DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY

Application of empirical orthogonal functions to evaluate ozone simulations with regional and global models

Citation

Fiore, Arlene M., Daniel J. Jacob, Rohit Mathur, and Randall V. Martin. 2003. "Application of Empirical Orthogonal Functions to Evaluate Ozone Simulations with Regional and Global Models." Journal of Geophysical Research 108, issue D14: ACH 10-1-ACH 10-8.

Published version

https://doi.org/10.1029/2002JD003151

Link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:14117799

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at
https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFlOWI2M2EwYTg

Accessibility

https://accessibility.huit.harvard.edu/digital-accessibility-policy

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story

Application of empirical orthogonal functions to evaluate ozone simulations with regional and global models

Arlene M. Fiore and Daniel J. Jacob
Department of Earth and Planetary Sciences and Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA

Rohit Mathur
Environmental Modeling Center, MCNC, Research Triangle Park, North Carolina, USA

Randall V. Martin
Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts, USA

Received 7 November 2002; revised 11 February 2003; accepted 24 March 2003; published 31 July 2003.

[1] Empirical orthogonal functions are used together with standard statistical metrics to evaluate the ability of models with different spatial resolutions to reproduce observed patterns of surface ozone (O_3) in the eastern United States in the summer of 1995. We examine simulations with the regional Multiscale Air Quality Simulation Platform model (horizontal resolution of 36 km^2) and the global GEOS-CHEM model ($2° \times 2.5°$ and $4° \times 5°$). As the model resolution coarsens, the ability to resolve local O_3 maxima ($O_3 \geq 90$ ppbv) is compromised, but the spatial correlation improves. This result shows that synoptic-scale processes modulating O_3 concentrations are easier to capture in models than processes occurring on smaller scales. Empirical orthogonal functions (EOFs) derived from the observed O_3 fields reveal similar modes of variability when averaged onto the three model horizontal resolutions. The EOFs appear to represent (1) an east-west pattern associated with frontal passages, (2) a midwest-northeast pattern associated with migratory high-pressure systems, and (3) a southeast stagnation pattern linked to westward extension of the Bermuda High. All models capture the east-west and southeast EOFs, but the midwest-northeast EOF is misplaced in GEOS-CHEM. GEOS-CHEM captures the principal components of the observational EOFs when the model fields are projected onto these EOFs, implying that it can resolve the contribution of the EOFs to the observed variance. We conclude that coarse-resolution global models can successfully simulate the synoptic conditions leading to high-O_3 episodes in the eastern United States. *INDEX TERMS:* 0345 Atmospheric Composition and Structure: Pollution—urban and regional (0305); 0368 Atmospheric Composition and Structure: Troposphere—constituent transport and chemistry; 0394 Atmospheric Composition and Structure: Instruments and techniques; 3364 Meteorology and Atmospheric Dynamics: Synoptic-scale meteorology; *KEYWORDS:* empirical orthogonal functions, principal components, air pollution, ozone, regional modeling, global modeling

Citation: Fiore, A. M., D. J. Jacob, R. Mathur, and R. V. Martin, Application of empirical orthogonal functions to evaluate ozone simulations with regional and global models, *J. Geophys. Res.*, *108*(D14), 4431, doi:10.1029/2002JD003151, 2003.

1. Introduction

[2] The design of policies to reduce photochemical O_3 smog relies upon the ability of air quality models to simulate accurately the meteorological and chemical processes governing O_3 production. Current regional air quality models can reproduce seasonal surface O_3 statistics much better than specific events [*Hogrefe et al.*, 2000; *Kasibhatla and Chameides*, 2000; *Sistla et al.*, 2001]. These studies call into question the traditional air quality planning approach of using a given historical episode to examine the benefits of emission reductions [*Hogrefe et al.*, 2001b]. Calculation of seasonal O_3 statistics requires longer simulations, making coarsely resolved models appealing from a computational standpoint. There is also increasing interest in using global models for regional air quality applications to examine the intercontinental transport of pollution [*Yienger et al.*, 2000; *Wild and Akimoto*, 2001; *Fiore et al.*, 2002a, 2002b; *Li et al.*, 2002] and to eliminate dependence on arbitrary boundary conditions. In the present study, we examine the ability of a global model at two different resolutions to simulate regional O_3 pollution over the United States, as compared to a regional model, through an analysis of a number of observational statistics including space-time patterns of pollution episodes.

Copyright 2003 by the American Geophysical Union.
0148-0227/03/2002JD003151

[3] The computational cost of an O_3 smog model is largely determined by its horizontal resolution and the effort associated with numerically solving the stiff ordinary differential equations that describe the chemical transformations at each grid square. Previous studies examining model sensitivity to horizontal resolution have focused on the urban-to-regional transition. *Sillman et al.* [1990] argued that degrading the horizontal resolution should cause a systematic positive bias in the O_3 simulation, because dilution of precursor NO_x emissions increases the O_3 production efficiency per unit NO_x oxidized [*Liu et al.*, 1987]. A simulation by *Kumar et al.* [1994] for southern California found that degrading horizontal resolution from 5×5 km^2 to 10×10 km^2 to 20×20 km^2 causes overestimates of O_3 concentrations in the urban core. *Liang and Jacobson* [2000] have shown that the bias in coarse-resolution O_3 simulations can in fact go either way, since the numerical dispersion of pollution plumes in low-NO_x regions of a coarse-grid model decreases the O_3 production efficiency in the latter regions. Multiscale simulations with nested grid models can help to reduce this bias [*Kumar et al.*, 1994; *Kumar and Russell*, 1996].

[4] Our analysis uses surface O_3 fields measured at the Environmental Protection Agency's Aerometric Information Retrieval System (EPA AIRS) monitoring sites in the eastern United States for the summer of 1995 to compare and evaluate simulations of O_3 pollution from (1) the regional Multiscale Air Quality Simulation Platform (MAQSIP) model (36 km^2 resolution) [*Kasibhatla and Chameides*, 2000; *Hogrefe et al.*, 2001a, 2001b], (2) the global GEOS-CHEM model at $2° \times 2.5°$ resolution, and (3) GEOS-CHEM at $4° \times 5°$ resolution. We have previously used the GEOS-CHEM model to examine intercontinental transport of O_3 pollution and the origin of the O_3 background in the United States [*Fiore et al.*, 2002a, 2002b; *Li et al.*, 2002]. *Fiore et al.* [2002a] showed that GEOS-CHEM captures regional high-O_3 episodes in the eastern United States during the summer of 1995 and reproduces the "piston effect" (the highest O_3 values exhibit the largest response to reductions in U.S. anthropogenic emissions from 1980 to 1995) [*Lefohn et al.*, 1998]. We analyze here standard statistical metrics and empirical orthogonal functions (EOFs) to assess the ability of GEOS-CHEM to simulate the observed regional structure of O_3 pollution episodes in the eastern United States as compared to the regional MAQSIP model.

2. Model Description

2.1. MAQSIP Model

[5] The MAQSIP [*Odman and Ingram*, 1996] regional photochemical model is driven by meteorological fields from the MM5 model and is similar to the CMAQ (Community Multi-scale Air Quality) modeling system [*Byun and Ching*, 1999]. The MAQSIP simulation analyzed here covers the eastern United States (Figure 1a), has 36×36 km^2 resolution, and extends over June–August 1995 [*Kasibhatla and Chameides*, 2000; *Hogrefe et al.*, 2001a, 2001b]. There are 22 vertical sigma levels up to 100 hPa, with a surface layer 38 m deep and 11 layers below 1500 m. Boundary conditions for O_3 are 35 ppbv everywhere. Ozone chemistry is simulated with the Carbon Bond Mechanism [*Gery et al.*, 1989; *Kasibhatla et al.*, 1997]. The emission inventory is described by *Houyoux et al.* [2000]. Biogenic emissions are from the Biogenic Emission Inventory System 2 (BEIS2) [*Geron et al.*, 1994]. The simulation for the summer of 1995 has been evaluated previously with surface O_3 observations from the AIRS network [*Kasibhatla and Chameides*, 2000; *Hogrefe et al.*, 2001b]. *Hogrefe et al.* [2001a, 2001b] found that while seasonal mean concentrations are captured well, the variance is underestimated with the largest bias in the southern part of the domain. They found that the meteorological processes in MM5 are well represented at synoptic timescales but not at the shorter diurnal and intraday timescales, so that O_3 cannot be simulated properly at the latter scales.

2.2. GEOS-CHEM Model

[6] The GEOS-CHEM global model of tropospheric chemistry [*Bey et al.*, 2001] uses assimilated meteorological data from the NASA Goddard Earth Observing System (GEOS). The simulation for June–August 1995 presented here was described previously by *Fiore et al.* [2002a]. It uses GEOS-CHEM version 3.3 (http://www-as.harvard.edu/chemistry/trop/geos). There are 20 vertical sigma layers, with five layers located below 2 km and a surface layer of roughly 100 m depth. Simulations were conducted with both $2° \times 2.5°$ and $4° \times 5°$ horizontal resolution. Anthropogenic emissions for summer 1995 in the eastern United States are from the Southern Appalachian Mountain Initiative (SAMI) inventory [*Ogburn et al.*, 2000; E. H. Pechan and Associates, Inc., Development of SAMI 1990 base year emissions inventory, draft report prepared for Southern Appalachian Mountains Initiative, January 1999]. The chemical mechanism is described by *Bey et al.* [2001] and builds upon the heritage from *Lurmann et al.* [1986].

[7] *Fiore et al.* [2002a] presented an extensive evaluation of the $2° \times 2.5°$ GEOS-CHEM simulation with observations for O_3 and precursor species over the United States. In general, the model was found to capture the observed O_3 frequency distributions and the relationships between O_3 and its precursors with no evident systematic bias. The model was shown to respond to fossil fuel emission changes from 1980 to 1995 in a manner consistent with observed trends. Weaknesses included (1) excessive convective mixing over the Gulf of Mexico and the Caribbean, causing the model to overestimate background O_3 over the southeastern states, and (2) poor resolution of the gradient in mixed layer depths between land and ocean, resulting in an overestimate of O_3 in coastal urban grid boxes.

3. Ozone Data for the Summer of 1995

[8] Hourly records of O_3 measurements for June–August 1995 between 1300 and 1700 local time (LT) were obtained from 650 EPA AIRS sites over the eastern United States. We focus on observations during the afternoon hours, which are representative of a relatively deep mixed layer and are thus best suited for model evaluation [*Fiore et al.*, 2002a]. For comparison purposes, we average all available AIRS observations over the respective horizontal grids of the GEOS-CHEM and MAQSIP models, and obtain one average observed value per grid square per afternoon. For comparison with GEOS-CHEM, the AIRS data were aver-

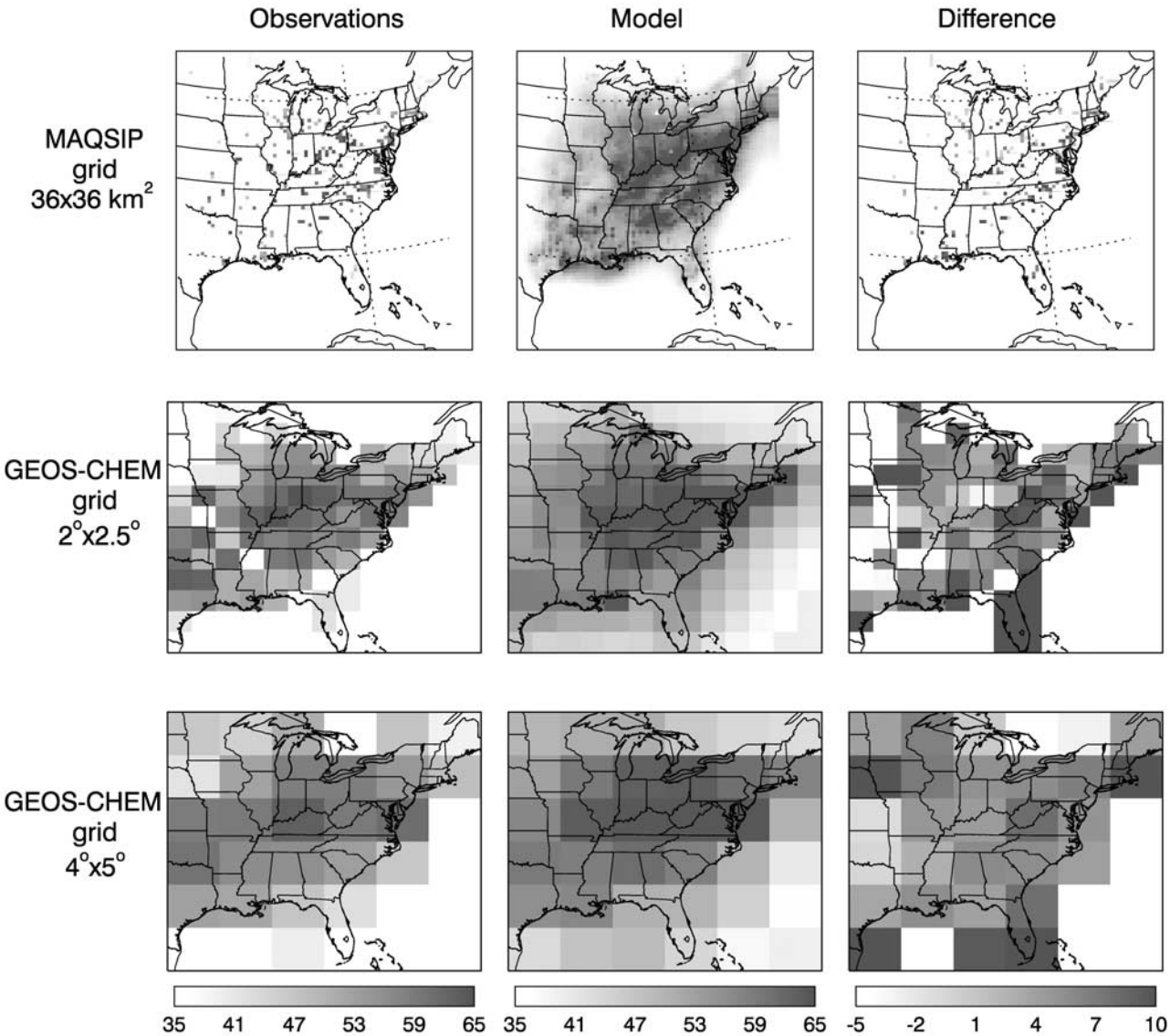

Figure 1. Mean afternoon (1300–1700 LT) surface O_3 concentrations (ppbv) in June–August 1995. (left column) AIRS observations averaged onto the MAQSIP (36 × 36 km^2) and GEOS-CHEM (2° × 2.5° and 4° × 5°) model grids. (middle column) Simulated fields from the MAQSIP and GEOS-CHEM models. (right column) Difference between observed and simulated fields (model bias). The color scale is designed to resolve regional features rather than extreme concentrations and saturates. The range of observed values is 34–73 (MAQSIP), 31–66 (GEOS-CHEM 2° × 2.5°), and 33–63 (GEOS-CHEM 4° × 5°). The range of simulated values is 17–103 (MAQSIP), 34–74 (GEOS-CHEM 2° × 2.5°), and 32–67 (GEOS-CHEM 4° × 5°). The maximum differences are −28 and +43 ppbv (MAQSIP), −6 and +26 ppbv (GEOS-CHEM 2° × 2.5°), and −3 and +12 ppbv (GEOS-CHEM 4° × 5°). See color version of this figure at back of this issue.

aged first on a 0.5° × 0.5° grid in order to remove some of the location bias in the observations resulting from the clustering of monitors in urban locations and then averaged onto the GEOS-CHEM grid. Only grid squares that contain data for every afternoon of summer 1995 are considered for further analysis.

[9] The left column of Figure 1 shows the mean observed O_3 concentrations. Values are highest in the industrial Midwest and along the mid Atlantic seaboard. Concentrations are lowest along the Gulf of Mexico, the upper Midwest, and northern New England. Further discussion of the features of the O_3 observations is presented below in the context of comparison to model results.

4. Model Versus Observations: General Statistics

[10] Figure 1 compares simulated and observed mean O_3 concentrations for the MAQSIP, GEOS-CHEM 2° × 2.5°, and GEOS-CHEM 4° × 5° models. The spatial statistics for the fields in Figure 1 are presented in Table 1. The overestimates in coastal urban grid boxes in GEOS-CHEM were previously discussed (section 2). All three models are too

Table 1. Spatial Statistics for Mean Afternoon O_3 in Surface Air Over the Eastern United States in Summer 1995[a]

	MAQSIP grid 36 × 36 km^2 (n = 263)		GEOS-CHEM grid 2° × 2.5° (n = 80)		GEOS-CHEM grid 4° × 5° (n = 28)	
Statistic	Model	Observed	Model	Observed	Model	Observed
Mean, ppbv	60	56	59	52	56	51
Variance, ppbv^2	89	62	42	66	55	69
Model bias ±1σ, ppbv	4 ± 8		6 ± 6		5 ± 4	
Correlation coefficient r[b]	0.59		0.70		0.90	
Slope of regression line[c]	1.20		0.80		0.88	

[a]Statistics are calculated from the mean June–August 1995 afternoon (1300–1700 LT) O_3 concentrations shown in Figure 1. The observations are averaged over the corresponding model grid. Only grid squares where observations are available for every summer day (June–August 1995) are used for the statistics.

[b]Correlation coefficient of simulated versus observed concentrations for the data in Figure 1.

[c]Slope a of the regression line $[O_3]_{model} = a[O_3]_{obs} + b$, calculated with the reduced major axis method [*Hirsch and Gilroy*, 1984] for the data in Figure 1.

low over the south-central United States and too high in the southeast. The higher correlation between model and observed O_3 fields as horizontal resolution decreases reflects a smoothing of the data by averaging over the coarser grid and does not imply that the coarser models offer a more accurate simulation [*Fiore et al.*, 2002a]. It does imply that the models fare better at capturing the larger-scale features of the O_3 distribution. Spatial averaging removes the smaller scale variability from the observations and the model is then able to capture a larger percentage of this reduced variance.

[11] The correlations between simulated and observed daily summer afternoon O_3 time series for each grid cell in the MAQSIP and GEOS-CHEM models are displayed in Figure 2. The correlation coefficients (r) range from 0.13 to 0.86 for MAQSIP, from 0.16 to 0.85 for GEOS-CHEM 2° × 2.5°, and from 0.33 to 0.81 for GEOS-CHEM 4° × 5°. The highest correlations in MAQSIP are found along the eastern seaboard, as previously discussed by *Hogrefe et al.* [2001b]. The GEOS-CHEM model has the highest correlations with observations in the inland eastern states because it does not simulate well the coastal transition. The strong correlation in the southeastern region shows that the processes modulating O_3 concentrations are captured by GEOS-CHEM, despite the positive bias discussed above. The correlation for the MAQSIP model is low in the industrial midwest (southern Ohio and Indiana). This issue has been extensively analyzed by *Hogrefe* [2000] without reaching a conclusive explanation. As a check on whether the higher resolution MAQSIP model yields a better prediction of the larger-scale features, we averaged the 36 km^2 MAQSIP O_3 simulation onto both the 2° × 2.5° and 4° × 5° GEOS-CHEM grids. We then correlated the daily afternoon O_3 time series for these coarse-resolution MAQSIP O_3 concentrations with the AIRS data averaged onto the same coarse grids. The correlation does improve (0.27–0.87 for the 2° × 2.5° grid; 0.54–0.87 for the 4° × 5° grid), but the distribution of high correlations near the coast and low correlations in the industrial midwest remains similar to that of the 36 km^2 MAQSIP simulation.

5. Model Versus Observations: Empirical Orthogonal Functions (EOFs)

5.1. Method

[12] In this section we further evaluate model results with a principal component analysis (PCA) that decomposes the O_3 fields over the eastern United States into empirical orthogonal functions (EOFs) characterizing space-time patterns of variability [*Yarnal*, 1993]. A general description of the application of PCA to geophysical fields is given by

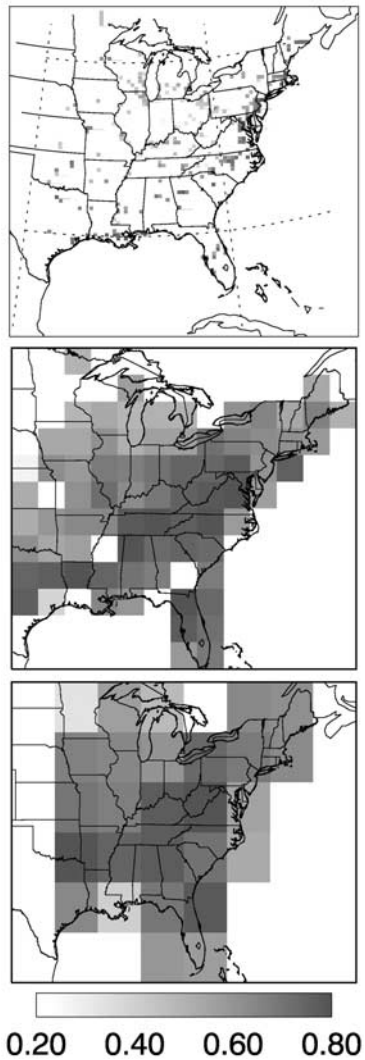

Figure 2. Correlation coefficients (r) between the simulated and observed daily afternoon (1300–1700 LT) mean surface O_3 concentrations during the summer of 1995. The color bar is saturated at 0.2 and 0.8. See color version of this figure at back of this issue.

Wilks [1995]. We provide a brief summary here with an explanation of our particular application to surface O_3 concentrations.

[13] PCA is typically conducted on the "anomalies" (the difference from some mean value) of geophysical fields. In our case, we remove the summer mean O_3 concentration field defined by the mean values in each grid square. We then construct a correlation matrix of elements r_{ij} representing the correlation of anomalies between grid square i and grid square j over the 3-month record. The EOFs (also referred to as "principal component loadings") are the eigenvectors of the correlation matrix [*Wilks*, 1995] and represent independent modes of variability. The eigenvalues give the amount of variance explained by each eigenvector. The sum of all eigenvalues equals the total variance in the original data. The first EOF is defined as that accounting for the maximum amount of variance in the data set. Successive EOFs explain the maximum variance not accounted for by preceding EOFs. Much of the variance can often be represented with only the first few EOFs, making PCA a useful data reduction technique. Projection of the original daily afternoon O_3 concentration fields onto the EOFs defines the principal components (PCs) [*Wilks*, 1995] (also referred to as principal component scores); these time series describe the temporal variation in the contribution of the associated EOF to the total variance in the original data.

[14] *Eder et al.* [1993] previously applied PCA to daily maximum surface O_3 concentrations over the eastern United States from April through October 1985–1990 in order to isolate regions with statistically unique O_3 characteristics. They subjected the first six EOFs (which explained 64% of the total variance) to an orthogonal Varimax rotation in order to better segregate the areas with similar O_3 concentration characteristics. They found that two major patterns of anticyclonic activity modulate O_3 concentrations over the eastern United States: (1) westward extension of the Bermuda High, creating stagnant conditions that often persist for 3–4 days over the southeastern and mid Atlantic states, and (2) migratory anticyclones originating in Canada and traversing the northeastern United States, causing fluctuations in O_3 concentrations on a shorter timescale of 1–2 days. Other studies have also underscored the importance of synoptic-scale weather systems, such as frontal passages or quasi-stationary high-pressure systems, in contributing to the observed variability of O_3 concentrations over the eastern United States [*Logan*, 1989; *Vukovich*, 1995, 1997; *Cooper and Moody*, 2000].

5.2. Application

[15] We follow the general approach of *Eder et al.* [1993] and extract the EOFs and PCs from the AIRS observations and model fields for June–August 1995 discussed in the previous section. We find that our first three Varimax-rotated EOFs together usually account for over 50% of the variance; the EOFs beyond the third individually account for less than 10% of the total variance and are excluded from further consideration based upon a scree graph analysis [*Wilks*, 1995]. The results are shown in Figures 3a–3c for the three model grids: the left column shows the EOFs derived from aggregating the AIRS data onto each model grid and the center column shows the EOFs derived from the O_3 concentrations simulated by each model. The fraction of variance explained by the first three EOFs increases as the spatial resolution decreases, both in the observations and in the models, presumably because the spatial averaging removes finer-scale, less regionally coherent contributions to the O_3 variability.

[16] The first EOF identifies a longitudinal gradient as a major factor of variability for O_3 in the eastern United States (east-west EOF, top left panel of Figure 3a). When the principal component time series (solid line, top right panel) peaks, this east-west EOF is strongly expressed, and O_3 concentrations are elevated along the east coast (where the EOF is shown in red) and are low in the central part of the country (where the EOF is shown in blue). The pattern is reversed when the time series is minimum, with high O_3 concentrations over the central states and lower concentrations along the east coast. This east-west EOF appears to reflect the passage of cold fronts traversing the region along the EOF gradient from west to east; examination of daily surface weather maps from the summer of 1995 [*National Weather Service*, 1995] confirms that cold fronts were present on days when the PC time series peaks and that stagnant conditions (conducive to O_3 buildup) prevailed over the central states on days corresponding to troughs in the PC time series. *Vukovich* [1995, 1997] and *Cooper and Moody* [2000] previously emphasized the role of frontal systems in determining O_3 variability over the eastern United States.

[17] The second EOF highlights the industrial Midwest and the Northeast (midwest-northeast EOF, middle left panel of Figure 3a) as a coherent region. The corresponding PC (solid line, middle right panel of Figure 3a) peaks on 18 June, 12–13 July, and 1 August, corresponding to three major O_3 episodes over the mid Atlantic states in the summer of 1995 [*Ryan et al.*, 1998]. The troughs of this time series correspond to clean conditions in the midwest-northeast region. Inspection of daily surface weather maps [*National Weather Service*, 1995] suggests that airflow was from the northwest on these days; this result is consistent with previous findings that the lowest O_3 concentrations in the northeast are associated with clean air from northern Canada [*Logan*, 1989].

[18] The pattern of our midwest-northeast EOF suggests that a pollution transport pathway links the Midwest to the Northeast, causing these regions to vary together. *Eder et al.* [1993] found that O_3 variability in this region is governed by migratory high-pressure systems that originate in Canada and typically spend less than 2 days over the region. Further evidence for this hypothesis is provided by *Moody et al.* [1998] who found from trajectory cluster analysis that high summertime O_3 concentrations over central Massachusetts are highly correlated with transport from the west (more so than with transport from the southwest). This westerly transport pathway is characterized by sunny conditions and moderate wind speeds, conducive to O_3 production and accumulation [*Moody et al.*, 1998]. A similar conclusion was reached by *Sistla et al.* [2001], who summarize an analysis of meteorological fields from the summer of 1995 [*Sonoma Technology Inc.*, 1999].

[19] The third EOF (southeast EOF, bottom left panel of Figure 3a) isolates the southeastern United States as a region that varies coherently. *Logan* [1989] previously showed that there is little correlation between O_3 concen-

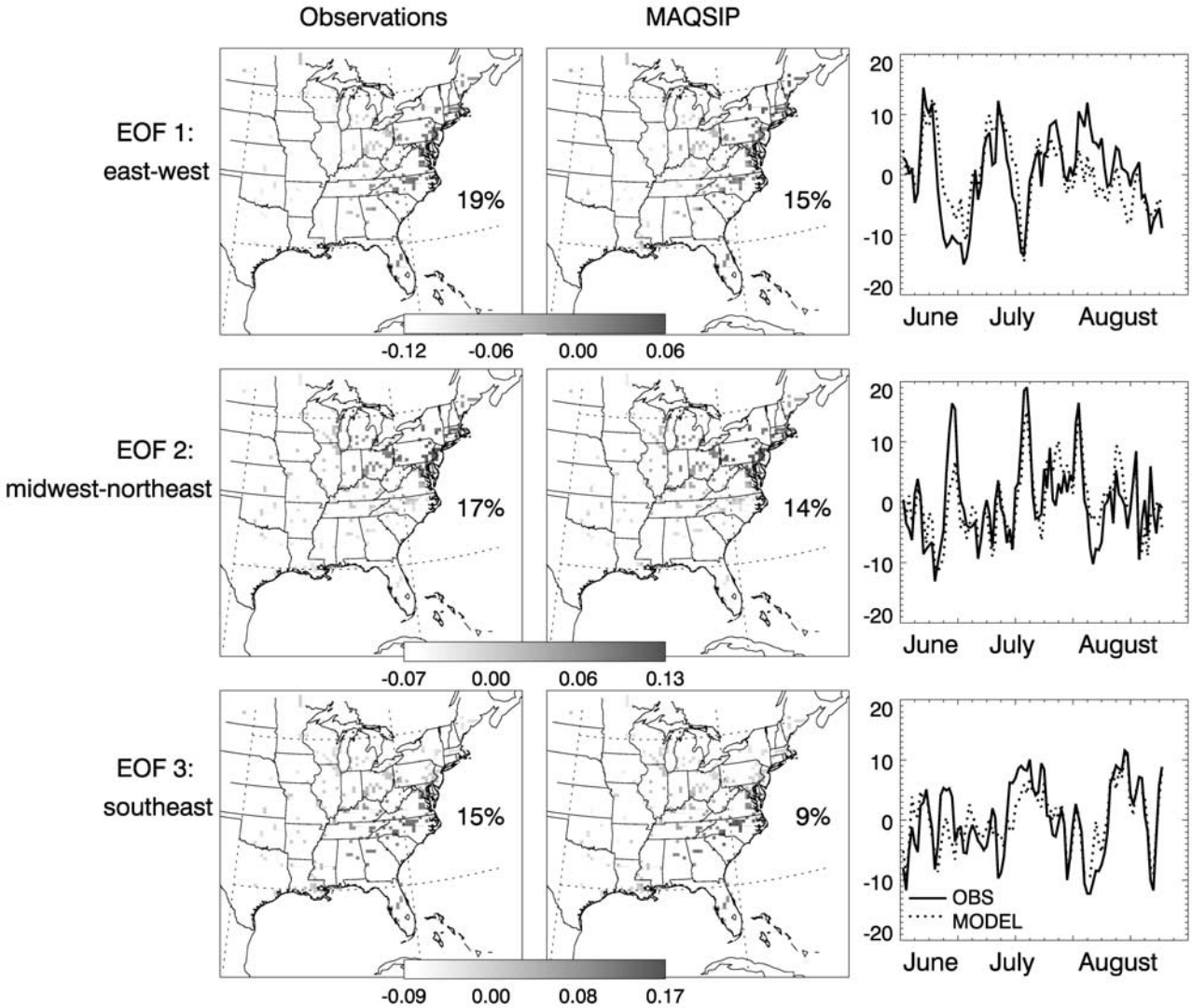

Figure 3a. (left and middle columns) The first three Varimax-rotated EOFs and (right columns) associated PCs as derived from daily afternoon average (1300–1700 LT) surface O_3 concentrations for the summer of 1995 (left) in the observations averaged on the MAQSIP model grid and (middle) in the MAQSIP model. The percentage of total variance explained by the individual EOFs is shown. The right column shows the PCs (temporal variation in the strength of the EOFs, solid lines) derived from the observations and the corresponding time series (dotted lines) obtained from projecting the simulated daily summer afternoon O_3 concentrations onto the (left column) observationally derived EOFs. See color version of this figure at back of this issue.

trations in the southeast and in the northeast and attributed this result to the different processes responsible for producing low O_3 concentrations in the two regions: the influence of clean air from Canada in the northeast versus clean air from the Atlantic in the southeast. The mid July peak in the southeast PC time series (solid line, bottom right panel of Figure 3a) corresponds to a documented high-O_3 event over the southeast region [*Olszyna et al.*, 1998], suggesting that the southeast EOF is most likely associated with regional stagnation. Indeed, the peaks in the southeast PC on 15 and 19 July and 17, 18, and 30 August are all associated with relatively weak high pressures over the region, while the troughs coincide with rain according to daily weather maps [*National Weather Service*, 1995]. Our findings are consistent with those of *Eder et al.* [1993], who postulated that variability in O_3 concentrations in this region is controlled by stagnant conditions associated with the westward extension of the quasi-permanent Bermuda High pressure.

[20] The EOFs from the AIRS observations at $2° \times 2.5°$ and $4° \times 5°$ resolution (left columns of Figures 3b and 3c) are similar to those at the 36×36 km^2 resolution, but the midwest-northeast and southeast EOFs are switched in order of importance at the $4° \times 5°$ resolution (Figures 3b and 3c). We attribute this switch to the larger scales of motion responsible for the southeast EOF than for the midwest-northeast EOF; the latter tend to be smoothed out over the coarse grid. Examination of the PCs for the AIRS observations averaged over both resolutions of the GEOS-CHEM grid (Figures 3b and 3c) reveals a similar temporal pattern as over the MAQSIP grid, but the ampli-

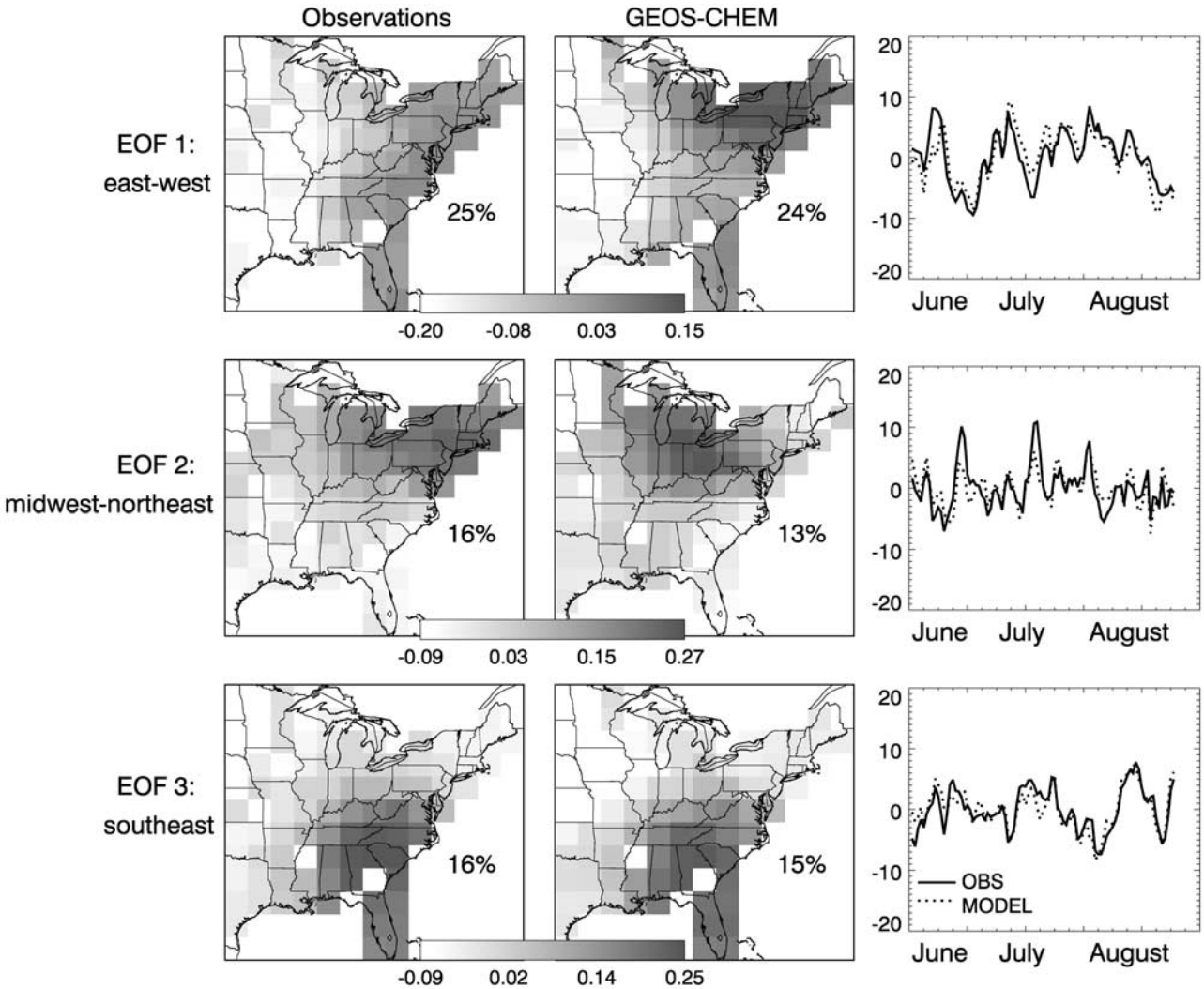

Figure 3b. Same as Figure 3a but for the 2° × 2.5° resolution of GEOS-CHEM. See color version of this figure at back of this issue.

tude of the PCs is dampened as resolution increases. The similar features in the EOFs and PCs derived from the observations at all horizontal resolutions indicate that the same fundamental processes are responsible for the observed variability of O_3 at all scales.

[21] We now turn to the simulations of the EOFs by the models. Both the MAQSIP and GEOS-CHEM models capture the longitudinal gradient of the east-west EOF (upper middle panels of Figure 3a), but the GEOS-CHEM model additionally highlights the Northeast within this EOF. It is possible that the GEOS-CHEM model is mixing together the first two observationally derived EOFs, placing a northeast enhancement on top of the longitudinal gradient. The MAQSIP model successfully captures the midwest-northeast EOF, but the GEOS-CHEM model displaces this EOF to the west. The southeast EOF is reproduced at all model resolutions (bottom middle panels of Figures 3a–3c).

[22] We correlate the simulated and observed EOFs to evaluate how well the models capture the spatiotemporal patterns of O_3 variability. Correlation statistics are given in Table 2; a slope of unity accompanied by a high coefficient of determination (r^2) signals a good match. The slope reveals whether the model exhibits any bias in its representation of the underlying processes, whereas the r^2 gives the amount of variance in the observationally derived EOFs that is captured by the model. The MAQSIP model reproduces well all three EOFs (r^2 of 0.76–0.86, slopes of 1.0). The GEOS-CHEM model at 2° × 2.5° resolution captures the east-west and southeast EOFs, but the resolution of the midwest-northeast EOF is degraded. In the 4° × 5° model, the east-west EOF is degraded, and the midwest-northeast EOF is missed completely; the southeast EOF is still well reproduced.

[23] We examined the ability of the models to reproduce the time series describing the strength of each EOF on every summer day (1 June through 30 August). The time series shown as dotted lines in the right column of Figures 3a–3c were obtained by projecting the simulated O_3 fields onto the observationally derived EOFs. This approach avoids introducing the model bias from the simulated EOFs into this evaluation. The corresponding statistics are shown in Table 2. MAQSIP reproduces the variability in the

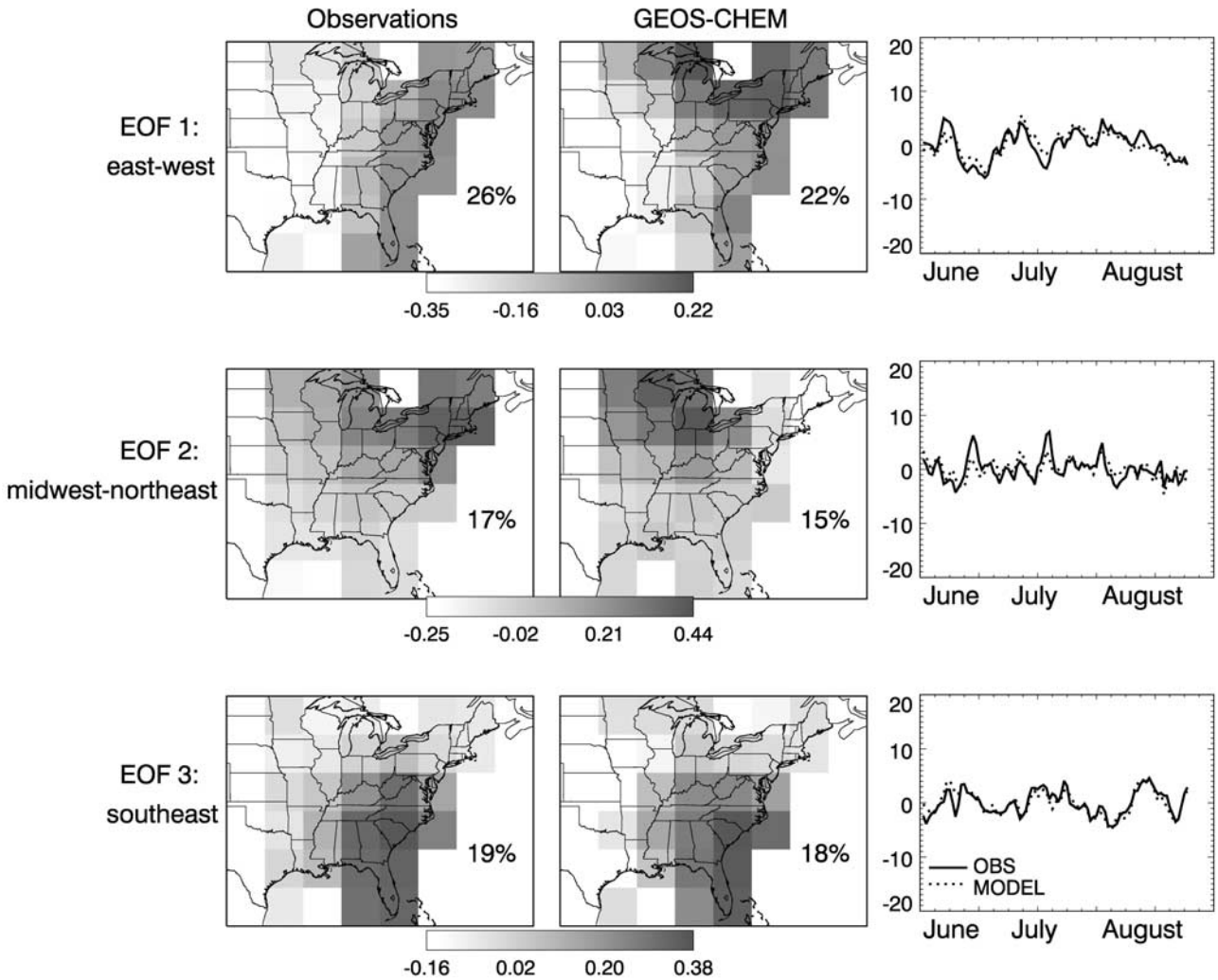

Figure 3c. Same as Figure 3a but for the 4° × 5° resolution of GEOS-CHEM. See color version of this figure at back of this issue.

observed PCs for all three EOFs (r^2 = 0.57–0.68). Although GEOS-CHEM does not capture the midwest-northeast EOF, it has some success in matching the observationally derived PC (r^2 = 0.44–0.54) when the simulated O_3 fields are projected onto the EOF derived from the observations. It does even better at capturing the temporal variability in the east-west and southeast PCs (associated with frontal passages and southeast stagnation, respectively), with r^2 in the range of 0.68–0.78. We conclude that all three models capture the conditions responsible for the timing of regional scale high-O_3 episodes.

6. Conclusions

[24] We have conducted a statistical analysis of observed and simulated afternoon O_3 concentrations in the eastern

Table 2. Correlation Statistics for Simulated Versus Observed O_3 Empirical Orthogonal Functions[a]

	Correlation Statistic	MAQSIP Grid 36 × 36 km^2		GEOS-CHEM 2° × 2.5°		GEOS-CHEM 4° × 5°	
		r^2	Slope	r^2	Slope	r^2	Slope
East-west	EOF	0.86	1.0	0.74	1.2	0.55	1.2
	PC	0.60	0.9	0.68	1.0	0.71	0.9
Midwest-northeast	EOF	0.76	1.0	0.27	1.0	0.01	1.1
	PC	0.57	0.8	0.54	0.8	0.44	0.8
Southeast	EOF	0.80	1.0	0.90	1.0	0.88	1.0
	PC	0.68	0.7	0.78	1.0	0.77	1.0

[a]Table 2 gives coefficients of determination (r^2) and slopes of regression lines for the simulated versus observed empirical orthogonal functions (EOFs) and their associated time series (PC). PC comparisons are for the modeled and observed data projected onto the EOFs derived from observations. The reduced major axis method is used to calculate slopes of regression lines (lines of organic correlation) [*Hirsch and Gilroy*, 1984]. Table 2 summarizes the ability of the MAQSIP and GEOS-CHEM models to reproduce the three principal EOFs (east-west, midwest-northeast, and southeast) in the O_3 observations over the eastern United States in summer 1995 averaged over the corresponding model grids.

United States for the summer of 1995 at three horizontal resolutions: (1) 36 × 36 km^2 with the regional MAQSIP model, (2) 2° × 2.5° with the global GEOS-CHEM model, and (3) 4° × 5° GEOS-CHEM. We find that degrading the horizontal resolution does not introduce systematic bias in the O_3 simulation. Previous work has shown that although the coarse resolution models cannot resolve local maxima, they can describe successfully regional scale high-O_3 events [*Fiore et al.*, 2002a]. Here we find that temporal and spatial correlations between simulated and observed O_3 concentrations (averaged onto the corresponding model grid) improve as resolution coarsens. We conclude that the large-scale processes modulating O_3 concentrations are easier to capture in models than processes occurring at smaller scales, irrespective of model resolution.

[25] We applied principal component analysis (PCA) to identify the fundamental modes of variability (EOFs) in the observations at the three horizontal resolutions and in the corresponding models. The first three Varimax-rotated EOFs in the observations capture over 50% of the variance in O_3 concentrations over the eastern United States in summer 1995, and they are similar at all three spatial resolutions. They appear to represent (1) cold fronts moving longitudinally across the eastern United States (east-west EOF), (2) transport from the midwest to the northeast associated with migratory high-pressure systems traversing the region (midwest-northeast EOF), and (3) southeast stagnation associated with westward extension of the Bermuda high-pressure system (southeast EOF). The similarity between the EOFs and PCs derived from the observations at all three horizontal resolutions indicates that the same fundamental processes are responsible for the observed variability of O_3 at all three scales.

[26] A PCA of the simulated O_3 fields at the three resolutions reveals that the MAQSIP model matches closely the EOFs and PCs derived from observations. The GEOS-CHEM model at 2° × 2.5° resolution reproduces the southeast and east-west EOFs but displaces the midwest-northeast EOF. At the 4° × 5° resolution, GEOS-CHEM still captures the southeast EOF, but degrades the east-west EOF and loses the midwest-northeast EOF. Even when GEOS-CHEM does not properly describe the observed EOFs, it still captures the corresponding PCs (r^2 = 0.44–0.78) when the model fields are projected onto the observationally derived EOFs. We conclude from this result (the time series in Figure 3) that the coarse models capture the synoptic-scale processes responsible for the observed temporal variability of O_3 concentrations, including high-O_3 episodes. We also conclude that analysis of EOFs derived from observed versus simulated O_3 fields provides a useful metric for evaluating models.

[27] **Acknowledgments.** The authors would like to acknowledge S.C. Wang and S.C. Wofsy (Harvard U.) and B.K. Eder (NOAA and U.S. EPA) for insightful discussions and P. Kasibhatla (Duke U.) for providing code to grid the AIRS data onto the MAQSIP grid. We are grateful for support from the Electric Power Research Institute (EPRI) and an NSF graduate fellowship to A.M. Fiore.

References

Bey, I., D. J. Jacob, R. M. Yantosca, J. A. Logan, B. D. Field, A. M. Fiore, Q. Li, H. Y. Liu, L. J. Mickley, and M. G. Schultz, Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, *J. Geophys. Res.*, *106*, 23,073–23,095, 2001.

Byun, D. W., and J. K. S. Ching (Eds.), Science algorithms of the EPA Models-3 Community Multi-scale Air Quality (CMAQ) modeling system, *Rep. EPA/600/R-99/030*, Off. of Res. and Dev., U.S. Environ. Protect. Agency, Research Triangle Park, N. C., 1999.

Cooper, O. R., and J. L. Moody, Meteorological controls on ozone at an elevated eastern United States regional background monitoring site, *J. Geophys. Res.*, *105*, 6855–6869, 2000.

Eder, B. K., J. M. Davis, and P. Bloomfield, A characterization of the spatiotemporal variability of non-urban ozone concentrations over the eastern United States, *Atmos. Environ.*, *27A*(16), 2645–2668, 1993.

Fiore, A. M., D. J. Jacob, I. Bey, R. M. Yantosca, B. D. Field, A. C. Fusco, and J. G. Wilkinson, Background ozone over the United States in summer: Origin, trend, and contribution to pollution episodes, *J. Geophys. Res.*, *107*(D15), 4275, 10.1029/2001JD000982, 2002a.

Fiore, A. M., D. J. Jacob, B. D. Field, D. G. Streets, S. D. Fernandes, and C. Jang, Linking ozone pollution with climate change: The case for controlling methane, *Geophys. Res. Lett.*, *29*(19), 1919, doi:10.1029/2002GL015601, 2002b.

Geron, C. D., A. B. Guenther, and T. E. Pierce, An improved model for estimating emissions of volatile organic compounds from forests in the eastern United States, *J. Geophys. Res.*, *99*, 12,773–12,791, 1994.

Gery, M. W., G. Z. Whitten, J. P. Killus, and M. C. Dodge, A photochemical kinetics mechanism for urban and regional scale computer modeling, *J. Geophys. Res.*, *94*, 12,925–12,956, 1989.

Hirsch, R. M., and E. J. Gilroy, Methods of fitting a straight line to data: Examples in water resources, *Water Res. Bull.*, *20*, 705–711, 1984.

Hogrefe, C., "Evaluating the regional-scale photochemical modeling systems and using them for regulatory policy making," Ph.D. thesis, 175 pp., Dep. of Earth and Atmos. Sci., State Univ. of N. Y. at Albany, Albany, 2000.

Hogrefe, C., S. T. Rao, I. G. Zurbenko, and P. S. Porter, Interpreting the information in ozone observations and model predictions relevant to regulatory policies in the eastern United States, *Bull. Am. Meteorol. Soc.*, *81*, 2083–2106, 2000.

Hogrefe, C., S. T. Rao, P. Kasibhatla, G. Kallos, C. J. Tremback, W. Hao, D. Olerud, A. Xiu, J. McHenry, and K. Alipaty, Evaluating the performance of regional-scale photochemical modeling systems: Part I—Meteorological predictions, *Atmos. Environ.*, *35*, 4159–4174, 2001a.

Hogrefe, C., S. T. Rao, P. Kasibhatla, W. Hao, G. Sistla, R. Mathur, and J. McHenry, Evaluating the performance of regional-scale photochemical modeling systems: Part II—Ozone predictions, *Atmos. Environ.*, *35*, 4175–4188, 2001b.

Houyoux, M. R., J. M. Vukovich, C. J. Coats Jr., N. J. M. Wheeler, and P. Kasibhatla, Emission inventory development and processing for the seasonal model for regional air quality, *J. Geophys. Res.*, *105*, 9079–9090, 2000.

Kasibhatla, P., and W. L. Chameides, Seasonal modeling of regional ozone pollution in the eastern United States, *Geophys. Res. Lett.*, *27*, 1415–1418, 2000.

Kasibhatla, P., W. L. Chameides, B. Duncan, M. Hoyoux, C. Jang, R. Mathur, T. Odman, and A. Xiu, Impact of inert organic nitrate formation on ground-level ozone in a regional air quality model using the Carbon Bond Mechanism 4, *Geophys. Res. Lett.*, *24*, 3205–3208, 1997.

Kumar, N., and A. G. Russell, Multiscale air quality modeling of the northeastern United States, *Atmos. Environ.*, *30*, 1099–1116, 1996.

Kumar, N., M. T. Odman, and A. G. Russell, Multiscale air quality modeling: Application to southern California, *J. Geophys. Res.*, *99*, 5385–5397, 1994.

Lefohn, A. S., D. S. Shadwick, and S. D. Ziman, The difficult challenge of attaining EPS's new ozone standard, *Environ. Sci. Technol.*, *32*, 276A–282A, 1998.

Li, Q., et al., Transatlantic transport of pollution and its effects on surface ozone in Europe and North America, *J. Geophys. Res.*, *107*(D13), 4166, doi:10.1029/2001JD001422, 2002.

Liang, J., and M. Z. Jacobson, Effects of subgrid segregation on ozone production efficiency in a chemical model, *Atmos. Environ.*, *34*, 2975–2982, 2000.

Liu, S. C., M. Trainer, F. C. Fehsenfeld, D. D. Parrish, E. J. Williams, D. W. Fahey, G. Hubler, and P. C. Murphy, Ozone production in the rural troposphere and the implications for regional and global ozone distributions, *J. Geophys. Res.*, *92*, 4191–4207, 1987.

Logan, J. A., Ozone in rural areas of the United States, *J. Geophys. Res.*, *94*, 8511–8532, 1989.

Lurmann, F. W., A. C. Lloyd, and R. Atkinson, A chemical mechanism for use in long-range transport/acid deposition computer modeling, *J. Geophys. Res.*, *91*, 10,905–10,936, 1986.

Moody, J. L., J. W. Munger, A. H. Goldstein, D. J. Jacob, and S. C. Wofsy, Harvard Forest regional-scale air mass composition by Patterns in Atmospheric Transport History (PATH), *J. Geophys. Res.*, *103*, 13,181–13,194, 1998.

National Weather Service, Daily weather maps, weekly series, Nat. Meteorol. Cent. Clim. Anal. Cent., Natl. Ocean and Atmos. Admin., Washington, D. C., 1995.

Odman, M. T., and C. I. Ingram, Multiscale Air Quality Simulation Platform (MAQSIP): Source code documentation and validation, *Tech. Rep. ENV-96TR002-v1.0*, 83 pp., MCNC, Research Triangle Park, N. C., 1996.

Ogburn, C., L. Boothe, M. Mullen, and P. F. Brewer, Emissions inventories for the southern Appalachian Mountains initiative, paper presented at Annual Meeting, Air and Waste Mange. Assoc., Salt Lake City, Utah, June, 2000.

Olszyna, K. J., W. J. Parkhurst, and J. F. Meagher, Air chemistry during the 1995 SOS/Nashville intensive determined from level 2 network, *J. Geophys. Res.*, *103*, 31,143–31,153, 1998.

Ryan, W. F., B. G. Doddridge, R. R. Dickerson, R. M. Morales, K. A. Hallock, P. T. Roberts, D. L. Blumenthal, J. A. Anderson, and K. L. Civerolo, Pollutant transport during a regional O_3 episode in the mid-Atlantic states, *J. Air Waste Manage. Assoc.*, *48*, 786–797, 1998.

Sillman, S., J. A. Logan, and S. C. Wofsy, A regional scale model for ozone in the United States with subgrid representation of urban and power plant plumes, *J. Geophys. Res.*, *95*, 5731–5748, 1990.

Sistla, G., W. Hao, J.-Y. Ku, G. Kallos, K. Zhang, H. Mao, and S. T. Rao, An operational evaluation of two regional-scale ozone air quality modeling systems over the eastern United States, *Bull. Am. Meteorol. Soc.*, *82*, 945–964, 2001.

Sonoma Technology, Inc., Analysis of data from the 1995 NARSTO-northeast study, in *Three-Dimensional Analysis of Aircraft and Upper-Air Meteorological Data*, vol. V, *Final Rep. STI-95426-1836-FR*, 125 pp., Coord. Res. Counc., Atlanta, Ga., 1999.

Vukovich, F. M., Regional-scale boundary layer ozone variations in the eastern United States and their association with meteorological variations, *Atmos. Environ.*, *29*, 2259–2273, 1995.

Vukovich, F. M., Time scales of surface ozone variations in the regional, non-urban environment, *Atmos. Environ.*, *31*, 1513–1530, 1997.

Wild, O., and H. Akimoto, Intercontinental transport of ozone and its precursors in a three-dimensional global CTM, *J. Geophys. Res.*, *106*, 27,729–27,744, 2001.

Wilks, D. S., *Statistical Methods in the Atmospheric Sciences*, Academic, San Diego, Calif., 1995.

Yarnal, B., *Synoptic Climatology in Environmental Analysis, A Primer*, Belhaven, London, 1993.

Yienger, J. J., et al., The episodic nature of air pollution transport from Asia to North America, *J. Geophys. Res.*, *105*, 26,931–26,945, 2000.

A. M. Fiore and D. J. Jacob, Department of Earth and Planetary Sciences and Division of Engineering and Applied Sciences, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA. (afiore@post.harvard.edu)

R. V. Martin, Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA.

R. Mathur, Environmental Modeling Center, MCNC, Research Triangle Park, NC 27709, USA.

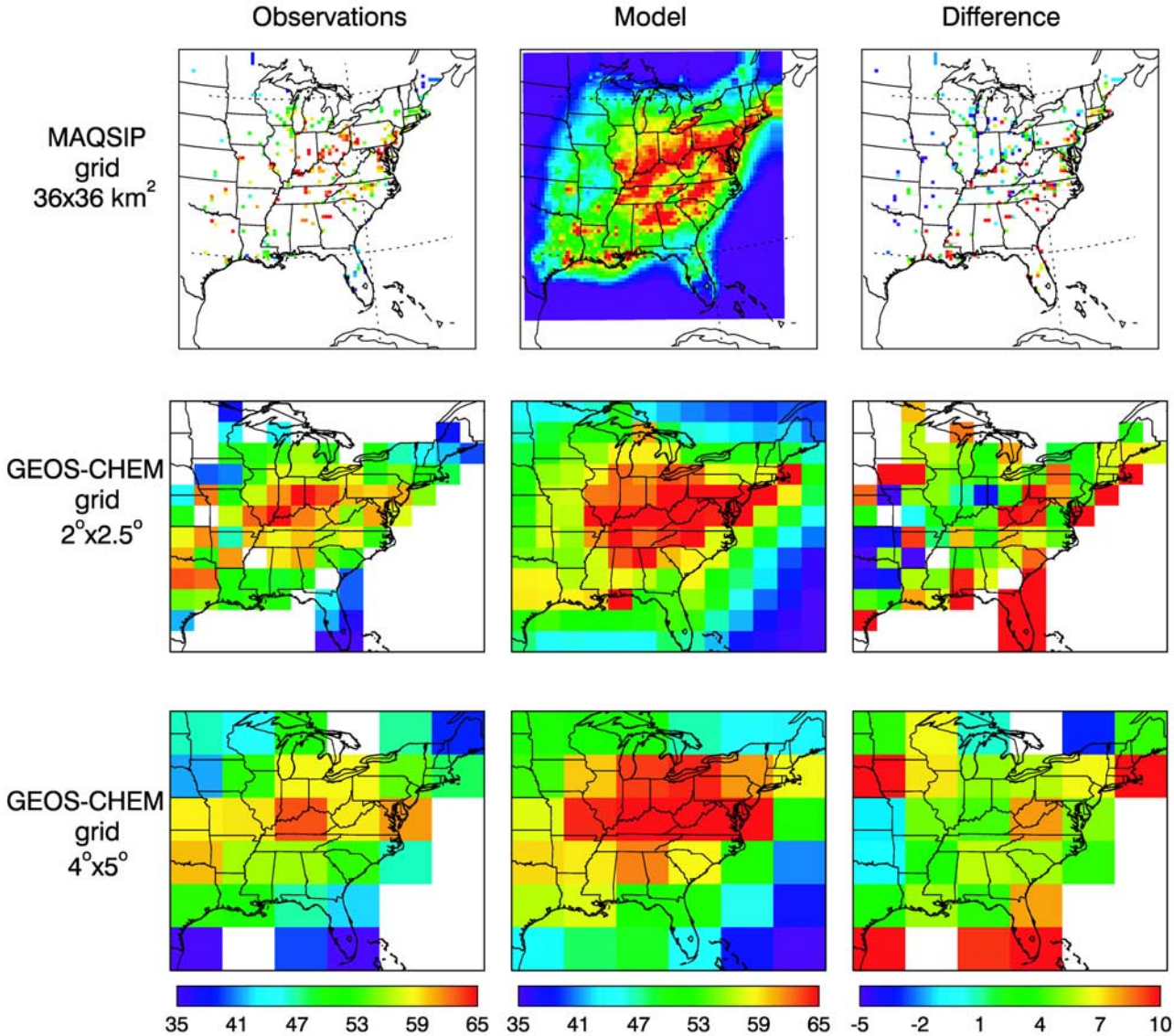

Figure 1. Mean afternoon (1300–1700 LT) surface O_3 concentrations (ppbv) in June–August 1995. (left column) AIRS observations averaged onto the MAQSIP (36 × 36 km^2) and GEOS-CHEM (2° × 2.5° and 4° × 5°) model grids. (middle column) Simulated fields from the MAQSIP and GEOS-CHEM models. (right column) Difference between observed and simulated fields (model bias). The color scale is designed to resolve regional features rather than extreme concentrations and saturates. The range of observed values is 34–73 (MAQSIP), 31–66 (GEOS-CHEM 2° × 2.5°), and 33–63 (GEOS-CHEM 4° × 5°). The range of simulated values is 17–103 (MAQSIP), 34–74 (GEOS-CHEM 2° × 2.5°), and 32–67 (GEOS-CHEM 4° × 5°). The maximum differences are −28 and +43 ppbv (MAQSIP), −6 and +26 ppbv (GEOS-CHEM 2° × 2.5°), and −3 and +12 ppbv (GEOS-CHEM 4° × 5°).

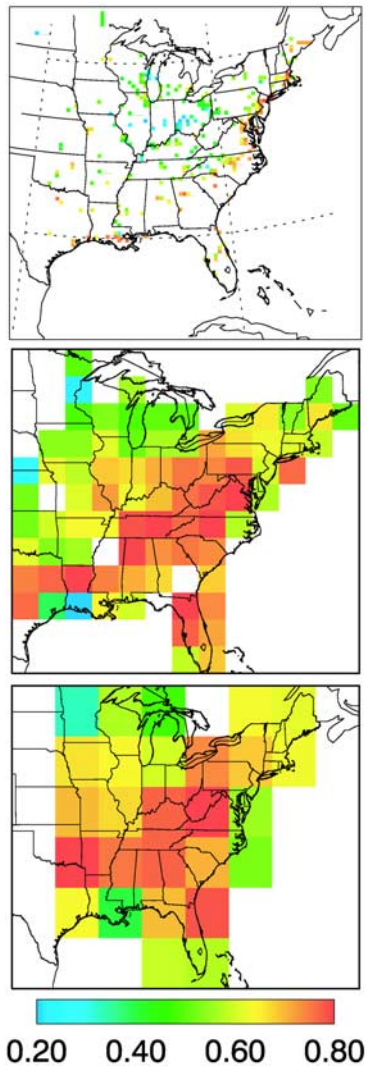

Figure 2. Correlation coefficients (r) between the simulated and observed daily afternoon (1300–1700 LT) mean surface O_3 concentrations during the summer of 1995. The color bar is saturated at 0.2 and 0.8.

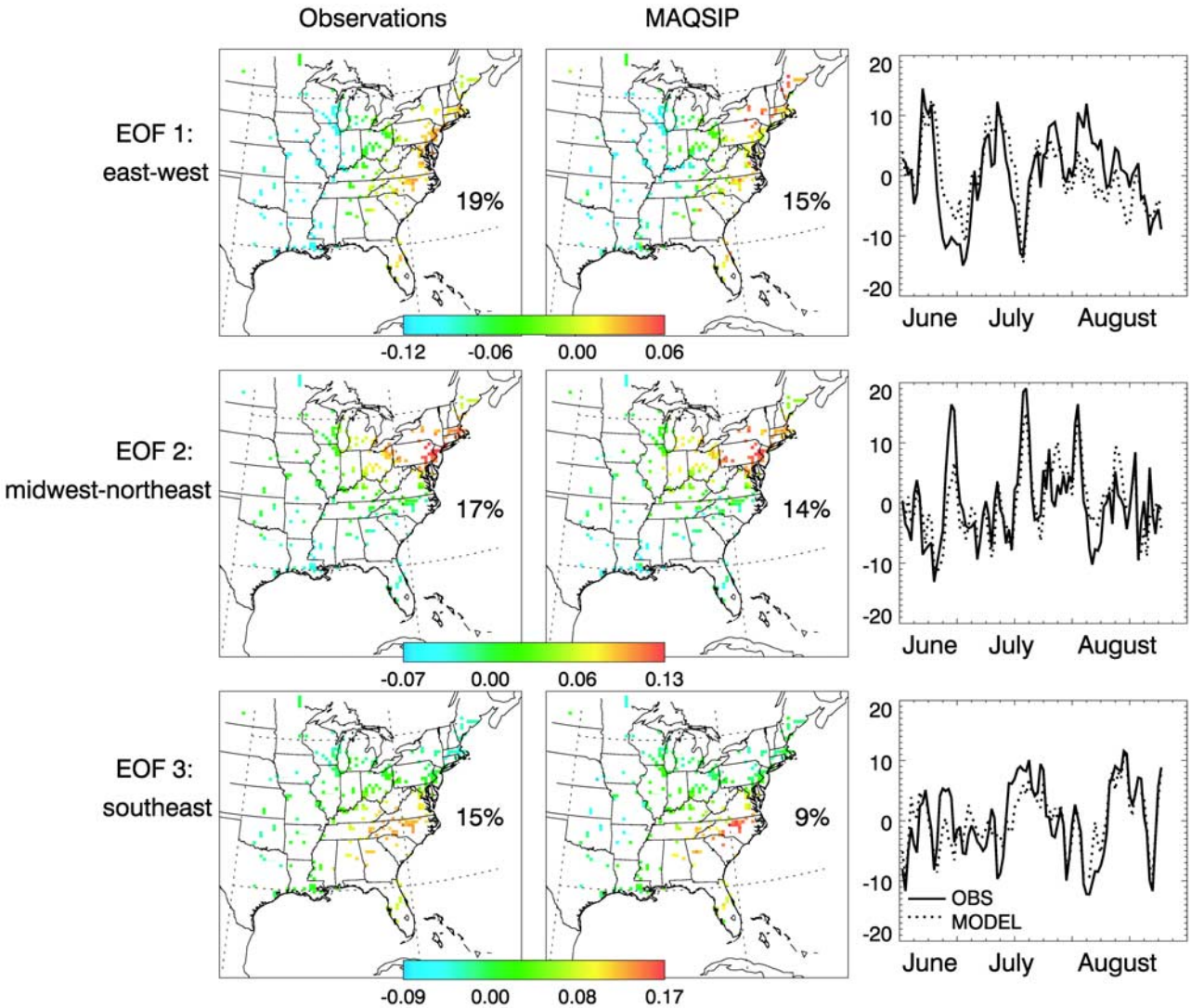

Figure 3a. (left and middle columns) The first three Varimax-rotated EOFs and (right columns) associated PCs as derived from daily afternoon average (1300–1700 LT) surface O_3 concentrations for the summer of 1995 (left) in the observations averaged on the MAQSIP model grid and (middle) in the MAQSIP model. The percentage of total variance explained by the individual EOFs is shown. The right column shows the PCs (temporal variation in the strength of the EOFs, solid lines) derived from the observations and the corresponding time series (dotted lines) obtained from projecting the simulated daily summer afternoon O_3 concentrations onto the (left column) observationally derived EOFs.

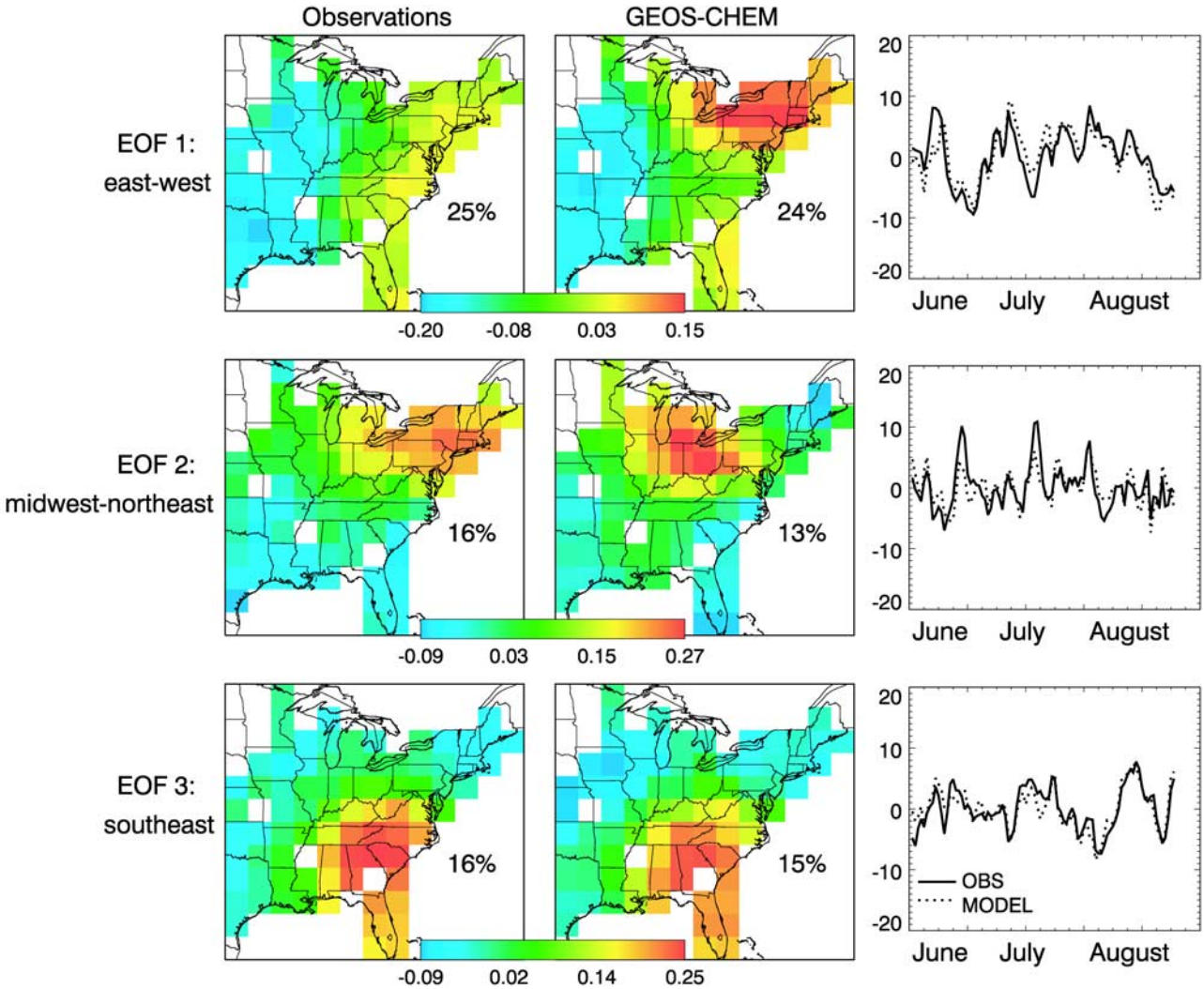

Figure 3b. Same as Figure 3a but for the 2° × 2.5° resolution of GEOS-CHEM.

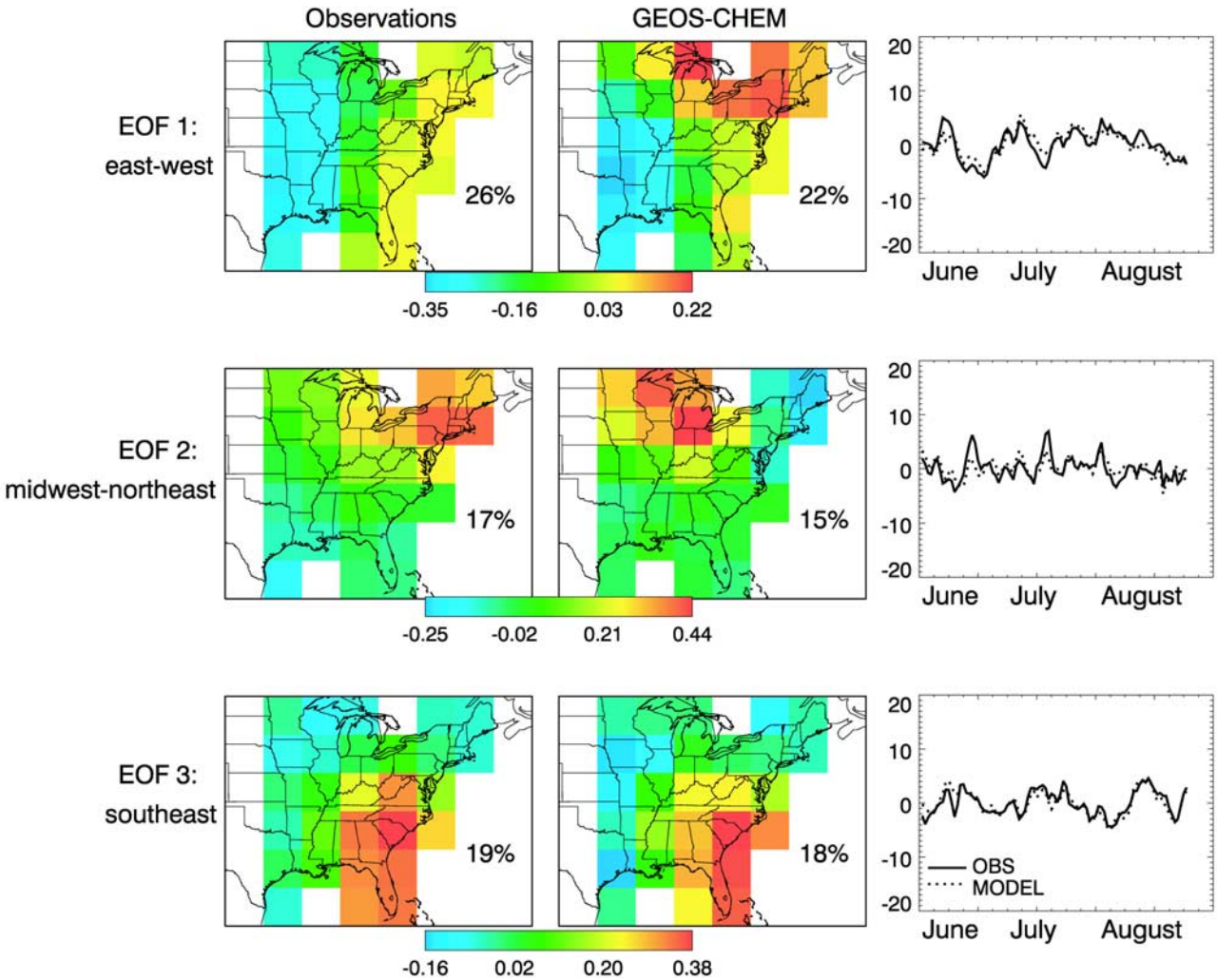

Figure 3c. Same as Figure 3a but for the 4° × 5° resolution of GEOS-CHEM.