Storm of the Century? Insights from a Massive Ensemble Forecast Experiment

Peter S. Dailey[1], Robert G. Fovell[2]* and Steven Van Horn[2]
[1]AIR-Worldwide, Boston
[2]University of California, Los Angeles

1. Introduction

In March, 1993, a "Nor'easter" dubbed "Superstorm 1993" and the "Storm of the Century" (SOC) brought massive snows and damaging winds to a large portion of the Eastern United States. It was born to be a very significant event, having occurred during a period in which the Available Potential Energy (APE) in the Northern Hemisphere was far in excess of its climatological values (Bosart et al. 1996). True to its moniker, the SOC event ranked first on Kocin and Uccellini's (2004; "KU") list of 70 storms over a roughly 100 year period, as measured by their North East Snow Impact Scale (NESIS). But was the SOC just a one in a century storm or a more or less rare event? Given substantially similar initial conditions, what is the probability that an even more severe Nor'easter could have been produced? Many parties, particularly property insurers, are interested in this question, but the answer is not clear from the all-too-short historical record.

Below we describe an (as yet incomplete) effort to address the recurrence interval and exceedance probability questions via a massive forecast ensemble experiment using a mesoscale numerical model. Our working hypothesis is that the actual SOC event is but one sample from a population of storms that could have transpired given slightly different initial conditions. We further expect the population is at least roughly normally distributed with respect to distinguishing characteristics such as snow and wind production and intensity, and that it can be reproduced via randomly perturbing the initial conditions.

To address this hypothesis, the crucial step is to identify where the actual storm falls within its own distribution of potential events. If that distribution is indeed normal, it is most likely that the real case came from the middle of the pack, which means that many potential SOCs could have been even stronger. However, if the actual SOC event came from the particularly intense end of the spectrum, we contend it was about as strong as it could have been, and thus correspondingly rarer an event. In that case, *perturbations should be more likely to create a weaker than stronger storm.* It is clear that selection of the benchmark ensemble member best representing "truth" is a central and difficult issue in this work.

2. The SOC ensemble

The simulations described herein were made using release 3.7.2 of the Penn State/NCAR Mesoscale Model, version 5 (MM5). Two domains were employed, a 90 km mesh encompassing much of North America and a 30 km nest centered over the Eastern United States. The initial conditions were drawn from ECMWF reanalyses. Model options were selected as a result of a physics-based experiment, the winning combination having resulted in the smallest root mean square (RMS) sea-level pressure (SLP) error relative to the ECMWF analysis. The physics options selected were "simple ice" microphysics, the MRF boundary layer scheme, CCM2 radiation and no cumulus scheme. The evaluation metrics are discussed in more detail below.

The SOC event was defined as the 60 hour period between 12 UTC March 12 and 00 UTC March 15 (Fig. 1). To enhance ensemble diversity, we chose to incorporate simulations initiated at a variety of starting times leading up to the event. Simulations initialized at 12 UTC March 12 were the "0 h lead runs", while those commencing 6 hours earlier represented 6 h leads, etc., the time interval reflecting the ECMWF reanalysis frequency. Each lead time presented the model with a unique atmospheric state that might evolve along different, if parallel, tracks, ideally providing a legitimate sample of what might have transpired during the SOC period had the event played out differently. That said, this strategy necessarily requires that the model maintain skill over a reasonable period of time, and increases the importance of properly identifying the benchmark or truth

---

*Corresponding author address: Prof. Robert Fovell, UCLA Atmospheric and Oceanic Sciences, Los Angeles, CA 90095-1565. E-mail: rfovell@ucla.edu.

case. Lead times out to 60 h were considered; longer simulations were also made for reference purposes.

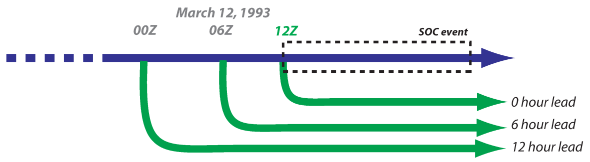

Fig. 1: Lead time strategy utilized in the SOC ensemble.

The *standard ensemble* consists of one *control run* for the each of the 16 lead times extending back to 18 UTC March 8th as well as a set of *perturbed runs* for the 0-60 h lead times. The control runs were initialized with the ECMWF fields at the appropriate times and integrated forward to 00 UTC March 15. Perturbed runs were created by decomposing the outer 90 km domain into individual vertical columns and then modifying each column separately with perturbations drawn from a uniform distribution bounded by $\pm X\%$, where X is a specified magnitude. Integer magnitudes between 1 and 9% were considered. For each column, the chosen perturbation was applied to both temperature and dewpoint at each level, in a manner that preserved both vertical lapse rate and relative humidity. Three trials for each perturbation magnitude were attempted, each utilizing a different, randomly selected seed. This resulted in a total of 27 perturbed runs for each lead time considered.

For each perturbed run, the vertical columns were recomposed into an initialization via Cressman objective analysis from the LITTLE_R package. Owing to discretization and possible smoothing, this procedure can potentially alter the initial fields even if no perturbation is actually applied. Thus, each of the 0-60 h lead times also possess an *unperturbed run*, which represents a simulation made following application of a 0% alteration to each column. In total, the standard ensemble consists of 16 control cases, 11 unperturbed runs, and 297 perturbed trials, for a total of 324 simulations.

Beyond the standard ensemble, three special ensembles have been created. For the 18, 42 and 60 h lead times, 81 additional trials were made using 5% perturbations and different random seeds. The 5% value was chosen as an intermediate magnitude having demonstrated ability to provoke useful diversity in the standard ensemble. Thus, the "18 h lead" ensemble contains a total of 110 cases (1 control, 1 unperturbed and 108 perturbed runs) sharing a common lead time, as do the "42 h lead" and "60 h lead" ensembles. One use for these extended simulation sets is to assess how well the standard ensemble perturbation strategy samples the potential range of variation at the selected lead times. The *entire ensemble* includes these special simulations and encompasses 567 simulations.

3. Evaluation metrics

The SOC was a snow and wind event, but the comparison of model outputs directly to observations of either is fraught with difficulty, if only owing to the irregular spacing and uncertainties inherent in the meteorological measurements. We decided at a more objective and easily computed field – SLP – can be useful for identifying the benchmark member. RMS SLP difference fields between each simulation and the ECMWF analysis represented on the 30 km MM5 model grid were computed. Constructing a SLP error measure (SLPER) started with summing the squared differences over a specified area (between 25° and 50°N east of 95°W longitude) and time (between 06 UTC March 13 and 06 UTC March 14, inclusive), capturing the cyclone while it was a closed circulation over the U.S. mainland and before it ceased producing appreciable snowfall in New England. The ensemble member with the minimum SLP error, expressed as average absolute pressure difference per gridpoint, would be most similar to the analysis with respect to *intensity*, *track* and *timing*, and thus a viable candidate to represent truth.

Other evaluation metrics will include two considering snowfall (TSLD2 and NESIS), one for wind (WINDMAX) and one for intensity (not considered further herein). TSLD2 is the summed snowfall over land in the 30 km domain during the SOC event, computed by accumulating precipitation recorded at the surface when and where the boundary layer temperature was subfreezing. NESIS is computed using area-averaged snow depths weighted by population. Census data were interpolated onto the 30 km grid and snow depths were calculated using a formula based on water content and temperature. WINDMAX is a nondimensional measure of areally-integrated estimated wind gusts over land exceeding a damage threshold set at 20 m s^{-1}, at which speed property damage typically begins.

4. Ensemble results

Figure 2 presents SLPER vs. lead time for the entire ensemble. Unsurprisingly, the smallest errors are associated with the shortest simulations. Those leads

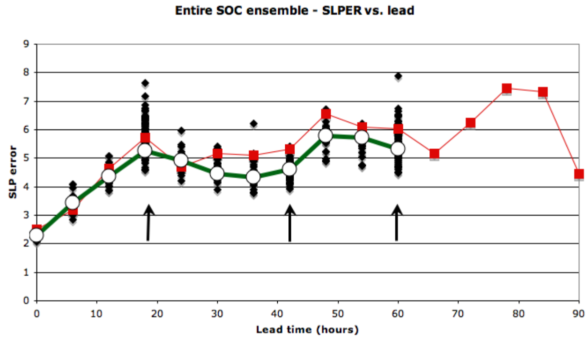

Fig. 2: SLPER (RMS SLP error, in millibars per gridpoint per evaluation time) vs. lead time for the entire SOC ensemble. Control, unperturbed and perturbed runs are indicated by the red squares, open black circles and black diamonds, respectively. Arrows indicate lead times having the extra (special ensemble) members.

also tended to have less spread, probably because the simulations had insufficient time to diverge further. The standard ensemble's benchmark run is identified as the simulation with the smallest SLPER. This was one of the 0 h lead members, having a 5% perturbation magnitude and an average absolute pressure error of 2.1 mb per gridpoint over the evaluation time period. For comparison, the 0 h lead control run has a SLPER of 2.5 mb, actually the largest error at this lead time. The SLPER temporal error growth rate lowered markedly after the 18 h lead.

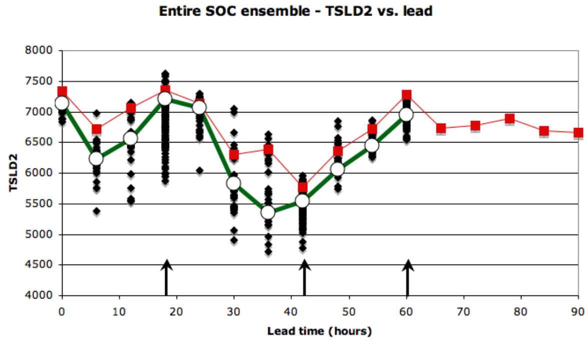

Fig. 3: As in Fig. 2 but for TSLD2 (total snow) vs. lead time for the entire SOC ensemble.

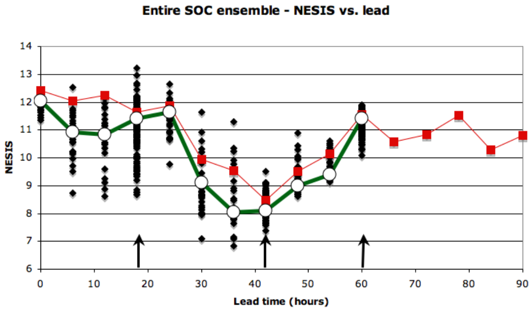

Fig. 4: NESIS values for the standard ensemble presented in rank order. The location of the benchmark run is identified.

Total snow production and estimated NESIS values during the event period is shown in Figs. 3 and 4. Note the smallest values for both are actually associated with the intermediate lead times (30-48 h). The case with the largest NESIS was one of the 18 h lead runs, also having received a 5% perturbation. Its value of 13.2 exceeds that of the actual SOC event (12.52), although calibration issues introduce some uncertainty in this comparison. The main point is the entire ensemble is composed of significant events, having a NESIS values exceeding 6.8 and thereby placing them in Kocin and Uccellini's categories 4 and 5 ("Crippling" and "Extreme"). All would rank among the top 7 cases in KU's Table 6. However, most of them are weaker than the benchmark case (NESIS = 12.3), suggesting that the SOC event fell well into the strong end of its own distribution. This is emphasized in Fig. 5, which presents the standard ensemble NESIS values in rank order; only 10 members of 324 exceed its NESIS value. The top three events in the ensemble came from the 18, 24 and 6 hour lead times, respectively.

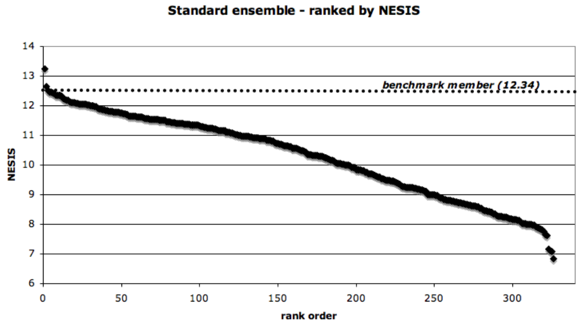

Fig. 5: As in Fig. 2 but for estimated NESIS vs. lead time for the entire SOC ensemble.

Figure 6 presents storm tracks for the benchmark case and NESIS extrema, as well as the former's snow depth footprint. Compared to the benchmark case, the NESIS maximum track is more inland in the southern and New England states. This storm generated more snowfall in population centers such as Atlanta, increasing its population-weighted total. The NESIS minimum case, in contrast, moved somewhat closer to the coast than the benchmark run.

The among-lead ensemble spread for the WINDMAX wind metric was large relative to the within-lead

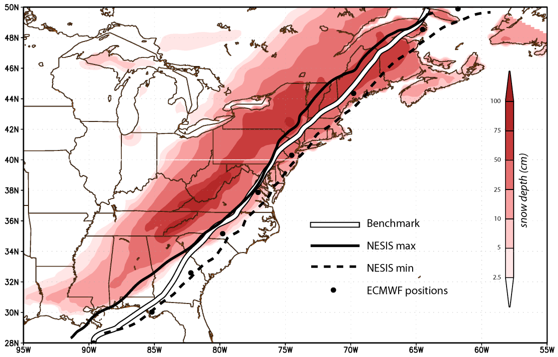

Fig. 6: Storm tracks for the benchmark and NESIS extrema cases, ECMWF analysis cyclone positions, and snow depth footprint for the benchmark case.

variation (Fig. 7), which is a favorable aspect for the ensemble. As with snow, the intermediate times generated the weakest wind events. One of the perturbed members of the 0 hour lead produced far and away the greatest amount of areal coverage for estimated wind damage; its wind footprint is shown in Fig. 8. As occurred in the actual event, the fastest winds generally occurred on the east side of the track, and/or in coastal areas exposed to air having passed over the smoother ocean surface.

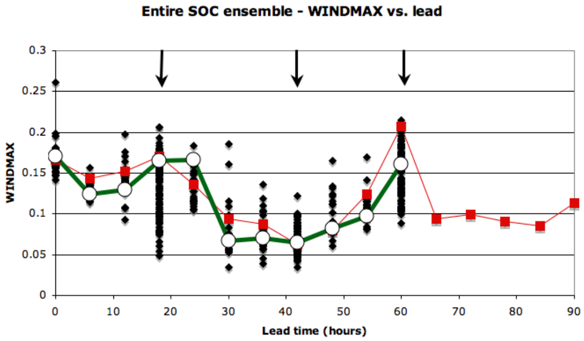

Fig. 7: As in Fig. 2 but for WINDMAX vs. lead time for the entire SOC ensemble).

Figure 9 represents a combined snow/wind metric, created by adding standardized NESIS and WINDMAX values for the entire ensemble. The top WINDMAX case presented in Fig. 8 also represented the ensemble's seventh largest NESIS value, and ranked first overall in this combined assessment. The benchmark case ranked fifth by this measure, meaning that only 4 of 567 members of the entire ensemble – 0.7% of the total – represented a greater snow and wind threat. If our strategy for reproducing the SOC storm's own distribution is valid and reasonable, the implication is the return period for storms of equal and greater magnitude is relatively long.

5. Examination of the perturbation strategy

In the standard ensemble as a whole, perturbations were as likely to produce greater snow producers as lesser ones, at least when compared to their lead time's unperturbed case. Figure 10 presents the distribution of the ratio between perturbed and unperturbed TSLD2 for all lead times in the standard ensemble. The mean of the 297 members is very nearly 1.0 and the distribution passes for normal. This demonstrates that the act of perturbing the simulations is fundamentally unbiased. The extreme cases were 24% stronger and 16% weaker than their respective unperturbed simulations.

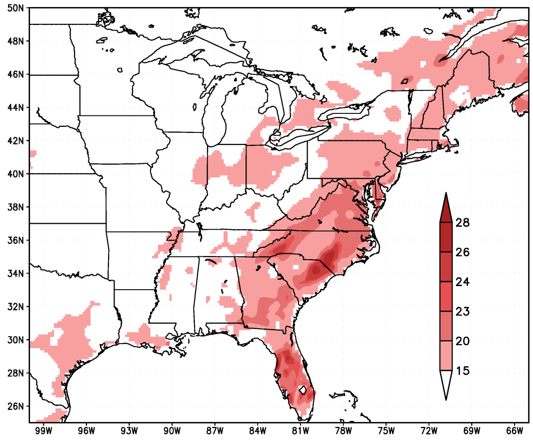

Fig. 8: Wind footprint for the WINDMAX case. Shaded field shows estimated maximum wind gust, in m s^{-1}).

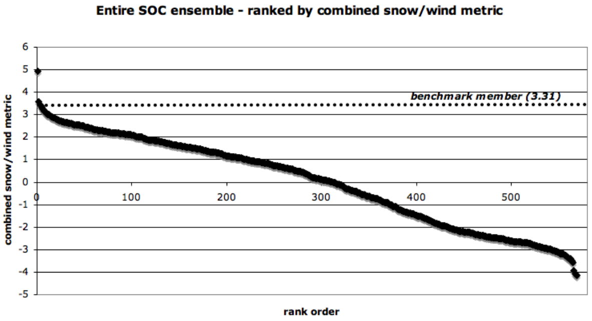

Fig. 9: Members of the entire SOC ensemble, ranked by combined snow/wind metric.

However, Figs. 2-4 show that for several lead times, there are relatively sizable differences between the control and unperturbed members (compare the red squares and open black circles). Curiously, virtually all of the unperturbed members have smaller SLP errors than their respective control runs. One concern is that the ECMWF analysis, coming from a relatively coarse grid, underestimates the actual intensity of the cyclone somewhat; this might result in the selection of a slightly weaker storm as the benchmark. It is seen that unperturbed members consistently generate less frozen precipitation, and smaller NESIS values, which raises questions.

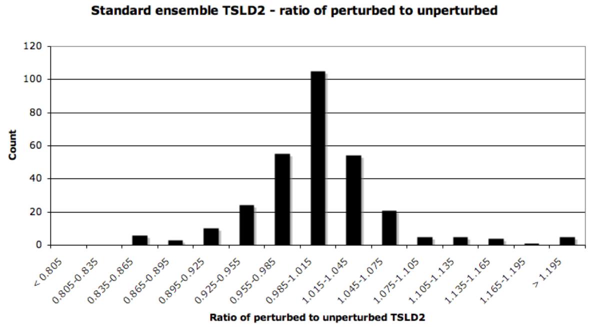

Fig. 10: Distribution of TSLD2 ratios between perturbed and unperturbed members for the standard ensemble (297 members).

A large control-unperturbed gap may be an indication that the lead time is particularly receptive to perturbation, which would be manifested by a relatively wider ensemble spread. Figure 11 presents the 29 simulations made for the 36 h lead, ranked by TSLD2. It is seen that most of the members generated less snow than the control run. However, 63% of the perturbed members (17 or 27) outproduced the *unperturbed run*, which is the most fair comparison in this situation since the simulations shared a common startup procedure. This makes the control-unperturbed gap seem less relevant.

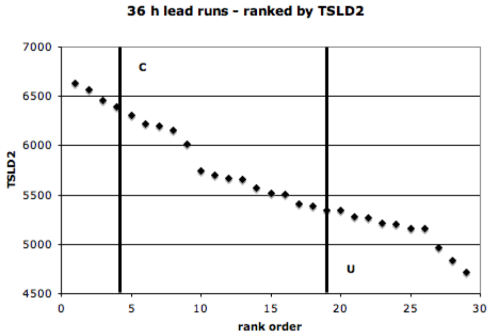

Fig. 11: TSLD2 totals for 36 h lead cases, in rank order. "C" and "U" designate the control and unperturbed cases, respectively.

For the 18 h lead special ensemble, host of many of the most prodigious snow producers, the majority of the perturbed members were in fact *weaker* than the unperturbed run. The distribution shown in Fig. 12 is clearly skewed towards perturbed to unperturbed snow ratios less than unity. In contrast, the distribution for the 42 h lead special ensemble is essentially normal (Fig. 13). This lead time contained many of the entire ensemble's weakest snow generators. Taken together, we believe the perturbation generation strategy has been shown to be adequate and free of systematic bias. The strategy used to generate the perturbations, and the conclusion that it is unbiased, is important to the next logical step, which involves estimating the recurrence interval of SOC-like storms.

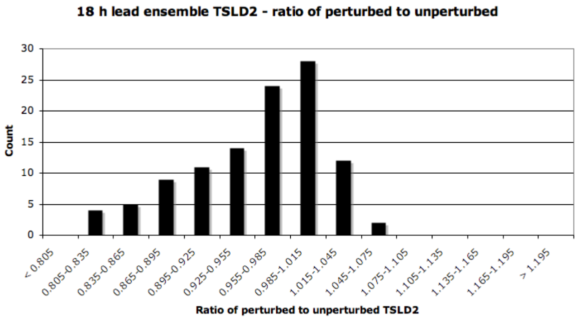

Fig. 12: Distribution of ratios between perturbed and unperturbed members for the 18 h lead ensemble (108 members).

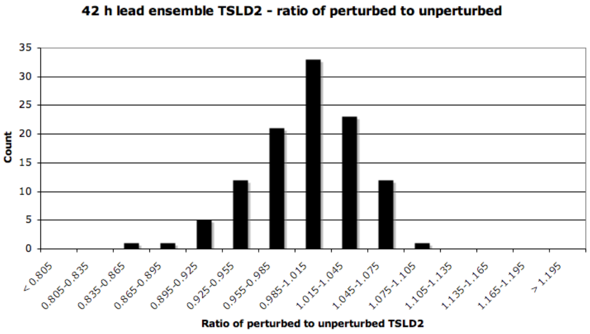

Fig. 13: As in Fig. 13, but for the 42 h lead ensemble.

6. Future work

Present and future work will be conducted in a number of areas. Specifically, we will consider:

- *Performing regional analyses within the SOC ensemble.* Just as there was no single ensemble member that was most intense in every possible characteristic (snow, wind, etc.), no one case can be the top event at every location. We are currently subdividing the ensemble to ascertain the recurrence interval of an SOC-like event on a local and regional basis.

- *Identifying precursors of the extreme events.* An ensemble of this size can serve as a "poor man's adjoint", indicating regions of sensitivity controlling snow production and wind damage. Analysis of the most and least intense ensemble members will identify geographic areas and fields that influenced the evolution of the storm the most.
- *Evaluating different ways of generating initial perturbations.* The perturbations considered herein were applied directly to the temperature and humidity fields, under specified constraints. Other ways of perturbing the initial conditions could be pursued that might prove more efficient at enhancing ensemble spread. We believe that the ratio of within lead time to among lead time variance should be as large as possible, as this justifies the usage of multiple lead times and leads to greater confidence in the fidelity of the ensemble.
- *Considering alternative benchmark identification strategies.* The outcome of this study is sensitively dependent on which ensemble member is selected as the benchmark or truth case. Different strategies for identifying this member should be formulated and evaluated.
- *Examining "underachievers".* The SOC event was one of the most dramatic snowstorms in recorded history, which means the storm made much of its available resources. It would be instructive to identify potential underachievers, or storms that failed to realize their potential. Our working hypothesis would suggest that if a storm could have been stronger, its benchmark case should reside in the weaker portion of the potential storm distribution.

7. Summary

A massive forecast ensemble was created to study the so-called Storm of the Century, the epochal 1993 snowstorm that ranked as the most significant snow-producing event of the 20th century. The ensemble was created by perturbing initial temperature and moisture fields for simulations spanning a set of lead times. The focus was primarily on snow production (including population-weighted measures) and the magnitude and areal extent of potentially damaging winds. A benchmark member was chosen on the basis of an independent, objective criterion involving the root mean square difference between simulated and analyzed sea-level pressure fields.

The working hypothesis is that any given storm is a member of some distribution that may be Gaussian and can be reconstructed through ensemble modeling, realized by perturbing the initial conditions. If an event is about as strong as it could have been, it should reside well above the mean of its own distribution, and perturbations should be more likely to weaken than further intensity the storm. We find that with respect to snow and wind characteristics, the benchmark SOC member fell very near the top of its distribution, implying a relatively long recurrence interval for the event. This conclusion hinges on proper and objective identification of the benchmark case and demonstration that the perturbation strategy is unbiased. Suggestions for future work enhancing and extending this ensemble forecasting strategy were offered.

8. Reference

Bosart, L. F., G. J. Hakim, K. R. Tyle, M. A. Bedrick, W. E. Bracken, M. J. Dickinson, and D. M. Schultz, 1996: Large-scale antecedent conditions associated with the 12-14 March 1993 Cyclone ("Superstorm '93") over Eastern North America. *Mon. Wea. Rev.,* **124**, 1865-1891.

Kocin, P. J., and L. W. Uccellini, 2004: A snowfall impact scale derived from northeast storm snowfall distributions. *Bull. Amer. Meteor. Sci.,* February, 2004, 177-194.

Acknowledgments. This work was supported by AIR-Worldwide and NSF grant ATM-0554765. George Waldenberger assisted in the early stages of ensemble creation and evaluation. Additional computing support was provided by Aerospace Corporation.