

Interactive comment on “ANISORROPIA: the adjoint of the aerosol thermodynamic model ISORROPIA” by S. L. Capps et al.

J. Kelly (Referee)

kelly.james@epa.gov

Received and published: 19 September 2011

Capps et al. describe the development, evaluation, and initial application of ANISORROPIA, the discrete adjoint model of the widely used inorganic aerosol thermodynamic model, ISORROPIA. Given the complexity of the ISORROPIA model, the development of its adjoint is a major undertaking. Up to this point, the lack of an adjoint of ISORROPIA has been a barrier to the development of full adjoints of some 3D photochemical air quality models (AQMs). Adjoint AQMs could potentially provide critical information for a range of important model applications. The ACPD article is a nice presentation of an important first step in the development and application of the adjoint of ISORROPIA. However, some important issues (described below) need to be addressed before I could recommend acceptance of the article into ACP. Revisions should focus on

C9098

better characterizing conditions where ANISORROPIA is unsuccessful or where linear adjoint sensitivities are of little value due to discontinuities and nonlinearities in the solution surface. The potential impact of such limitations on the applications that motivate ANISORROPIA's development should also be discussed.

General Comments:

1. On p. 23482 line 12, the authors indicate that ANISORROPIA converged in 70% of the 450 000 runs conducted. This issue needs to be clarified further. What is the percentage of unsuccessful attempts for standard ISORROPIA for these conditions? Is it correct that the failure percent for ANISORROPIA (30%) is much greater than that for ISORROPIA? If so, then a 3D simulation based on ANISORROPIA would differ from one with standard ISORROPIA. Additional 3D runs would need to be conducted in practical applications to investigate these differences and would reduce the computational efficiency of the approach implied in Section 3.3. Given the significant failure percent, more details should be provided on the conditions where ANISORROPIA is unsuccessful. For instance, are the failures driven entirely by high ionic strengths at low RH or are certain of the 10 solution regimes more susceptible to problems? Is the percentage of failures much greater at lower temperatures than the ones considered? A quantitative breakdown of the percentage of unsuccessful attempts as a function of T, RH, and solution regime could be informative. Also, the reasons that the algorithm fails should be clarified. The authors mention failures of convergence for the bisection method, but the bisection method should always converge with linear convergence rate for a continuous function. Is the failure of bisection due to an unspecified maximum number of iterations being exceeded? The text indicates that the stringent criterion for bisection is driven by the need for good input for the Newton-Raphson (N-R) routine. Since the N-R routine is only called for acidic aerosol, are the failures only occurring for the acidic cases? The authors also mentioned on p. 23482 that the adjoint code returns a flag for appropriate handling by the host model when the model is unsuccessful. However, it is unclear what the host model can do with this flag except count the number of failures (a similar comment applies to the statement on p. 23491, line 18).

C9099

2. The complex variable method (CVM) is implemented to evaluate ANISORROPIA adjoint sensitivities because of challenges arising from the “highly nonlinear, discontinuous solution surface of ISORROPIA”. Although this is a clever approach for avoiding problems with finite difference sensitivities, the complex differentiation formula of Squire and Trapp (1998) applies only to analytic functions, and Capp et al. should explain their use of the approach for a discontinuous solution surface. Along discontinuities in the surface, the adjoint sensitivities do not exist, right? Therefore I do not understand how CVM or ANISORROPIA can provide meaningful results at these locations without implementing an approach for smoothing the solution surface across the discontinuities. Aside from the discontinuities, the highly nonlinear nature of the solution surface calls into question the value of computing and evaluating linear sensitivities. The authors need to provide more details on the ISORROPIA solution surface and potential conditions where nonlinearities would limit the value of linear sensitivities. The sharp gradients between zones in Figure 6b appear to provide such information for one set of conditions, but these results are for a small part of the solution space that would be encountered in a 3D simulation. Perhaps other representative scenarios could be considered and included for reference in a supplement.

3. The motivation for developing ANISORROPIA is to calculate sensitivities that can be used in practical applications (e.g., inverse modeling, emission control strategy development, etc.) where a finite perturbation in a model parameter would generally be made based on the linear adjoint sensitivities. Given the nonlinearities and discontinuities in the ISORROPIA solution surface, it is unclear if successful outcomes would be reached by using linear sensitivities in such applications. At a minimum, much iteration would appear to be necessary in such studies, and this iteration would reduce the overall computational efficiency implied by the comparison of ANISORROPIA and ISORROPIA runtimes. Although the manuscript does not directly address the issue of using the sensitivities in practical applications, some results presented suggest that problems could occur. For instance, the sharp gradients in the adjoint sensitivity surface shown in Figure 6b suggest that perturbations to model parameters based on

C9100

the adjoint sensitivities might produce unintended results in many cases. The results showing differences between finite difference and adjoint sensitivities might also suggest that problems could occur in practical applications. Would it be worthwhile to consider idealized applications of ANISORROPIA where the linear sensitivities were used to perturb a model input parameter and then results were evaluated to determine conditions where the behavior expected based on the linear sensitivities actually occurs? I recognize that a full evaluation of these issues is beyond the scope of a single study, but this article should at least raise the key issues that will need to be addressed in practical applications with ANISORROPIA.

4. The results presented in the ternary diagrams in Figures 6 and 8 appear to disagree with results in Figure 7. In Figure 7, the ammonium-to-sulfuric acid (A-to-S) sensitivity equals 2 in Region I when the molar ratio of total ammonia to total sulfuric acid (R_s) exceeds 2. However, results in the ternary diagrams appear to indicate that the A-to-S sensitivity equals 2 when R_s is less than 2 in many cases. The situation could be rectified if the axes labels for ammonia and nitric acid were switched in the ternary diagrams. However, the discussion accompanying the figures seems to explain the results as they are plotted. For instance, lines 16-17 on p. 23489 appear to suggest that the A-to-S sensitivity equals 2 when the ratio of total nitric acid to sulfuric acid exceeds 2. Also, line 21 on p. 23487 appears to suggest that the A-to-S sensitivity equals 2 under acidic low-ammonia conditions. However, lines 11-12 on p. 23488 indicate the opposite: i.e., the A-to-S sensitivity equals 2 when total ammonia is in excess of sulfuric acid and $R_s > 2$. Yet, line 13 on p. 23488 states that the results in Figures 6 and 7 are consistent. The overall effect of these figures and statements is confusing. I apologize if I am misunderstanding something basic here, but please clarify or fix these sections as necessary.

5. The purpose of discussing inverse modeling in the manuscript is vague since this application is not considered in the study. In the Introduction (p. 23474, lines 16-18), the authors indicate that the utility of ANISORROPIA for conducting inverse modeling

C9101

would be thoroughly discussed in Section 5. However, Section 5 is the Conclusions section and barely mentions inverse methods. The Mathematical Background section (Section 2.1) focuses on describing the cost function used in the 4D-var framework and the adjoint-based evaluation of the gradient of the cost function. However, this cost function and assimilation technique are not applied in the current study, and so the section feels disconnected from other parts of the manuscript. Section 2.1 should be revised to focus directly on the mathematical background of the approaches used in applications in the current study or on better relating the 4D-var example to the approaches considered in the current study.

Specific Comments:

p. 23471, lines 1-2: It is unclear why the aqueous phase is specified here since the observations were apparently for dry particles. A similar comment applies to p. 23477 lines 6-8.

p. 23474, line 9-10: Is there a better alternative to “deliquescence curve”? Perhaps “water uptake curve”? Also, it would more accurate to state that CTMs neglect the formation of solid phases rather than stating that they treat particles as metastable solutions, because particle solutions are stable at high RH.

p. 23474, lines 23-24: By “produce a field of requisite perturbations in input parameters” do you mean something like “to yield sensitivities of model output variables to model input parameters that can be used to perturb model inputs in certain applications”?

p. 23477, lines 10-11: I think this could be better phrased as “CTMs often treat fine mode particles as existing in thermodynamic equilibrium with the surrounding gases based on the short equilibration time scales for these particles.” The bulk equilibrium calculations in CTMs limit their ability to represent size-dependent particle composition variations, and longer equilibration time scales than 20 min are often reported for sub-2.5 micron particles. Rather than justifying the typical treatment of aerosols in CTMs

C9102

here, the approach used in CTMs and reasoning behind that approach could just be stated.

p. 23478, lines 1-2: This seems like a strong statement. Is it true that K-M is more accurate than the Pitzer model? The studies by Kim et al. (1993a,b) actually state that the Pitzer method is superior up to $I = 20m$, and the Kim et al. studies do not consider updates by Clegg et al. (1992a,b).

p. 23479, lines 10 and 18: Please clarify “formed by hand” and “checkpoint”

p. 23480, lines 21-23: Has this intercomparison been performed? Can it be included here?

p. 23481, line 17: Is “post-convergence Newton-Raphson method” supposed to indicate that bisection output is being used as the initial guess for N-R?

p. 23482, line 7: Why not consider lower temperatures since they will frequently be encountered in 3D simulations?

p. 23483, lines 20-21: Can you reword for clarity?

p. 23484, equation (5): Since this approach could be used in the future, it might be worthwhile to state that (5) arises from setting the total derivative at the solution point to zero: $df = \frac{\partial f}{\partial x}|_{x=x_0} dx + \sum_{i=1}^3 \frac{\partial f}{\partial A_i}|_{x=x_0} dA_i = 0$ where $dx = Im(x_0)$ and $dA_i = Im(A_i)$ in this case.

p. 23484, line 20: Is the performance for results at 268-288 K similar to that at 288-308 K? These results could be included in a supplement for future reference for users who might be interested in low-T applications.

p. 23485, line 11: Does use of the same convergence criteria imply that the same number of unsuccessful attempts occurred for both models, and so failures did not impact timing results?

p. 23488, line 3-4: Can you provide a better explanation for the negative sensitivities

C9103

at low T?

Figure 7: The parts of the figure where the regions overlap may generate confusion. I would recommend clarifying these regions in the figure or text. Also, the figure legend refers to the “increasing gradient from left to right” but the vertical gradient is more clearly apparent.

p. 23488, line 21-22: As mentioned above, I think a supplement with figures for alternative scenarios could be a handy reference for future studies that apply ANISORROPIA to a wide range of conditions.

p. 23489, lines 13-14: Nowak et al. (2006) concluded that “the assumption of thermodynamic equilibrium was reasonable”; however, for model-observation comparisons, they reported a slope of 1.25, which can be considered high for a constrained comparison. The characterization in the current study that Nowak et al. “demonstrated the accuracy of ISORROPIA in predicting the partitioned concentrations” appears to be stretching the conclusions of that study.

p. 23491, line 5: Here it is suggested that precipitation of solid phases have a minor impact on results, but the previous page (p. 23490, line 23) suggests precipitation of salts has a significant impact on results. Please clarify.

p. 23491, lines 11-13. It is hard to understand the differences in results because significantly different versions of ISORROPIA are being compared. Are the results the same when the Fountoukis et al. (2009) finite difference calculations are repeated with ISORROPIA (rather than ISORROPIA-II)?

p. 23491, lines 17: Do results with unrealistic sensitivities imply that the success percentage of 70% reported here is an upper estimate (i.e., in addition to cases that do not converge, results from other challenging cases may not be usable)?

p. 23491, lines 18-21: Do the differences in sensitivities for the adjoint, backward difference, and central difference approaches imply that making finite perturbations in

C9104

model input parameters based on adjoint sensitivities will produce unintended behavior in practical applications?

References:

Clegg, S. L., and K. S. Pitzer: Thermodynamics of multicomponent, miscible, ionic-solutions - Generalized equations for symmetrical electrolytes. *J. Phys. Chem.* 96, 3513-3520, 1992a.

Clegg, S. L., K. S. Pitzer, and P. Brimblecombe: Thermodynamics of multicomponent, miscible, ionic solutions. II. Mixtures including unsymmetrical electrolytes. *J. Phys. Chem.* 96, 9470-9479, 1992b.

Kim, Y. P., Seinfeld, J. H., and Saxena, P.: Atmospheric Gas-Aerosol Equilibrium I. Thermodynamic Model, *Aerosol Sci. Tech.*, 19, 157–181, 1993a.

Kim, Y. P., Seinfeld, J. H., and Saxena, P.: Atmospheric Gas-Aerosol Equilibrium II. Analysis of Common Approximations and Activity Coefficient Calculation Methods, *Aerosol Sci. Tech.*, 19, 182–198, 1993b.

Fountoukis, C., Nenes, A., Sullivan, A., Weber, R., Van Reken, T., Fischer, M., Matas, E., Moya, M., Farmer, D., and Cohen, R. C.: Thermodynamic characterization of Mexico City aerosol during MILAGRO 2006, *Atmos. Chem. Phys.*, 9, 2141–2156, doi:10.5194/acp-9-2141-2009, 2009.

Nowak, J., Huey, L., Russell, A., Tian, D., Neuman, J., Orsini, D., Sjostedt, S., Sullivan, A., Tanner, D., Weber, R., Nenes, A., Edgerton, E., and Fehsenfeld, F.: Analysis of urban gas phase ammonia measurements from the 2002 Atlanta Aerosol Nucleation and Real-Time Characterization Experiment (ANARChE), *J. Geophys. Res.*, 111, D17308, doi:10.1029/2006JD007113, 2006.

Squire, W. and Trapp, G.: Using complex variables to estimate derivatives of real functions, *SIAM Rev.*, 40, 110–112, 1998.

C9105

C9106