H.P., 1995, *Evolution and Optimum Seeking*, John Wiley, New York

Zhong,S. and J.D. Fast, 2003: An evaluation of the MM5,RAMS and Meso-Eta models at subkilometer resolution using VTMX field campaign data in the Salt Lake valley. Mon. Wea. Rev., 131, 1301-1322.

FIGURE CAPTIONS

Figure 1 RAMS Grids for Salt Lake City simulation. See text for grid resolutions.

Figure 2 Variance (thin line) and best score (thick line) of the model runs as a function of generation for EXP1 (Generation 1 not on plot).

Figure 3 a) Maximum and minimum temperature soundings at SLC (10/26, 18Z) after 93 generations for EXP1 for the model (thick lines) and the observed data (thin lines) b) as in a) but for dewpoint.

Figure 4 Upper level scaled soil moisture as a function of generation for EXP1.

Figure 5 As in Fig. 2 but for EXP2.

Figure 6 Best score for EXP1 (thick dotted line), EXP2 (thick dashed line), EXP3 (thick solid line), and the variance for EXP3 (thin line)

Figure 7 a) Temperature and b) wind direction soundings on 1800 UTC 26 October 2000 at the SLC station for EXP3. The curves are for the 1st generation (solid line), the 117th generation (dashed line), and the target value (crossed line).

Figure 8 a) 412m temperature above SLC as a function of generation. b) As in Fig. 9 a), but for wind direction at 1115m.

Figure 9 Surface temperature (ºC, top), dewpoint values (ºC middle) and winds (m/s, bottom) on 1800 UTC 26 October 2000 (1pm) for the original run (left) and the EXP3 final run (right).

Fig. 10 Evolution of model parameters as a function of generation for EXP1 (dashed line), EXP2 (solid line), and EXP3 (dotted line) for the a) topographic factor 1,

b) topographic factor 2, c) topographic factor 3, d) topographic factor 4, e) vegetation fraction 1, f) vegetation factor 2, g) vegetation factor 3, and h) longwave factor.

Figure 11 a) Value of each model parameter in the first (black), seciond (grey), and third (white) dominant EXP3 eigenvectors. Each has been normalized by its standard deviation. See Table 1 for the variable corresponding to each number. b) Magnitudes of the first (thick line), second (dashed line), third (dotted line) and fourth (thin line) dominant EXP3 principal components as function of generation.

Fig. 12 a) Cumulative variance explained by EXP3 EP Eigenvectors. b) Original and reconstructed series for the Topographic factor 1. c) As ion b) but for the interior nudging time (bottom).

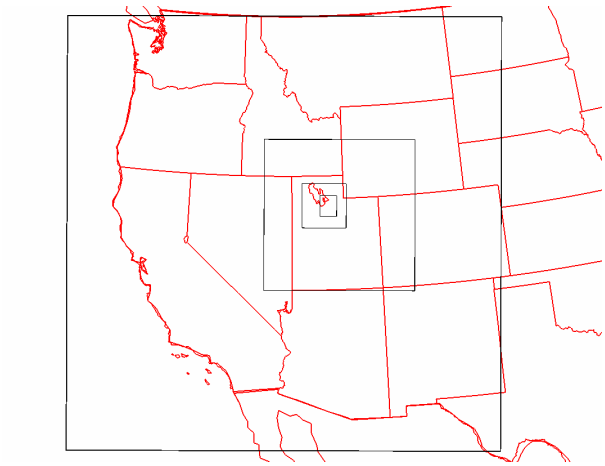

Figure 1 RAMS Grids for Salt Lake City simulation. See text for grid resolutions.

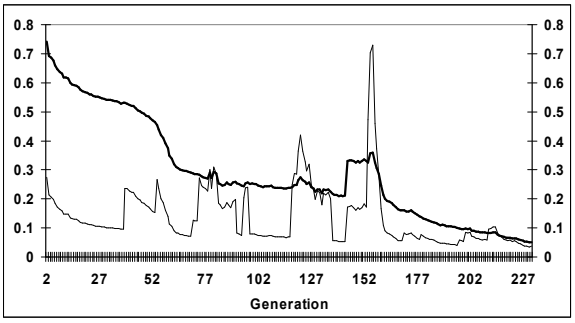

Figure 2 Variance (thin line) and best score (thick line) of the model runs as a function of generation for EXP1 (Generation 1 not on plot).

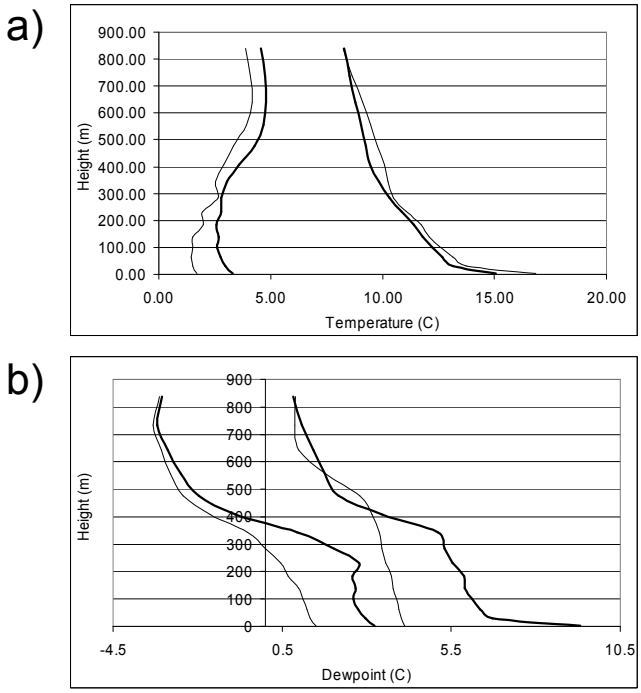

Figure 3 a) Maximum and minimum temperature soundings at SLC (10/26, 18Z) after 93 generations for EXP1 for the model (thick lines) and the observed data (thin lines) b) as in a) but for dewpoint.

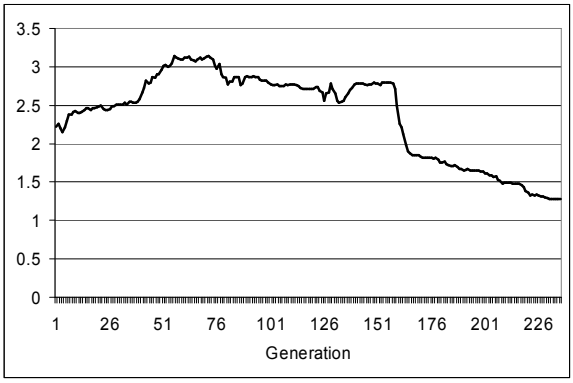

Figure 4 Upper level scaled soil moisture as a function of generation for EXP1.

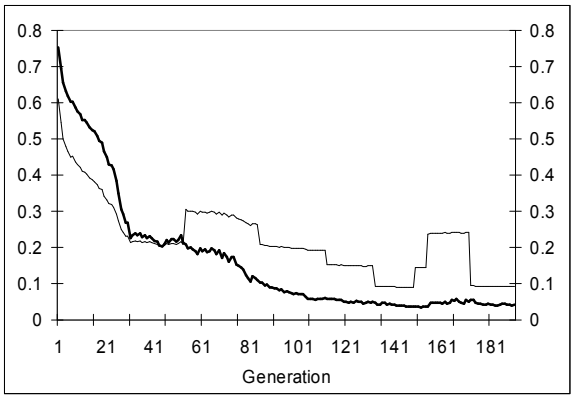

Figure 5 As in Fig. 2 but for EXP2.

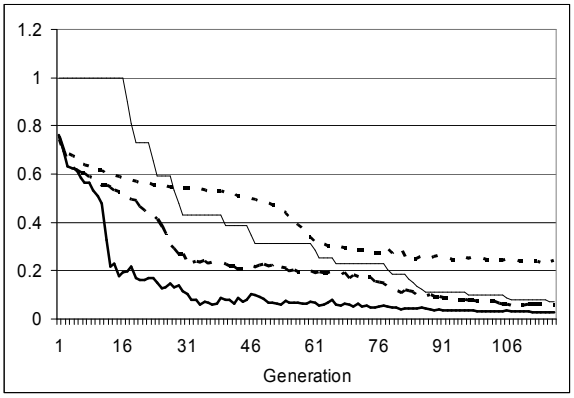

Figure 6 Best score for EXP1 (thick dotted line), EXP2 (thick dashed line), EXP3 (thick solid line), and the variance for EXP3 (thin line)

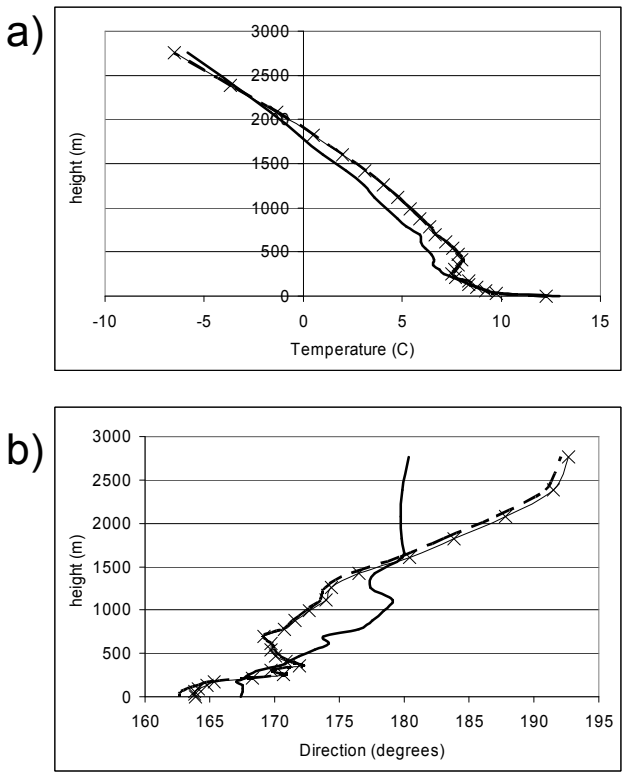

Figure 7 a) Temperature and b) wind direction soundings on Oct. 26, 18Z at the SLC station for EXP3. The curves are for the 1st generation (solid line), the 117th generation (dashed line), and the target value (crossed line).

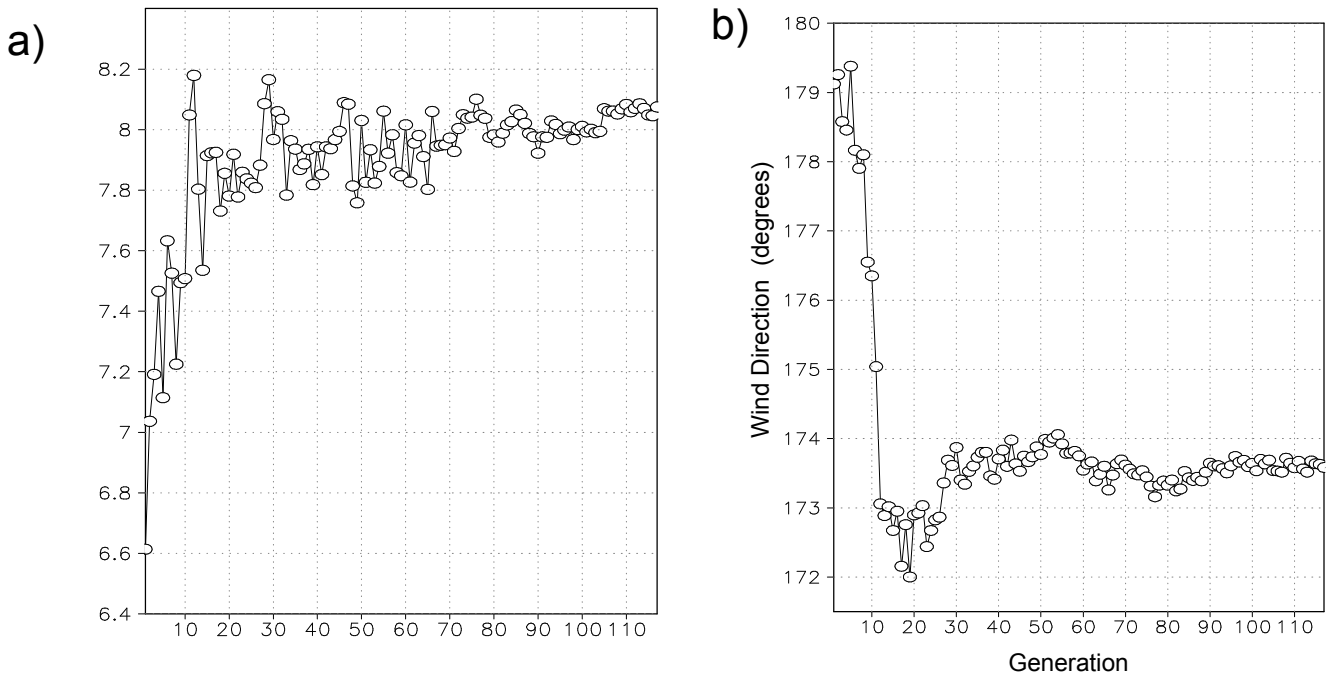

Figure 8 a) 412m temperature above SLC as a function of generation. b) As in Fig. 9 a), but for wind direction at 1115m.

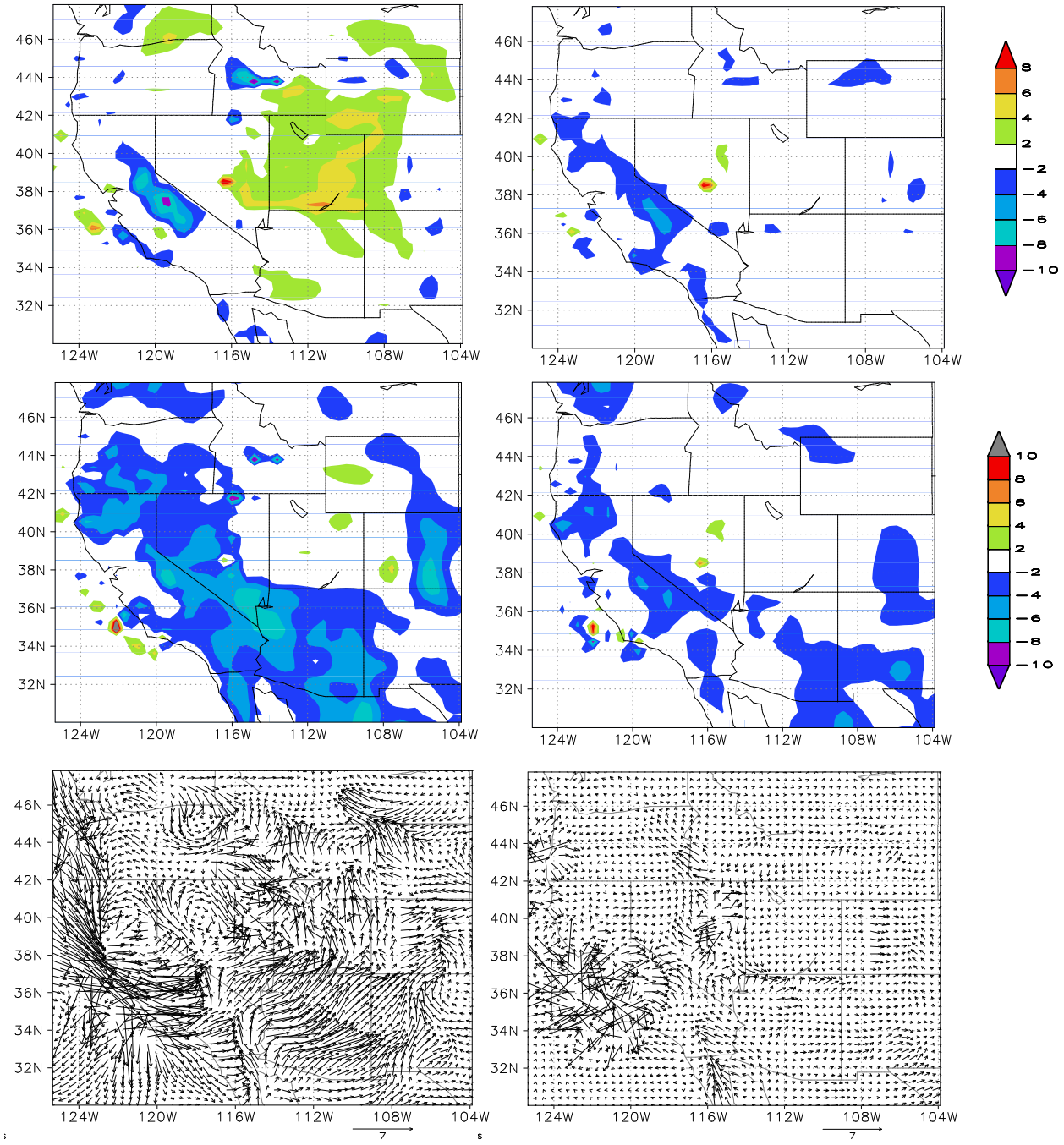

Figure 9 Surface temperature (ºC, top), dewpoint values (ºC middle) and winds (m/s, bottom) on October 26, 18z (1pm) for the original run (left) and the EXP3 final run (right).

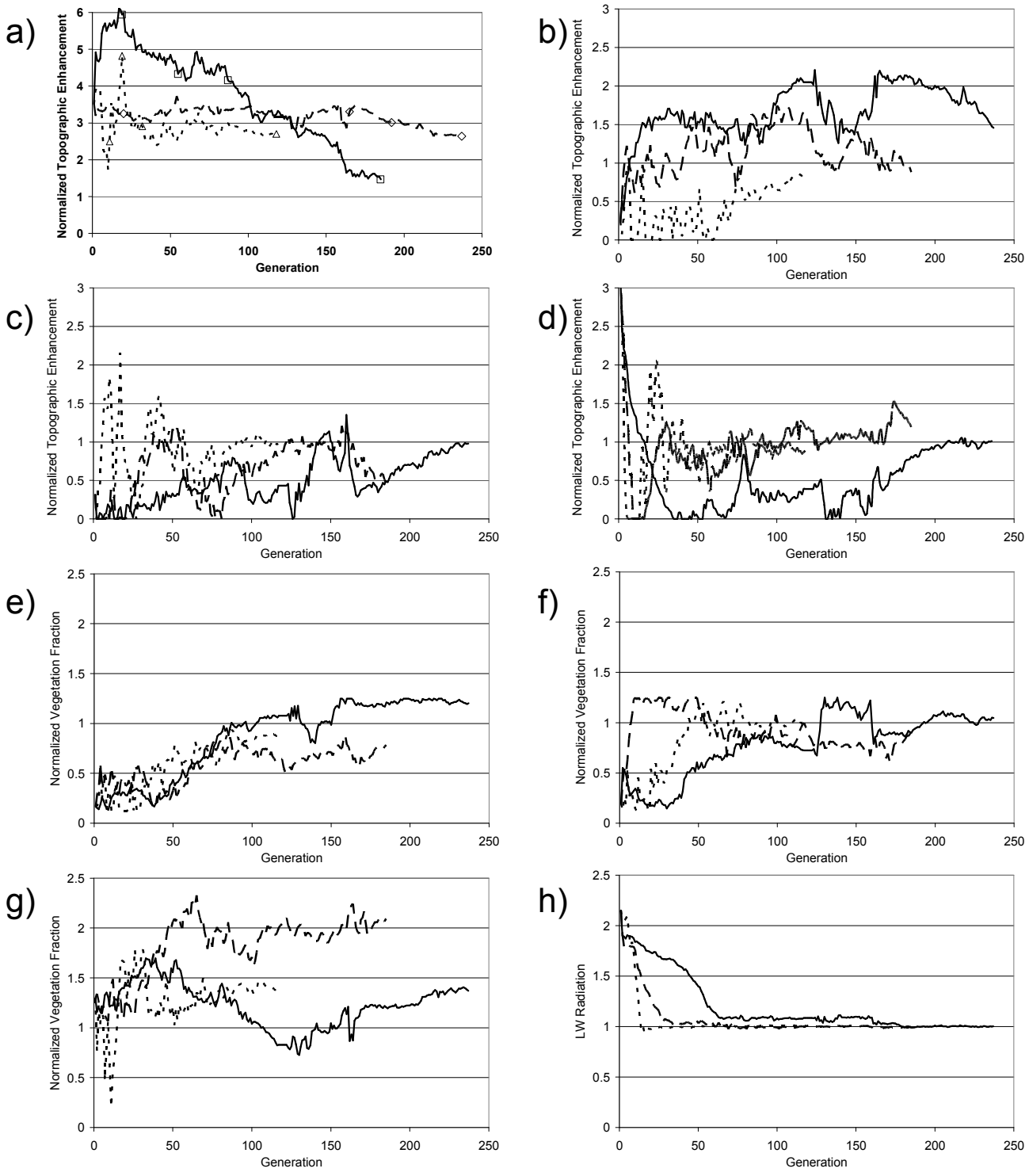

Fig. 10 Evolution of model parameters as a function of generation for EXP1 (dashed line), EXP2 (solid line), and EXP3 (dotted line) for the a) topographic factor 1, b) topographic factor 2, c) topographic factor 3, d) topographic factor 4, e) vegetation fraction 1, f) vegetation factor 2, g) vegetation factor 3, and h) longwave factor.

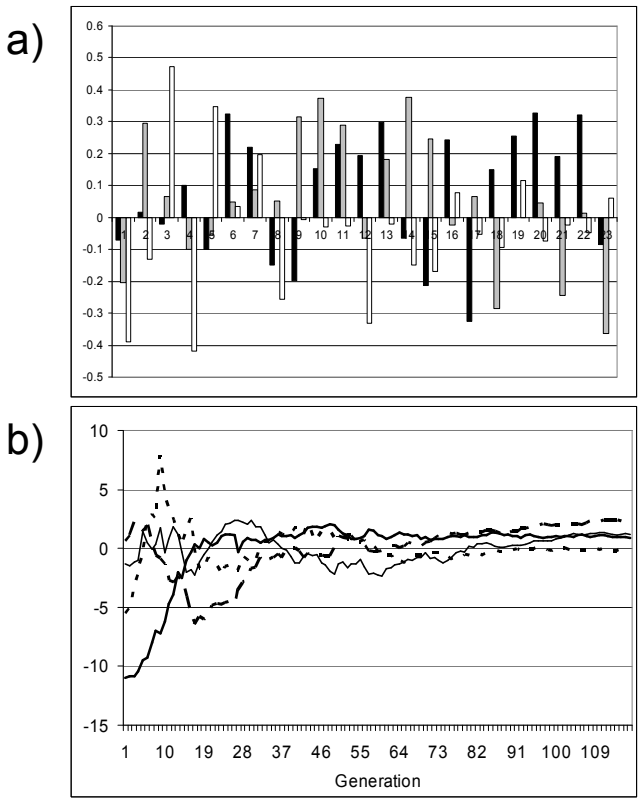

Figure 11 a) Value of each model parameter in the first (black), seciond (grey), and third (white) dominant EXP3 eigenvectors. Each has been normalized by its standard deviation. See Table 1 for the variable corresponding to each number. b) Magnitudes of the first (thick line), second (dashed line), third (dotted line) and fourth (thin line) dominant EXP3 principal components as function of generation.

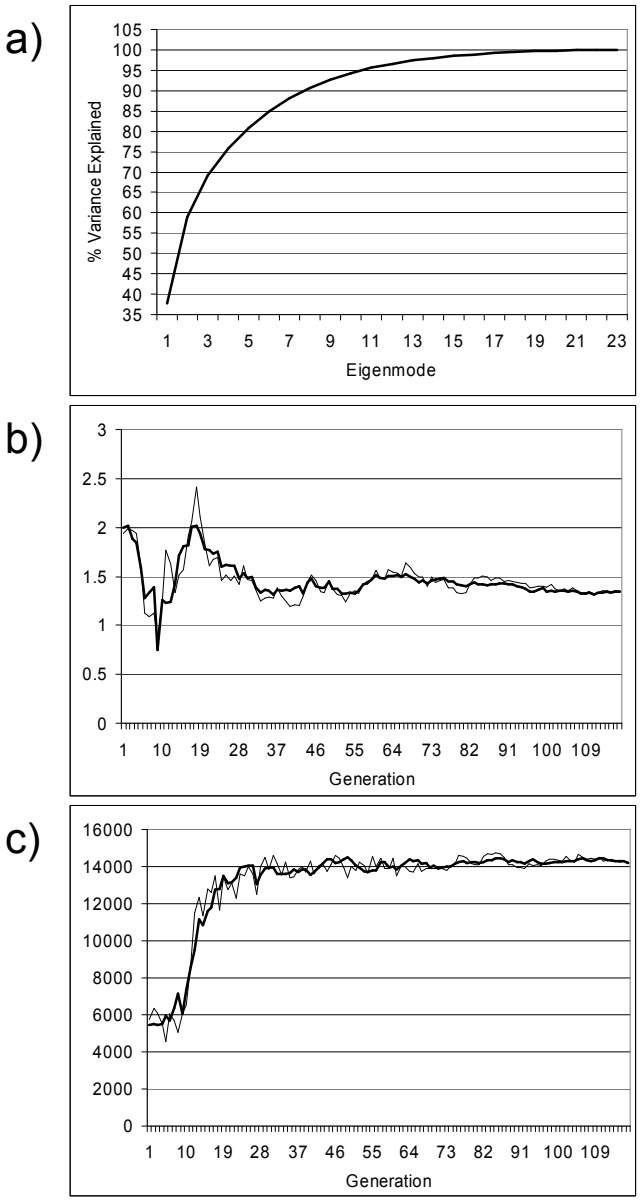

Fig. 12 a) Cumulative variance explained by EXP3 EP Eigenvectors. b) Original and reconstructed series for the Topographic factor 1. c) As ion b) but for the interior nudging time (bottom).

Variable	Target Value	Name
EP1	.5	Topo Init for Grid 1
EP2	.5	Topo Init for Grid 2
EP3	.5	Topo init for Grid 3
EP4	.5	Topo init for Grid 4
EP5	0	SGS Topo Roughness
EP6	.002	Minimum vertical velocity
EP7	0.1	Soil Roughness
EP8	0.31	Dry Albedo
EP9	0.14	Wert Albedo
EP10	0.8	Veg. Frac. #1 (Evergreen tree)
EP11	0.8	Veg. Frac. #2 (Short Grass)
EP12	0.1	Veg. Frac. #3 (Semi-Desert)
EP13	0.5	Soil Moisture (top)
EP14	0.5	Soil Moisture (bottom)
EP15	1	Convective Vert. Scale
EP16	1	Nocturnal vert. Scale
EP17	1	Pielke LW down
EP18	.100	Cloud scale factor
EP19	900	Lateral Nudging Timescale
EP20	28800	Interior Nudging timescale
EP21	28800	Top Nudging Timescale
EP22	0	RAMS grid shift –X
EP23	0	RAMS Grid Shift - Y

Table 1 RAMS parameters subject to adjustment by the evolutionary procedure.

Table 2 Relative improvement in RMSE values at Oct. 26, 18Z between the initial and final runs for all 4 grids. Each value is calculated as the difference between the final and initial RMSE values, divided by the initial values

	Temperature (K)	Relative Humidity (%)	Wind Speed (m/s)
Grid 1	0.407	0.382	0.454
Grid 2	0.816	0.559	0.682
Grid 3	0.943	0.935	0.918
Grid 4	0.922	0.928	0.901

Table 3 – Final values of the EP parameters for the 3 experiments

	VAR1	VAR2	VAR3
Topo Init for Grid 1	1.32	0.65	1.35
Topo Init for Grid 2	0.73	0.56	0.40
Topo init for Grid 3	0.49	0.25	0.50
Topo init for Grid 4	0	0	0.44
SGS Topo Roughness	0	0	0.84
Min. vertical velocity	.0213	.0132	.045
Soil Roughness	0.0009	0	0.0003
Dry Albedo	0.021	0.013	0.045
Wert Albedo	0.0889	0.1023	0.1197
Veg. Frac. #1 (Evergreen tree)	0.3798	0.1500	0.3274
Veg. Frac. #2 (Short Grass)	0.1177	0.1067	0.0982
Veg. Frac. #3 (Semi-Desert)	0.9626	0.6906	0.6588
Soil Moisture (top)	0.8374	0.7098	0.7681
Soil Moisture (bottom)	0.1372	0.1908	0.1406
Convective Vert. Scale	0.1893	0.8848	0.7142
Nocturnal vert. Scale	0.6405	0.4463	0.2073
Pielke LW down	0.3792	0.9046	0.6329
Cloud scale factor	.8817	.9970	1.1490
Lateral Nudging Timescale	0.1076	0.2276	0.0741
Interior Nudging timescale	1.0089	1.2402	1.0056
Top Nudging Timescale	1.0009	0.9935	1.0072
RAMS grid shift –X	0.8817	0.997	1.149
RAMS Grid Shift - Y	1297.9897	1310.6383	1307.901