NASA/TM-2021-104606/Vol. 55

Technical Report Series on Global Modeling and Data Assimilation, Volume 55

Randal D. Koster, Editor

A Phenomenon-Based Decomposition of Model-Based Estimates of Boreal Winter ENSO Variability

Siegfried Schubert, Young-Kwon Lim, Andrea Molod, and Allison Marquardt Collow

February 2021

NASA STI Program ... in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at http://www.sti.nasa.gov

- E-mail your question to help@sti.nasa.gov

- Phone the NASA STI Information Desk at 757-864-9658

- Write to:
  NASA STI Information Desk
  Mail Stop 148
  NASA Langley Research Center
  Hampton, VA 23681-2199

NASA/TM-2021-104606/Vol. 55

Technical Report Series on Global Modeling and Data Assimilation, Volume 55

Randal D. Koster, Editor

A Phenomenon-Based Decomposition of Model-Based Estimates of Boreal Winter ENSO Variability

Siegfried Schubert
Science Systems and Applications, Inc., Lanham, MD

Young-Kwon Lim
Universities Space Research Association, Columbia, MD

Andrea Molod
Goddard Space Flight Center, Greenbelt, MD

Allison Marquardt Collow
Universities Space Research Association, Columbia, MD

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

February 2021

Level of Review: This material has been technically reviewed by technical management.

Available from

NASA STI Program
Mail Stop 148
NASA's Langley Research
Center Hampton, VA
23681-2199

National Technical Information
Service 5285 Port Royal Road
Springfield, VA 22161
703-605-6000

Table of Contents

List of Figures:

Figure 1a: The variance decomposition of the 250mb eddy height for El Niño events that occurred during 1980-2016, averaged over the months December through March, with terms defined in eq. 2.1.2. See text for details. Units are m^2.

Figure 1b: Same as Fig. 1a, but for La Niña.

Figure 2a: The variance decomposition of the 250mb u-wind for El Niño events that occurred during 1980-2016, averaged over the months December through March, with terms defined in eq. 2.1.2. See text for details. Units are m^2/s^2.

Figure 2b: Same as Fig. 2a, but for La Niña.

Figure 3a: The variance decomposition of the precipitation for El Niño events that occurred during 1980-2016, averaged over the months December through March, with terms defined in eq. 2.1.2. See text for details. Units are mm^2/day^2.

Figure 3b: Same as Fig. 3a, but for La Niña.

Figure 4a: The variance decomposition of the precipitation over North America for El Niño events that occurred during 1980-2016, averaged over the months December through March, with terms defined in eq. 2.1.2. See text for details. Units are mm^2/day^2.

Figure 4b: Same as Fig. 4a, but for La Niña.

Figure 5a: The variance decomposition of T2m over North America for El Niño events that occurred during 1980-2016, averaged over the months December through March, with terms defined in eq. 2.1.2. See text for details. Units are $°K^2$.

Figure 5b: Same as Fig. 5a, but for La Niña.

Figure 6: The variance of the 250mb eddy height composite mean based on MERRA-2 (Y'^2, left panels) and M2AMIP ($\{X'^2\}$, right panels). See 2.1.11 in the text. Top panels are for El Niño, and bottom panels are for La Niña events that occurred during 1980-2016. Units: m^2.

Figure 7: The variance of the 250mb U-wind composite mean based on MERRA-2 (Y'^2, left panels) and M2AMIP ($\{X'^2\}$, right panels). See 2.1.11 in the text. Top panels are for El Niño, and bottom panels are for La Niña events that occurred during 1980-2016. Units: m^2/s^2.

Figure 8: The variance of the precipitation composite mean based on MERRA-2 (Y'^2, left panels) and M2AMIP ($\{X'^2\}$, right panels). See 2.1.11 in the text. Top panels are for El Niño, and bottom panels are for La Niña events that occurred during 1980-2016. Units: mm^2/day^2.

Figure 9: The variance of the North American precipitation composite mean based on MERRA-2 (Y'^2, left panels) and M2AMIP ($\{X'^2\}$, right panels). See 2.1.11 in the text. Top panels are for El Niño, and bottom panels are for La Niña events that occurred during 1980-2016. Units: mm^2/day^2.

Figure 10: The variance of the North American T2m composite mean based on MERRA-2 (Y'^2, left panels) and M2AMIP ($\{X'^2\}$, right panels). See 2.1.11 in the text. Top panels are for El Niño, and bottom panels are for La Niña events that occurred during 1980-2016. Units: $°K^2$.

Figure 11: E2E results for 250mb eddy height for El Niño (left panels) and La Niña (right panels) averaged over the months December through March for the events that occurred during 1980-2016. Top panel: E2E variance based on MERRA-2. Middle panel: E2E (Term 3 +Term 4) variance based on M2AMIP. Units: m^2. Bottom panel: E2E signal (Term 3) / E2E noise (Term 4) based on M2AMIP.

Figure 12: E2E results for 250mb u-wind for El Niño (left panels) and La Niña (right panels) averaged over the months December through March for the events that occurred during 1980-2016. Top panel: E2E variance based on MERRA-2. Middle panel: E2E (Term 3 +Term 4) variance based on M2AMIP. Units: m^2/s^2. Bottom panel: E2E signal (Term 3) / E2E noise (Term 4) based on M2AMIP.

Figure 13: E2E results for precipitation for El Niño (left panels) and La Niña (right panels) averaged over the months December through March for the events that occurred during 1980-2016. Top panel: E2E variance based on MERRA-2. Middle panel: E2E (Term 3 +Term 4) variance based on M2AMIP. Units: mm^2/day^2. Bottom panel: E2E signal (Term 3) / E2E noise (Term 4) based on M2AMIP.

Figure 14: E2E results for precipitation over North America for El Niño (left panels) and La Niña (right panels) averaged over the months December through March for the events that occurred during 1980-2016. Top panel: E2E variance based on MERRA-2. Middle panel: E2E (Term 3 +Term 4) variance based on M2AMIP. Units: mm^2/day^2. Bottom panel: E2E signal (Term 3) / E2E noise (Term 4) based on M2AMIP.

Figure 15: E2E results for T2m over North America for El Niño (left panels) and La Niña (right panels) averaged over the months December through March for the events that occurred during 1980-2016. Top panel: E2E variance based on MERRA-2. Middle panel: E2E (Term 3 +Term 4) variance based on M2AMIP. Units: $°K^2$. Bottom panel: E2E signal

(Term 3) / E2E noise (Term 4) based on M2AMIP.

Figure 16: The signal-to-total variance ratios for precipitation for the composite mean (S/T $|_{composite}$, left panels), and event-to-event variability (S/T $|_{E2E}$, right panels). Top row is for El Niño and the bottom row is for La Niña for the events that occurred during 1980-2016. See Section 2 for details.

Figure 17: Same as Fig. 16, except for 250mb eddy height.

Figure 18: Same as Fig. 16, except for 250mb u-wind.

Figure 19: The signal-to-total variance ratios for precipitation (top four panels) and T2m (bottom four panels) over North America for the composite mean (S/T $|_{composite}$, left panels), and event-to-event variability (S/T $|_{E2E}$, right panels). Top row in each set of four is for El Niño and the bottom row is for La Niña for the events that occurred during 1980-2016. See Section 2 for details.

Figure 20a: The El Niño January (top panels), February (middle panels) and average of January through March (bottom panels) conditional correlations between MERRA-2 and M2AMIP for the 250mb eddy height field for the events that occurred during 1980-2016. The correlations are decomposed into the terms associated with event-to-event (E2E) variability (left panels) and the composite mean (right panels). Values not significant at the 1% level based on a Monte Carlo approach to assess ensemble uncertainty and/or correlations with absolute values less than 0.2 are masked out. Positive values are contoured.

Figure 20b: Same as Fig. 20a except for La Niña.

Figure 21a: The El Niño January (top panels), February (middle panels) and average of January through March (bottom panels) conditional correlations between MERRA-2 and M2AMIP for the 250mb u-wind field for the events that occurred during 1980-2016. The correlations are decomposed into the terms associated with event-to-event (E2E) variability (left panels) and the composite mean (right panels). Values not significant at the 1% level based on a Monte Carlo approach to assess ensemble uncertainty and/or correlations with absolute values less than 0.2 are masked out. Positive values are contoured.

Figure 21b: Same as Fig. 21a except for La Niña.

Figure 22a: The El Niño January (top panels), February (middle panels) and average of January through March (bottom panels) conditional correlations between MERRA-2 and M2AMIP for precipitation for the events that occurred during 1980-2016. The correlations are decomposed into the terms associated with event-to-event (E2E) variability (left panels) and the composite mean (right panels). Values not significant at the 1% level based on a

Monte Carlo approach to assess ensemble uncertainty and/or correlations with absolute values less than 0.2 are masked out. Positive values are contoured.

Figure 22b: Same as Fig. 22a except for La Niña.

Figure 23: The conditional correlations between MERRA-2 and M2AMIP for precipitation over North America for El Niño (a) and La Niña (b) events that occurred during 1980-2016. In each set of six panels, the results are shown for January (top panels), February (middle panels) and average of January through March (bottom panels). The correlations are decomposed into the terms associated with event-to-event (E2E) variability (left panels) and the composite mean (right panels). Values not significant at the 1% level based on a Monte Carlo approach to assess ensemble uncertainty and/or values with absolute correlations less than 0.2 are masked out. Positive values are contoured.

Figure 24: The conditional correlations between MERRA-2 and M2AMIP for T2m over North America for El Niño (a) and La Niña (b) events that occurred during 1980-2016. In each set of six panels, the results are shown for January (top panels), February (middle panels) and average of January through March (bottom panels). The correlations are decomposed into the terms associated with event-to-event (E2E) variability (left panels) and the composite mean (right panels). Values not significant at the 1% level based on a Monte Carlo approach to assess ensemble uncertainty and/or correlations with absolute values less than 0.2 are masked out. Positive values are contoured.

Abstract

Climate models are now routinely being used to simulate and predict climate variability on time scales ranging from sub-seasonal to seasonal and longer. As such, there are now long histories of such simulations and predictions spanning multiple decades and multiple ensemble members, both of which are crucial for separating climate signal from climate noise. A key focus of such runs has been the El Niño-Southern Oscillation (ENSO), spurred by recent improvements in our ability to predict such events, though questions remain as to how well climate models do beyond simply always predicting the "canonical" atmospheric response to an ENSO event—something simple statistical models already do reasonably well. This is a critical issue that needs addressing, given the importance of event-to-event differences for predicting regional impacts of ENSO teleconnections, and the need to justify the expense of running sophisticated climate models. Unfortunately, current diagnostic tools are not well suited for quantifying the different sources of variability associated with specific phenomena such as ENSO. More generally, while much effort has focused on addressing model bias, less has been done to address errors in second moment statistics—an issue whose importance is gaining increased attention particularly as we build climate prediction systems capable of taking advantage of forecasts of opportunity—a capability that requires reliable estimates of forecast uncertainty.

In this report, we outline a phenomenon-based statistical decomposition of climate variance (in essence a detailed variance budget) that is specifically tailored to address the above questions by separating the variability (both the signal and noise) into that tied to the long-term average impact of a particular phenomenon (the composite mean) and the event-to-event (E2E) variability about the composite mean. In addition, we provide related decompositions of the correlations that allow us to quantify how much of the agreement with observations (the skill) comes from the composite mean as opposed to from the E2E variability. As an example, we present the results of such a decomposition for ENSO based on simulations with the GEOS atmospheric general circulation model (AGCM), with a focus on the monthly mean impacts over North America during boreal winter (December – March). Here we take advantage of existing GEOS AGCM simulations that were produced as companion simulations to MERRA-2 for the period 1980-2016. Comparisons are made throughout with MERRA-2.

Insights into the quality of the model's ENSO-related variability, including that of the upper tropospheric eddy (deviations from the zonal mean) height and zonal wind as well as the near surface air temperature and precipitation over North America, are facilitated by providing observational constraints on selected components of the variance budget. A key finding—one that highlights the advantage of the variance decomposition— is that the monthly mean impacts of La Niña on the upper tropospheric circulation over the North Pacific/North American region during boreal winter are, for the most part, currently not predictable with the GEOS AGCM beyond what can be achieved from predicting the canonical (composite) La Niña response, while there is some evidence that we can predict

E2E differences in the responses over the North Pacific/North American region for El Niño. The extent to which these results are indicative of fundamental limits to ENSO prediction are, however, unclear, given a number of deficiencies in the model's ability to reproduce ENSO-related variability, including a too strong composite signal over North America and excessive unpredictable E2E noise (especially for El Niño) over much of the North Pacific and North America. As such, it is clear our findings about the predictability of ENSO need to be verified with other models, including through hindcasts carried out with coupled models in which the predictability of the SST is also considered.

We believe that the variance decomposition detailed here should provide an important metric (one that focuses on second moments) for evaluating model performance. Such an evaluation is currently not a trivial task, since it requires long, multi-decadal histories of simulations/hindcasts and large enough ensembles to overcome statistical sampling errors. The availability of such simulations and hindcasts, however, should become more routine as computing resources increase in the coming years.

1. Introduction

Despite the substantial amount of progress made in understanding and modeling the El Niño–Southern Oscillation (ENSO) phenomenon, there are still considerable unknowns about what controls the characteristics (e.g., strength, temporal evolution, spatial pattern, etc.) of specific ENSO events and our ability to predict these characteristics. As reported in a recent World Meteorological Organization report (WMO, 2015), "ENSO remains a problem worth solving – there are few if any natural semi-regular climate signals whose prediction can have such widespread impact – but we are not there yet. We continue to be surprised by the diversity of ENSO events. Regular oscillations in the 1960s and 1970s shaped early understanding of ENSO. The 1980s and 1990s were dominated by El Niño phases—including the large 1982/83 and 1997/98 events. The recent period has been accompanied by changes in the mean state, reduced variability and the so-called "Modoki" El Niño."

Concerns about the predictability of ENSO and the nature of ENSO diversity were expressed in the same WMO 2015 report: "Perplexed by the apparent failures of ENSO forecasts in 2012 and again in 2014, several researchers ask whether changes in ENSO reflect larger shifts or changes in the planet-wide climate system." Some progress in our understanding of ENSO prediction was reported by Sohn et al. (2019), who found that the 6-month prediction skill in APEC Climate Center multi-model ensemble (MME) forecasts depends on both the strength and the flavor of ENSO. Stratifying the sea surface temperature (SST) into that associated with a typical ENSO and its residual, they found that the typical ENSO is the major source of predictability of tropical Pacific SST, while the residual ENSO variability acts to limit tropical rainfall predictability. Also relevant to the predictability question is the study by Imada et al. (2015), which found that, compared to the conventional eastern Pacific El Niño, the central Pacific El Niño has more limited predictability. They suggested that the relatively small amplitude of the central Pacific El Niño (which is thus more affected by atmospheric noise) is the reason for the limited predictability.

Concerns about our lack of understanding of ENSO diversity and our ability to predict it are, in fact, reflected in a broad range of studies addressing the case-to-case variability of ENSO (e.g., Kumar and Hoerling 1997; Hoerling and Kumar 1997; Kao and Yu 2009; Kug et al. 2009; Capotondi et al. 2015), the nonlinearity of ENSO (e.g., Hoerling et al. 1997; Hoerling et al. 2001; Zhang et al. 2014; Frauen et al. 2014), and the question of whether complex dynamical models are even capable of outperforming much simpler statistical models (Kumar et al. 1996; Barnston et al. 1999; Kirtman et al. 2000; Goddard et al. 2005). There is however general agreement that, despite limitations in the skill of current forecast models, climate forecasts are nevertheless more accurate during El Niño and La Niña events compared with ENSO neutral periods (e.g., Goddard and Dilley, 2005). Furthermore, there is evidence that the skill of precipitation and temperature forecasts over North America is higher for El Niño events than for La Niña events (Chen et al. 2017), and that this appears

to reflect the fact that the El Niño response over North America is stronger than the La Niña response, resulting in higher signal-to-noise ratios for El Niño events (e.g., Hoerling et al. 2001; Frauen et al. 2014).

Digging deeper into the ability to predict ENSO impacts, Chen et al. (2017) concluded (based on an analysis of North American Multi-Model Ensemble (NMME) predictions): "For probability composites, all models have superior performance in predicting ENSO precipitation patterns than temperature patterns." They also found that "predictive skill varies with month. All models, as well as NMME, have greater ACC for February prediction, and this is seen for both P and T anomaly composites under either El Niño or La Niña condition." Lim et al. (2020) carried out an in-depth analysis of the underlying reasons for the within-season variations in the boreal winter skill found by Chen et al. (2017), focusing on why the skill of January forecasts is so much poorer than that of February forecasts in the NASA GEOS model (forecasts that are contributed to the NMME). That study found that the relatively poor skill of the January forecasts in the GEOS model was the result of biases in the January climatological stationary waves rather than the result of errors in the tropical Pacific El Niño heating anomalies in January. Furthermore, evidence was provided that the relatively poor simulation of the observed January climatology, which is characterized by a strengthened North Pacific jet and enhanced ridge over western North America, can be traced back to biases in the January climatological heating over the Tibet region and the tropical western Pacific.

Given the still on-going uncertainties about the nature of ENSO diversity and our ability to predict variations in ENSO character, we believe an important step in making progress on both of those research fronts is to develop new metrics that are specifically tailored to: (i) separate the event-to-event variations from the canonical (composite mean) ENSO response, and (ii) quantify the separate contributions of each to forecast skill. The variance decomposition presented here does just that. By distinguishing between the contributions to the total variability coming from the composite mean ENSO response and event-specific ENSO character, it allows a direct and quantitative assessment of how well models perform in simulating (or predicting) both. While we focus on ENSO, the variance (and covariance) decomposition presented here can be applied to any recurring climate phenomenon, with the main caveat being that there must be enough recurrences of the event in the climate record to provide reliable statistics. More generally, this approach can provide an important metric of performance for climate models, especially those that are used for climate prediction, given that an estimate of the uncertainty is a key component of the forecast product.

Section 2 presents the methodology used for the variance (Section 2.1) and the covariance (Section 2.2) decomposition. The datasets used in this study are described in Sections 3.1 (MERRA-2) and 3.2 (the model simulations). The results are presented in Section 4. This includes the complete variance budget (Section 4.1), a focus on the variance associated with the composite mean (section 4.2), a focus on the event-to-event variance (Section 4.3), the

signal-to-total variance ratios (Section 4.4), and the correlations with observations/MERRA-2 (Section 4.5). Section 5 provides a summary and discussion.

2. Methodology

The underlying assumption is that there exists a particular recurring phenomenon (PHEN), whose variability and impacts we are trying to isolate. Furthermore, we assume that there exists an ensemble of realizations (model simulations or predictions) of the phenomenon in question. With that in mind we begin by defining three averaging operators:

$[\ \ ]$ *is the long term (climate) mean*

$\langle\ \ \rangle$ *is the composite mean over periods when PHEN is active*

$\{\ \ \}$ *is the ensemble mean*

Then, for a monthly mean[1] simulated quantity X, the anomaly with respect to the long-term ensemble mean can be decomposed into:

$$X - [\{X\}] = X^* + \{X'\} + \{\hat{X}\}, \text{ where} \qquad 2.1.1$$

$X^* = X - \{X\}$ is the anomaly about the ensemble mean,

$X' = \langle X\rangle - [X]$ is the anomaly of the PHEN composite about the long term mean, and

$\hat{X} = X - \langle X\rangle$ is the anomaly about the PHEN composite.

It is important to note that what we are considering in our variance decomposition (see 2.1.2) are the anomalies and variability with respect to the long-term climate mean $[\ \ ]$, but conditioned on (averaged over) periods, $\langle\ \ \rangle$, when the phenomenon of interest is active.

[1] We consider here monthly means, though this is somewhat arbitrary (we could, for example, have used seasonal means instead). Our choice of monthly means allows for an assessment of within-season differences in the ENSO variability.

2.1 Decomposition of the variance

With the above definitions, the square of an anomaly with respect to the long-term ensemble mean ($X - [\{X\}]$) averaged over all PHEN events and all ensemble members (the total variance) is:

$$\{\langle (X - [\{X\}])^2 \rangle\} =$$

$$\{(X' - \{X'\})^2\} + \{X'\}^2 + \langle \{\hat{X}\}^2 \rangle + \{\langle (X^* - \langle X^* \rangle)^2 \rangle\} \qquad 2.1.2$$

Composite noise	**Composite signal**	**E2E signal**	**E2E noise**
(1)	(2)	(3)	(4)

The first term on the RHS of 2.1.2 is the intra-ensemble variance of the composite mean and, as such, represents the inherent unpredictable noise associate with that composite. The second term on the right-hand side (RHS) is the signal associated with the composite mean event. Terms three and four deal with the signal and noise, respectively, associated with the event-to-event (referred to hereafter as E2E) differences from the composite mean. We note that the sum of terms two and three in 2.1.2 is the ***total signal***:

$$\langle (\{X\} - [\{X\}])^2 \rangle = \{X'\}^2 + \langle \{\hat{X}\}^2 \rangle. \qquad 2.1.3$$

We make such a separation of the signal with the idea that it is the signal associated with the E2E variability ($\langle \hat{X}^2 \rangle$) that separates the prediction capabilities of dynamical models from those of simple statistical/empirical methods, which primarily provide information on the ENSO composite ($\{X'\}^2$).

Also, the ***total noise*** (intra-ensemble variance) is the sum of terms (1) and (4) in 2.1.2:

$$\langle\{X^{*2}\}\rangle \;=\; \{(X' - \{X'\})^2\} \quad + \quad \{\langle(X^* - \langle X^*\rangle)^2\rangle\}. \qquad 2.1.4$$

It is also useful to note that the ***total E2E variance*** is the sum of terms 3 (signal) and 4 (noise),

$$\langle\{\hat{X}\}^2\rangle \quad + \quad \{\langle(X^* - \langle X^*\rangle)^2\rangle\} = \quad \{\langle\hat{X}^2\rangle\} \qquad 2.1.5$$

Furthermore, the ***total composite mean variance*** is the sum of terms 1 (composite noise) and 2 (composite signal),

$$\{(X' - \{X'\})^2\} \quad + \{X'\}^2 \;=\; \{X'^2\}. \qquad 2.1.6$$

In order to facilitate the assessment of predictability, we define two signal-to-total variance ratios for the two sources of signal defined in 2.1.3 (E2E and composite). The first quantifies the potential predictability of individual events (E2E) and involves the ratio of term 3 in 2.1.2 to the total variance:

$$\mathbf{S/T}\,|_{\mathbf{E2E}} = \langle\{\hat{X}\}^2\rangle \,/\{\langle(X - [\{X\}])^2\rangle\}. \qquad 2.1.7$$

The second quantifies the potential predictability of the composite and involves the ratio of term 2 in 2.1.2 to the total variance:

$$\mathbf{S/T}\,|_{\mathbf{composite}} = \{X'\}^2/\{\langle(X - [\{X\}])^2\rangle\}. \qquad 2.1.8$$

Also, the total S/T is the sum of 2.1.7 and 2.1.8:

$$\mathbf{S/T}\,|_{\mathbf{total}} = \left(\langle\{\hat{X}\}^2\rangle + \{X'\}^2\right) \,/\, \{\langle(X - [\{X\}])^2\rangle\}. \qquad 2.1.9$$

It is important to note that some (but not all) components of the above variance decomposition (2.1.2) can be constrained by the observations. In particular, the following ensemble quantities involving X (the model quantity) have companion terms that can be estimated from the observations (Y). These are the total variance:

$$\{\langle(X - [\{X\}])^2\rangle\} \Rightarrow \langle(Y - [Y])^2\rangle, \qquad 2.1.10$$

the variance of the composite mean (2.1.6):

$$\{X'^2\} \Rightarrow Y'^2 \ , \qquad 2.1.11$$

and the total E2E variability (2.1.5):

$$\left\{\langle\hat{X}^2\rangle\right\} \Rightarrow \langle\hat{Y}^2\rangle \ . \qquad 2.1.12$$

We will in the following sections take advantage of 2.1.10-2.1.12 to provide some constraints on the quality of the different components of the model's variance budget.

2.2 Decomposition of the covariance

Here we examine the covariance between the ensemble mean simulated/predicted $\{X\}$ and the observed (Y) anomalies with respect to the long-term mean[ ], conditioned over all time periods making up the composite ⟨ ⟩:

$$\sigma^2_{\{X\}Y} \equiv \langle(\{X\} - [\{X\}])(Y - [Y])\rangle = \langle\{\hat{X}\}\hat{Y}\rangle + Y'\{X'\}. \qquad 2.2.1$$

Here, the first term on the RHS of 2.2.1 is the covariance associated with E2E variability, and the second term on the RHS is the covariance associated with the composite mean. By normalizing 2.2.1 by the variances (conditioned on PHEN):

$$\sigma^2_{\{X\}} = \langle(\{X\} - [\{X\}])^2\rangle$$

$$\sigma^2_Y = \langle(Y - [Y])^2\rangle$$

we obtain the conditional correlation

$$\rho_{\{X\}Y} \equiv \frac{\sigma^2_{\{X\}Y}}{\sigma_{\{X\}}\sigma_Y} = \frac{\langle\{\hat{X}\}\hat{Y}\rangle}{\sigma_{\{X\}}\sigma_Y} + \frac{Y'\{X'\}}{\sigma_{\{X\}}\sigma_Y}, \qquad 2.2.2$$

decomposed into that associated with E2E variability (first term on the RHS of 2.2.2) and that associated with the composite mean (second term on the RHS of 2.2.2). Here again the quantities are decomposed as follows: $X - [X] = X^* + \{X'\} + \{\hat{X}\}$, $X^* = X - \{X\}$, $\hat{X} = X - \langle X\rangle$, $X' = \langle X\rangle - [X]$. Analogous definitions are used for the observations Y, though of course, there is no ensemble mean, so $Y - [Y] = Y' + \hat{Y}$, where $\hat{Y} = Y - \langle Y\rangle$ and $Y' = \langle Y\rangle - [Y]$.

3. Datasets

Our focus is on the winter months of December through March within the period 1980 through 2016—a period common to MERRA-2 and the AGCM simulations described below. In the following our phenomenon of interest (PHEN) is ENSO, and we will examine El Niño and La Niña composites separately. For that purpose, the set of 11 El Niño winters is [1982/83, 1986/87, 1987/88, 1991/92, 1994/95, 1997/98, 2002/03, 2004/05, 2006/07, 2009/10, 2015/16], and the set of 9 La Niña winters is [1984/85, 1988/89, 1995/96, 1998/99, 1999/00, 2000/01, 2007/08, 2010/11, 2011/12]. The following gives a brief overview of MERRA-2 and the GEOS AGCM simulations.

3.1 MERRA-2

The atmospheric reanalysis data used for this study is the Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2; Gelaro et al. 2017). MERRA-2, developed by NASA Goddard Space Flight Center (GSFC) / Global Modeling and Assimilation Office (GMAO), is an updated version of MERRA (Rienecker et al. 2011) that includes an improvement of the assimilating model's physical parameterizations of moist processes, turbulence, land and ocean surface processes, and gravity wave drag (Bosilovich et al. 2015; Molod et al. 2015; Gelaro et al. 2017; see also below). Other differences from MERRA include aerosol data assimilation, new developments in the representation of ozone, and the use of precipitation observations to force the land surface. The horizontal resolution of the MERRA-2 data is 0.625° longitude × 0.5° latitude. The key variables used here are 2-meter air temperature (T2m), precipitation, zonal wind, and geopotential height. We note, however, that the MERRA-2 precipitation used in this study for verification is an observationally-corrected product in which the precipitation generated by the atmospheric model underlying MERRA-2 was scaled to agree with gauge and satellite precipitation observations (Reichle et al. 2017).

In the following, we will use the words observations and MERRA-2 interchangeably with, of course, the understanding that MERRA-2 is a reanalysis product that combines a model-based first guess with observations, meaning that the reanalysis products are potentially impacted by model biases. However, of the quantities considered here, model biases are typically only a major problem for the precipitation product, which, again, is observationally-corrected.

3.2 M2AMIP simulations

The GEOS AGCM used here (Molod et al. 2015; Gelaro et al. 2017) includes the finite-volume dynamical core of Putman and Lin (2007), which uses a cubed sphere horizontal discretization at an approximate resolution of 0.5° × 0.625° and a vertical resolution

consisting of 72 hybrid-eta levels from the surface to 0.01 hPa. Recent upgrades to the physical parameterization schemes include increased re-evaporation of frozen precipitation and cloud condensate, changes to the background gravity wave drag, and an improved relationship between the ocean surface roughness and ocean surface stress (Molod et al. 2015). The model also includes a Tokioka-type trigger on deep convection as part of the Relaxed Arakawa-Schubert (RAS, Moorthi and Suarez 1992) convective parameterization scheme, which governs the lower limit on the allowable entrainment plumes (Bacmeister and Stephens 2011). A new glaciated land representation and seasonally-varying sea ice albedo have been implemented, leading to improved air temperatures and reduced biases in the net energy flux over these surfaces (Cullather et al. 2014). The model includes the catchment land surface model developed by Koster et al. (2000).

An ensemble of ten AMIP[2]-style simulations were performed for the time period spanning January 1980 through December 2016 (Collow et al. 2017). Initial conditions for each ensemble member were taken from different days of November 1979 during the MERRA-2 spin up period. An important feature of these simulations is that they were produced with the same version of the GEOS AGCM as was used to produce MERRA-2. In addition, they were run with the same SST data, greenhouse gases (GHGs), and other forcing as in MERRA-2, and they used the MERRA-2 grid resolution. Indeed, these simulations differed from MERRA-2 only in that they did not assimilate observations. This similarity offers the unique opportunity to assess how the observations influence various aspects of the model climate, though, of course, we note again that model errors may be implicit in the reanalysis itself, potentially clouding our interpretations.

We also take advantage of an additional set of 5 AMIP-style simulations that, while otherwise having the same setup as above, are forced with the SST and sea ice concentration used in the European Centre for Medium-Range Weather Forecasts interim reanalysis (Marquardt Collow et al. 2020). It should be noted that Marquardt Collow et al. (2020) found substantial differences in the sea ice concentrations in the two forcing datasets (from MERRA-2 and ERA-Interim), and that this results in substantial differences in the simulations over the Arctic—something to keep in mind as we analyze our results based on the combined set of 15 simulations. However, our focus is primarily on the middle latitudes and the tropics where the differences in the two sets of simulations appear to be minimal. In the following, we shall refer to the 15 AMIP runs made with this model as the *M2AMIP* simulations.

4. Results for ENSO during boreal winter

In the following (based on all 15 M2AMIP ensemble members), we examine the variance of the 250mb eddy (deviation from the zonal mean) height, the 250mb zonal wind, and the precipitation over a large region extending from 120°E to 0°W and from 30°S northward to

[2] AMIP stands for the Atmospheric Model Inter-comparison Project as described in Gates et al. (1992), but the acronym is now used more generally to refer to any long AGCM simulations forced with observed SST.

the pole—a region tailored to encompass the main ENSO impacts. Additional plots focus on the precipitation and two-meter air temperature (T2m) over North America.

Section 4.1 examines the various terms (the budget) of the variance decomposition to get an overall sense of the magnitudes and spatial distributions of the different budget terms (eq. 2.1.2) and any differences between El Niño and La Niña events. Next we look in more detail into the character and quality of the model's composite variance (in section 4.2) and the E2E variance (section 4.3). This is followed by an examination of signal-to-total variance ratios (section 4.4) and correlations with observations (section 4.4).

4.1 The variance budget

We present in Figs. 1-5 the various terms of the variance budget for each of several meteorological variables. Each figure is organized to show the individual components of the budget (see 2.1.2) in the top 4 panels, with the bottom two panels comparing the model's total variance with the observed/MERRA-2 variance, thereby providing an indication of the quality of the model's overall variability. We focus on the average of the variances for the four months of the extended winter season, December through March (DJFM)[3]. Such an average, we believe, should provide robust estimates of the various terms in the budget, though we acknowledge that there are substantial within-season (monthly) changes in the statistics (e.g., Chen et al. 2017; Lim et al. 2020). We will address some aspects of the within-season month-to-month differences in Section 4.5.

Figure 1 shows the DJFM mean variance decomposition for the 250 mb eddy height during El Niño events (Fig. 1a) and La Niña events (Fig. 1b). Term 2 (the composite signal) is simply the square of the ensemble mean of the composite mean anomaly, averaged over the 4 months. Term 1 indicates how much variability there is in the composite as a result of having different ensemble members (the noise in the composites). Clearly that noise variability (term 1) is small compared to the associated composite signal (term 2). The strength of the signal (versus that of the noise) will be quantified in Section 4.4 in terms of the ratio of the signal to the total variance. Term 3 is the signal (the variance about the ensemble mean) associated with individual El Niño/La Niña events (the E2E signal), while term 4 is the average intra-ensemble variance (E2E noise) during the El Niño/La Niña events. The E2E noise variability (term 4) is clearly the largest component of the variability in middle and high latitudes for both El Niño and La Niña.

Comparing the El Niño and La Niña model results (Figs. 1a and 1b) we see a remarkable similarity in both the pattern and the magnitude of the composite signals (term 2, upper right panels). There is less similarity in the E2E variance, with the E2E noise (term 4) substantially larger over the North Pacific for La Niña (cf. middle right panels of Figs 1a

[3] To be clear, the various budget terms are computed separately for each month, and then averaged. The choice of an extended winter (including March) is motivated by the work of Chen et al. (2017) and also helps to increase the sample size.

and 1b). Also, the La Niña E2E signal is farther west (over the Northwest Pacific) compared with that of El Niño, though of similar magnitude (cf. middle left panels of Figs 1a and 1b). The comparison with the MERRA-2 total variance (bottom panels of Figs. 1a and 1b) shows overall reasonable simulations of the variance pattern, although the free-running model overestimates the total variance. Another key difference is that the maximum over the North Pacific is situated too far west, especially for El Niño (bottom panels of Fig. 1a) for which the simulated maximum is centered over the Aleutian Islands whereas the MERRA-2 maximum values are to the southeast of that, just off the west coast of North America. It is likely that the excessive total variance in the North Pacific during both El Niño and La Niña is mostly due to excessive E2E noise (term 4 in Figs. 1a and b). We will come back to that in section 4.3. There is also excessive total variance over much of western Canada, especially for La Niña. It is also noteworthy that the composite signal (term 2) is somewhat larger than the E2E signal (term 3), with a tendency for the two signals to be somewhat phase-shifted with respect to each other.

In summary, the above results indicate that the model has excessive E2E noise. In addition, the model appears to be too linear in the sense that the differences in the total variances over the North Pacific between El Niño and La Niña seen in the observations (cf. lower right panels of Figs 1a and 1b) are much less apparent in the M2AMIP results (cf. lower left panels of Figs 1a and 1b). As such, the model appears to do better in reproducing the observed La Niña variance than the El Niño variance, reflecting the failure of the model to reproduce the more eastward location (just south of the Gulf of Alaska) of the maximum variability during El Niño— a deficiency that presumably affects the model's ability to predict El Niño impacts over North America. Nevertheless, the model produces a difference in the magnitude of the E2E noise over the North Pacific/western North America region (less for El Niño, more for La Niña) that is not inconsistent with recent studies indicating that El Niño impacts over North America are more predictable than those of La Niña (e.g., Chen et al. 2017). On the other hand, there seems to be little difference between El Niño and La Niña in either the composite or the E2E signals, something we will come back to in sections 4.2 and 4.3.

Figure 2 is the same as Fig. 1, except for the 250 mb zonal (u) -wind. For El Niño (Fig. 2a), the comparison between the total variance from the model and MERRA-2 (bottom two panels) shows that the maximum model variance in the North Pacific is larger than the observed and located too far west. The model does better for La Niña, with the location of the maximum variance to the northwest of Hawaii similar to the observed, though again the simulated maximum is too large (Fig 2b, bottom panels). For both El Niño and La Niña the zonal wind signal is predominately associated with the composite variance (term 2), while the noise (which is the largest component of the total variability) is for the most part associated with E2E variance (term 4). Also, it is worth noting that the region of maximum total variability in the North Pacific for the model is remarkably similar for El Niño and La Niña, again reflecting an unrealistic linearity of the model results. Over North America, it is clear that the signal associated with the composite mean (term 2) is the best hope for

skillful forecasts associated with both El Niño and La Niña, with the potential for skill largely confined to the southern tier of states and the border between the U.S. and Canada (top right panels of Figs 2a and 2b). The E2E signal (term 3) is especially weak over North America for La Niña (left middle panel of Fig. 2b).

Figure 3 is the same as Fig. 1, except for precipitation. Focusing first on El Niño (Fig. 3a), a striking feature of the total variability is the model's excessive variance in the tropical Pacific (cf. the bottom panels of Fig. 3a). The total precipitation variance in the tropical Pacific seems to have roughly equal contributions from the composite signal (Term 2), E2E signal (Term 3) and E2E noise (Term 4), and it is unclear which of these are excessive. We will come back to that in sections 4.2 and 4.3. MERRA-2 also shows a split in the maximum variance in the tropical Pacific, with the southern branch presumably linked to variations in the South Pacific Convergence Zone (SPCZ); such a southern branch is not evident in the model results. It is noteworthy that the maximum of the tropical Pacific precipitation E2E signal (term 3, Fig. 3a) is situated to the east of the maximum in the composite (term 2, Fig 3a), whereas the E2E noise (term 4, Fig. 3a) is located to the west of the composite maximum, suggesting these different sources of precipitation variability may lead to different (in particular, phase shifted) atmospheric responses over the North Pacific and North America.

Turning next to La Niña (Fig. 3b), we see that the total observed (MERRA-2) precipitation variance is more confined to the central and western tropical Pacific (compared with El Niño), with a band of higher variance extending from about 150E into the eastern Pacific just north of the equator and a shorter band just south of the equator, with the latter again (as we saw for El Niño) not captured by the model (cf. the bottom panels of Fig. 3b). In contrast to the results for El Niño, the model shows a more realistic level of total variance in the tropics, though the variance in the eastern tropical Pacific is somewhat excessive. Here, the maxima in the tropical Pacific of both the E2E signal (term 3) and E2E noise (term 4) lie to the west of the maximum in the composite signal (term 2).

The above precipitation results suggest that it is likely that the excessive variance noted earlier in the height and zonal wind over the North Pacific may in fact be driven by the excessive tropical Pacific precipitation (heating) variance, especially for El Niño (we will look more into the nature of the tropical precipitation variance in our discussions of Figs. 8 and 13). It is also noteworthy that for both El Niño and La Niña, much of the total variability over the Northern Hemisphere extratropical oceans, the west coast of North America, and the eastern United States appears to be due to E2E noise (term 4, middle right panels of Figs. 3a and b). In fact, with the contours used in Fig. 3, there is little evidence of any precipitation signal (composite or E2E) over North America. As such, we next focus more closely (i.e., with more refined contouring) on the North American region.

Figure 4 shows the results for precipitation over North America for both El Niño (Fig. 4a) and La Niña (Fig. 4b). We see that for both El Niño and La Niña, the variance over North

America is indeed primarily composed of E2E noise (term 4, right middle panels), with the largest values occurring over the southeastern U.S. and along the west coast extending from California north to Alaska. Also, for both El Niño and La Niña, the composite signal (Term 2) is considerably weaker and confined to a narrow region of the southeast U.S as well as along the west coast. The E2E signal (Term 3) is overall similar in both coverage and magnitude compared to that of the composite signal, although it has a greater regional extent over the southeast. There do appear to be some rather subtle differences between El Niño and La Niña, with La Niña showing somewhat greater spatial extent in the southeast composite signal (term 2, Fig. 4b), and El Niño showing an E2E signal maximum (term 3, Fig. 4a) that is more focused on the southeast including Florida (in the case of La Niña the maximum is located further to the west centered on Alabama). Similarly, for the E2E noise (Term 4), the maximum for La Niña is shifted somewhat to the west and north compared to that for El Niño, which is more centered on the southeast (cf. middle right panels of Figs. 4a and 4b). It is noteworthy that the magnitude of the composite noise (term 1), while small compared to the E2E noise (term 4), is similar in magnitude to that of the composite and E2E signals for both El Niño and La Niña. Overall, the total precipitation variance over North America is well simulated by the model for both El Niño and La Niña (bottom panels of Figs. 4a and 4b). In fact, there is remarkably little difference between the El Niño and La Niña precipitation variability over North America for both the model and MERRA-2, with some evidence of somewhat stronger variability over the southeastern U.S. for El Niño. This is surprising, given the substantial differences between El Niño and La Niña seen (particularly for the MERRA-2) for the eddy height and zonal wind variance just upstream of and over North America (Figs. 1 and 2).

Turning next to T2m over North America (Figure 5), we see that again (as for precipitation), the variance is dominated by E2E noise variability for both El Niño and La Niña (term 4, middle right panels of Figs 5a and 5b). The E2E noise is, for both warm and cold events, largest along a band extending from Alaska, across Canada, and into the eastern U.S., though it is somewhat larger and extends farther into the southeast U.S. for La Niña. These El Niño versus La Niña differences in the noise are reflected in both the model and the MERRA-2 total variances, though the model total variances are generally larger than the MERRA-2 values. The E2E signal (term 3) is spatially rather uniform (increasing to the north) but overall weak for both El Niño and La Niña (left middle panels of Figs. 5a and 5b). In contrast, the composite signal (Term 2) is somewhat larger but confined to the southeastern U.S. and northwest Canada and Alaska (top right panels of Figs. 5a and 5b). Surprisingly, the composite signal tends to be larger for La Niña than for El Niño (we will come back to that in section 4.2). Overall, the total variance (bottom panels of Figs. 5a and 5b) is reasonably well simulated but is somewhat larger than observed, with some evidence for somewhat larger values for La Niña compared with El Niño.

Comparing the results for precipitation and T2m, we see evidence of spatially more extensive and larger (relative to the noise) signals for T2m (see also sections 4.3 and 4.4). This is consistent with numerous studies that show greater forecast skill for T2m than

precipitation on subseasonal and seasonal time scales over the U.S. (e.g., Wang and Robertson 2018). This is, however, not a universal result, with, for example, Chen et al. (2017) finding greater skill in the North American Multimodel Ensemble (NMME) probability forecasts of the boreal winter composite ENSO precipitation patterns than found for the forecasts of temperature patterns. It is also noteworthy that while there is evidence for generally greater noise (primarily from E2E variability) during La Niña compared with El Niño for T2m (consistent with greater predictability for El Niño), this is not the case for precipitation, for which, if anything, the opposite seems to be the case.

Further insights into the quality of the M2AMIP variance budget, as well as insights into the reasons for the seemingly contradictory results of previous studies (mentioned above) concerning the relative skill of predicting precipitation and T2m over North America, can be gained by a more detailed look at the individual terms in the variance budget. As such, we next take a closer look at the composite (section 4.2) and E2E (section 4.3) variability, focusing in particular on comparisons with analogous results from MERRA-2.

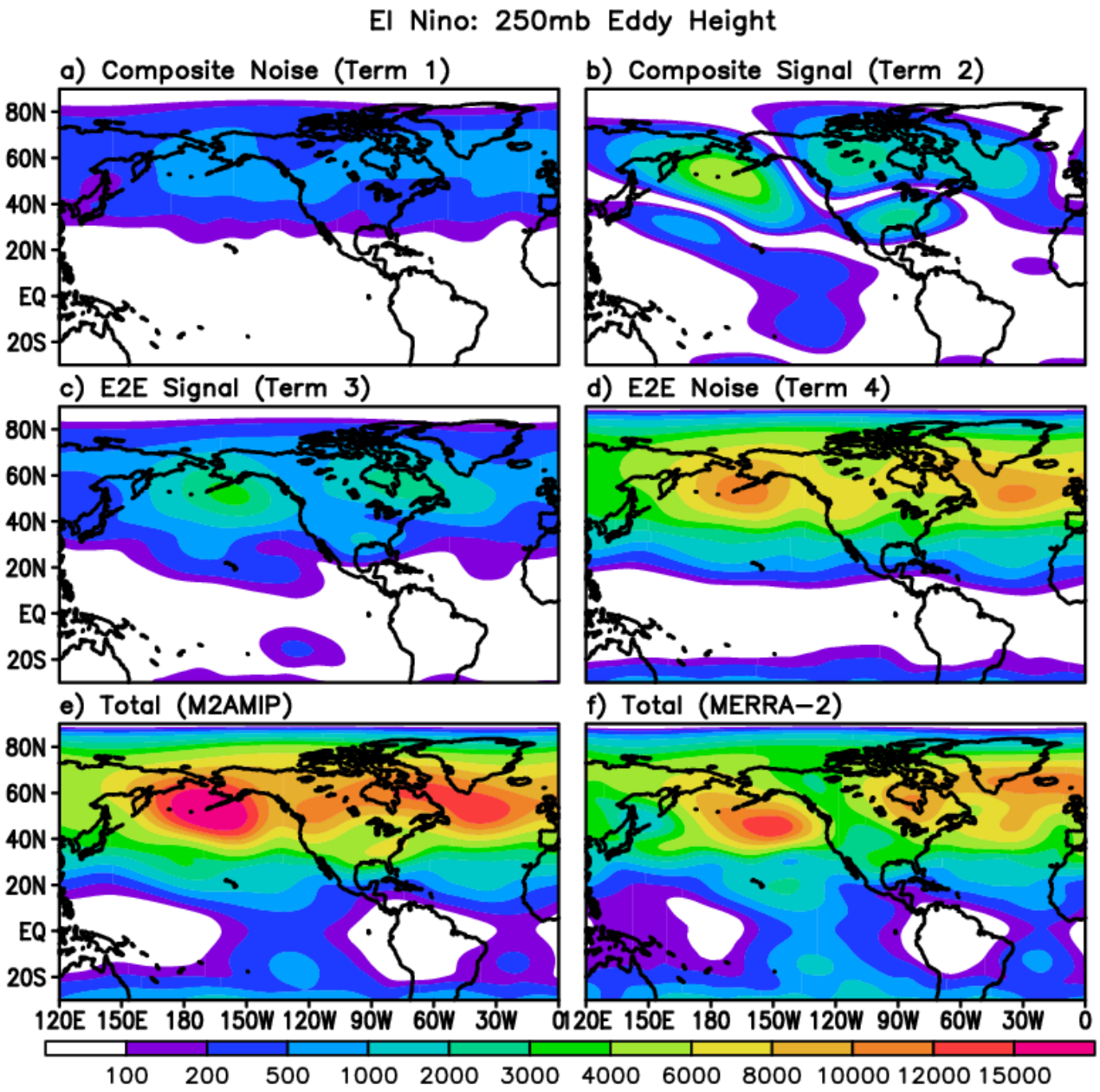

Figure 1a: The variance decomposition of the 250mb eddy height for El Niño events that occurred during 1980-2016, averaged over the months December through March, with terms defined in eq. 2.1.2. See text for details. Units are m^2.

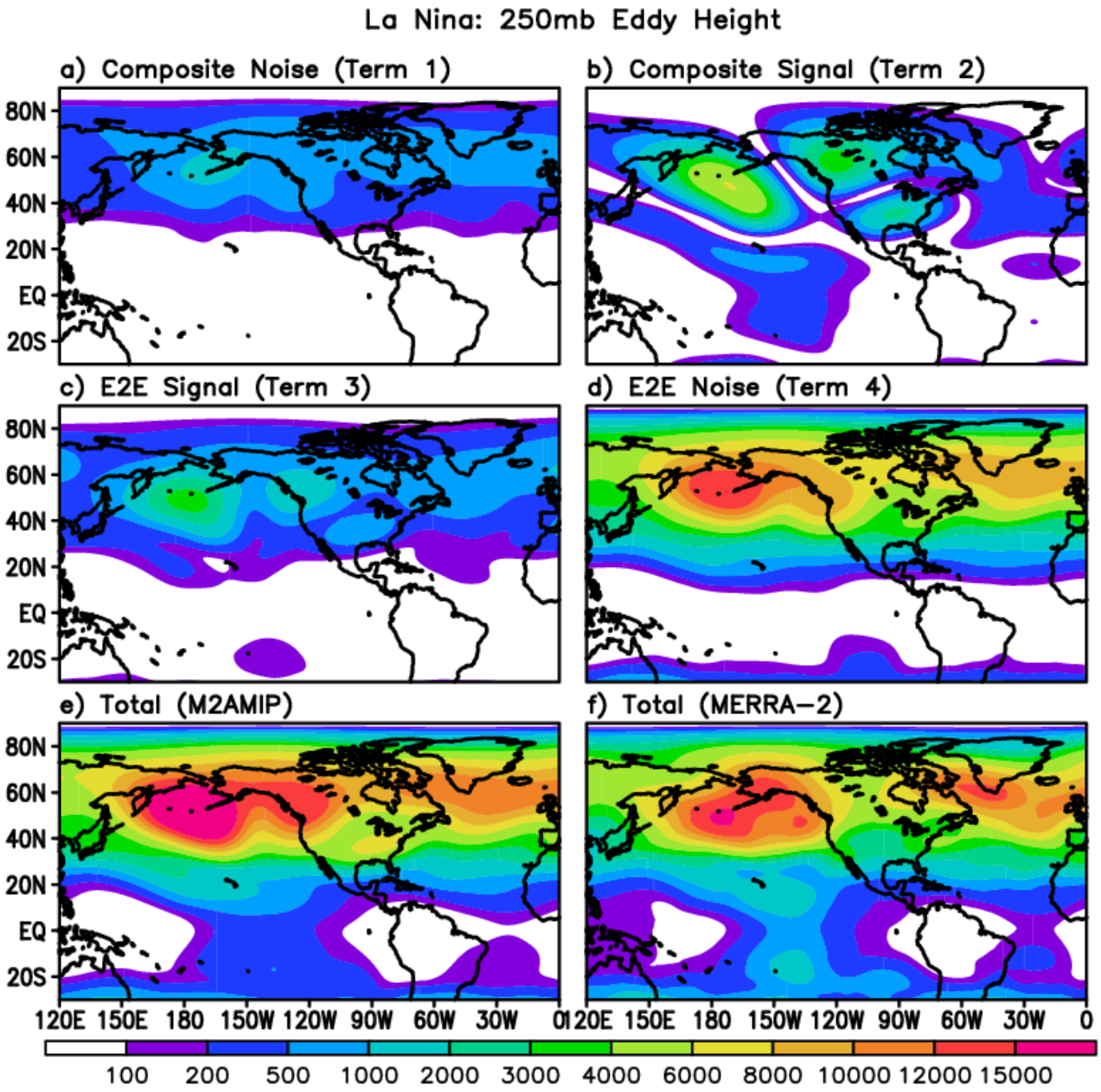

Figure 1b: Same as Fig. 1a, but for La Niña.

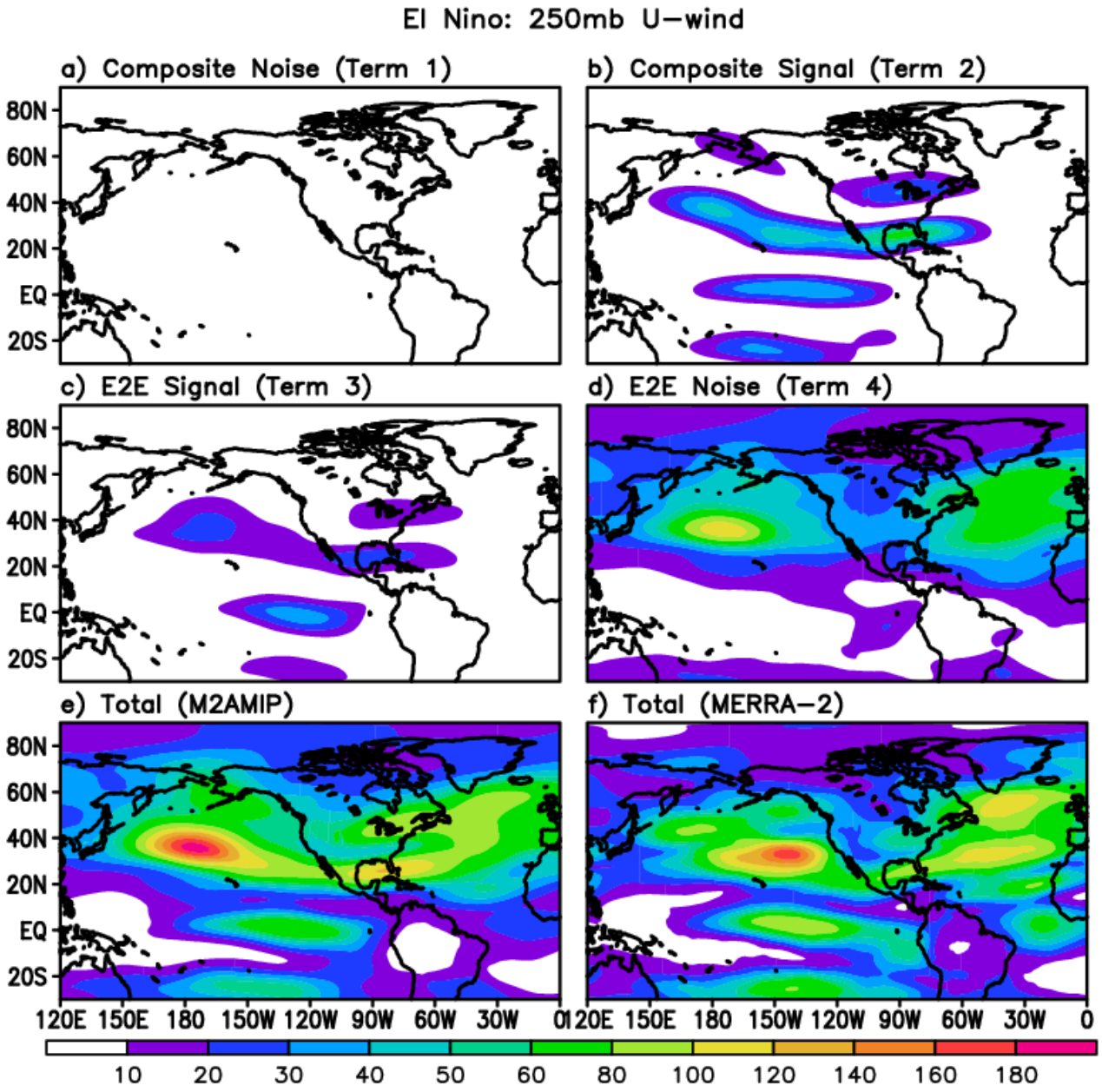

Figure 2a: The variance decomposition of the 250mb u-wind for El Niño events that occurred during 1980-2016, averaged over the months December through March, with terms defined in eq. 2.1.2. See text for details. Units are m^2/s^2.

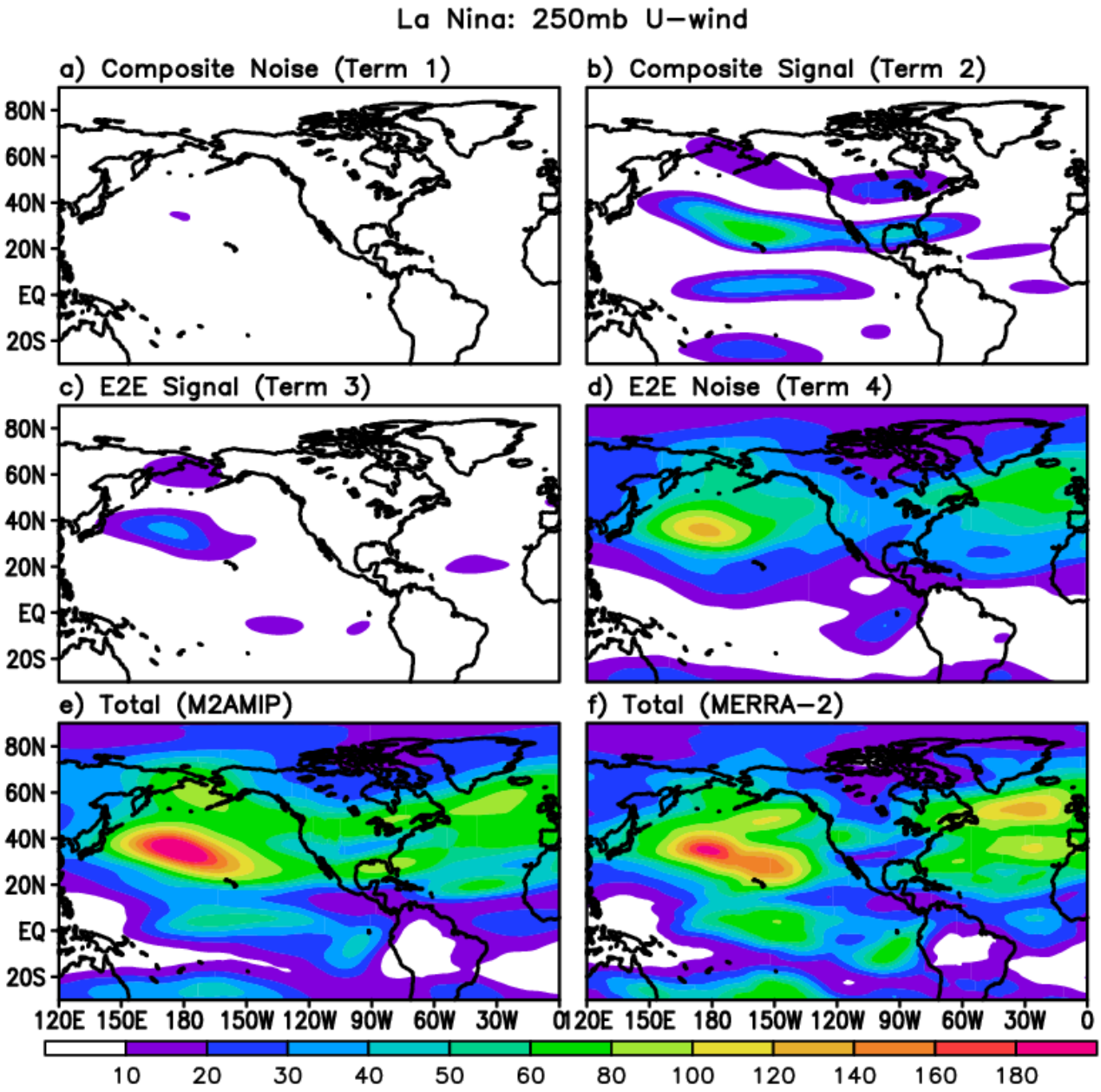

Figure 2b: Same as Fig. 2a, but for La Niña.

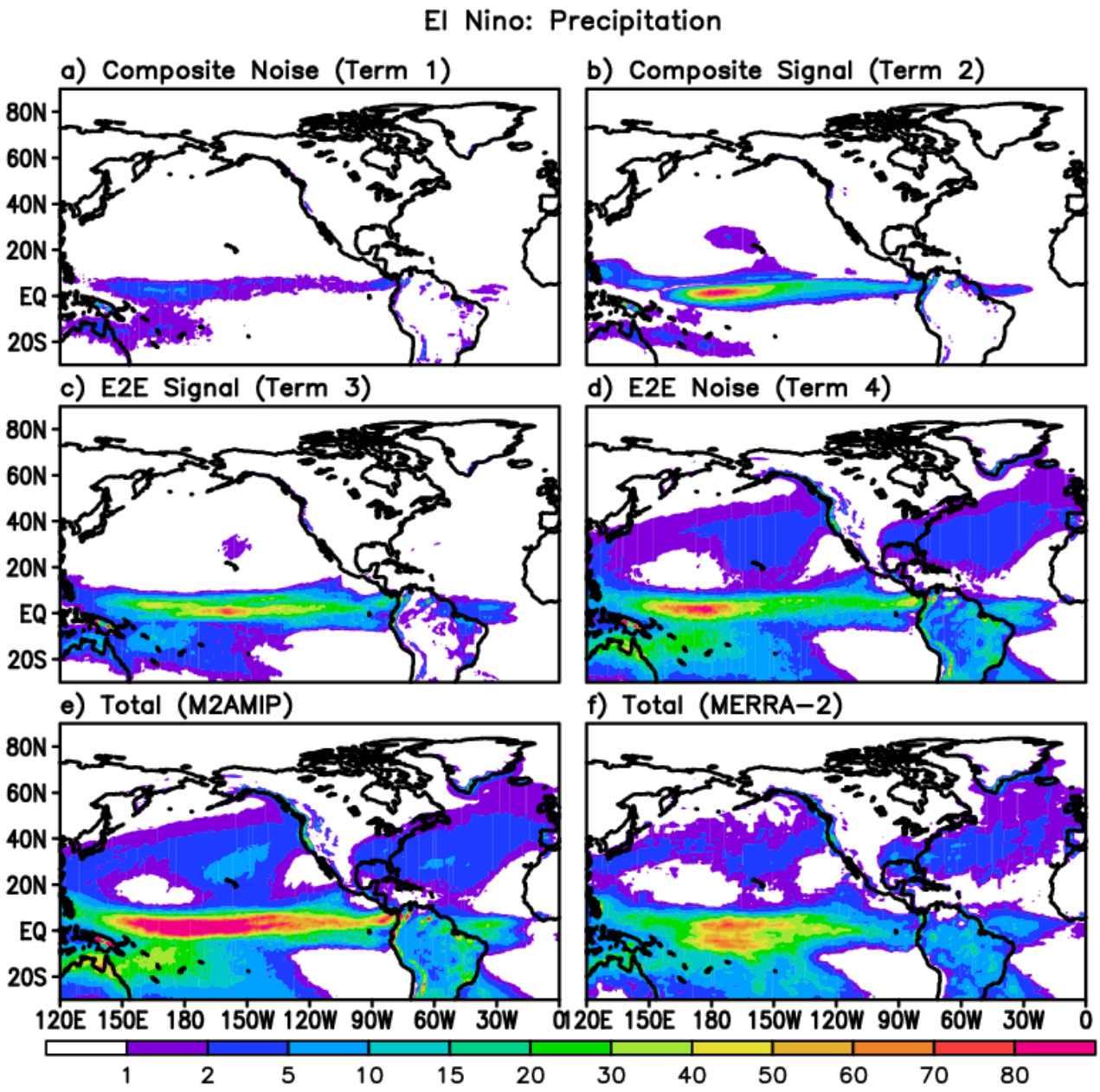

Figure 3a: The variance decomposition of the precipitation for El Niño events that occurred during 1980-2016, averaged over the months December through March, with terms defined in eq. 2.1.2. See text for details. Units are mm^2/day^2.

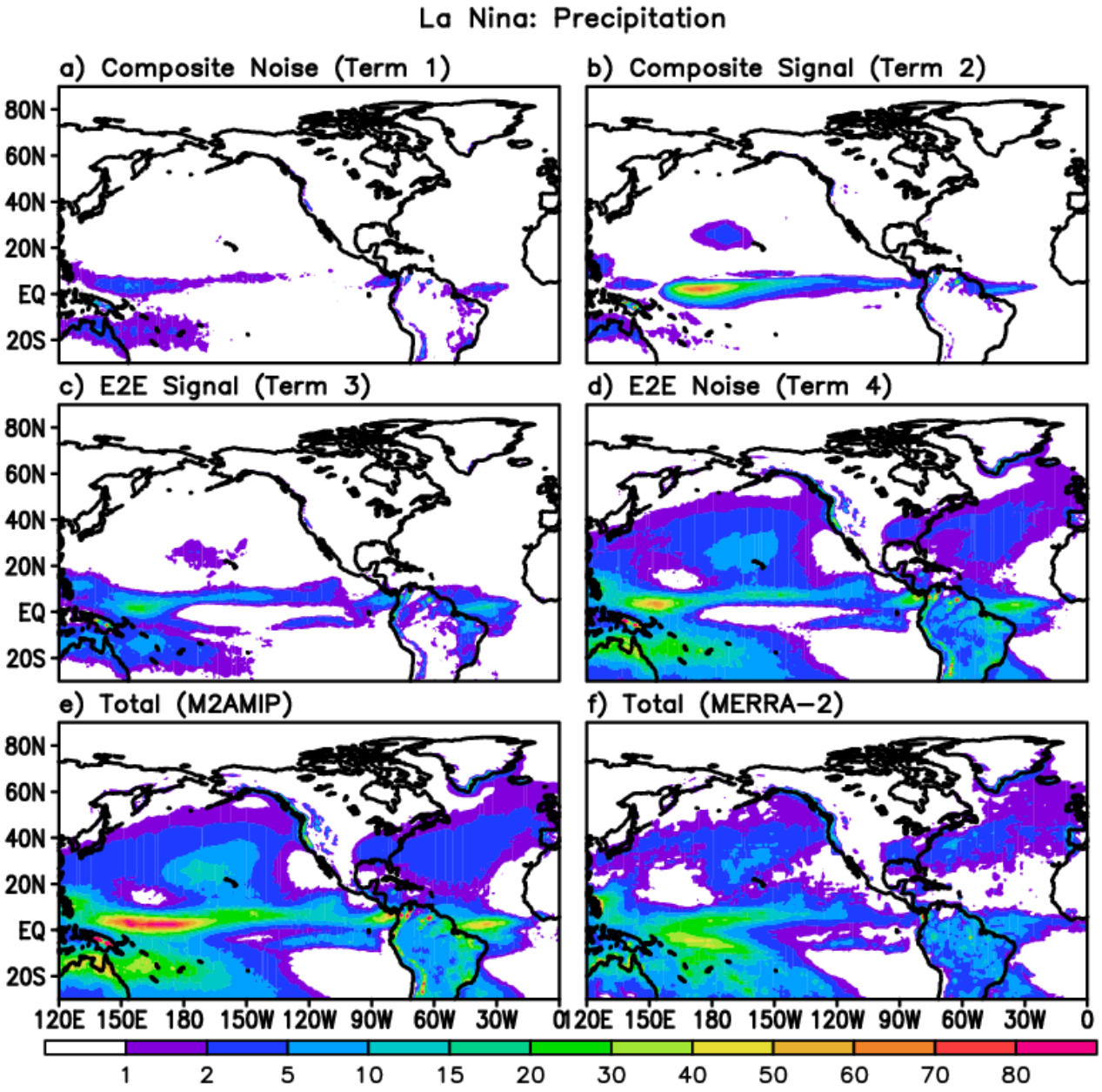

Figure 3b: Same as Fig. 3a, but for La Niña.

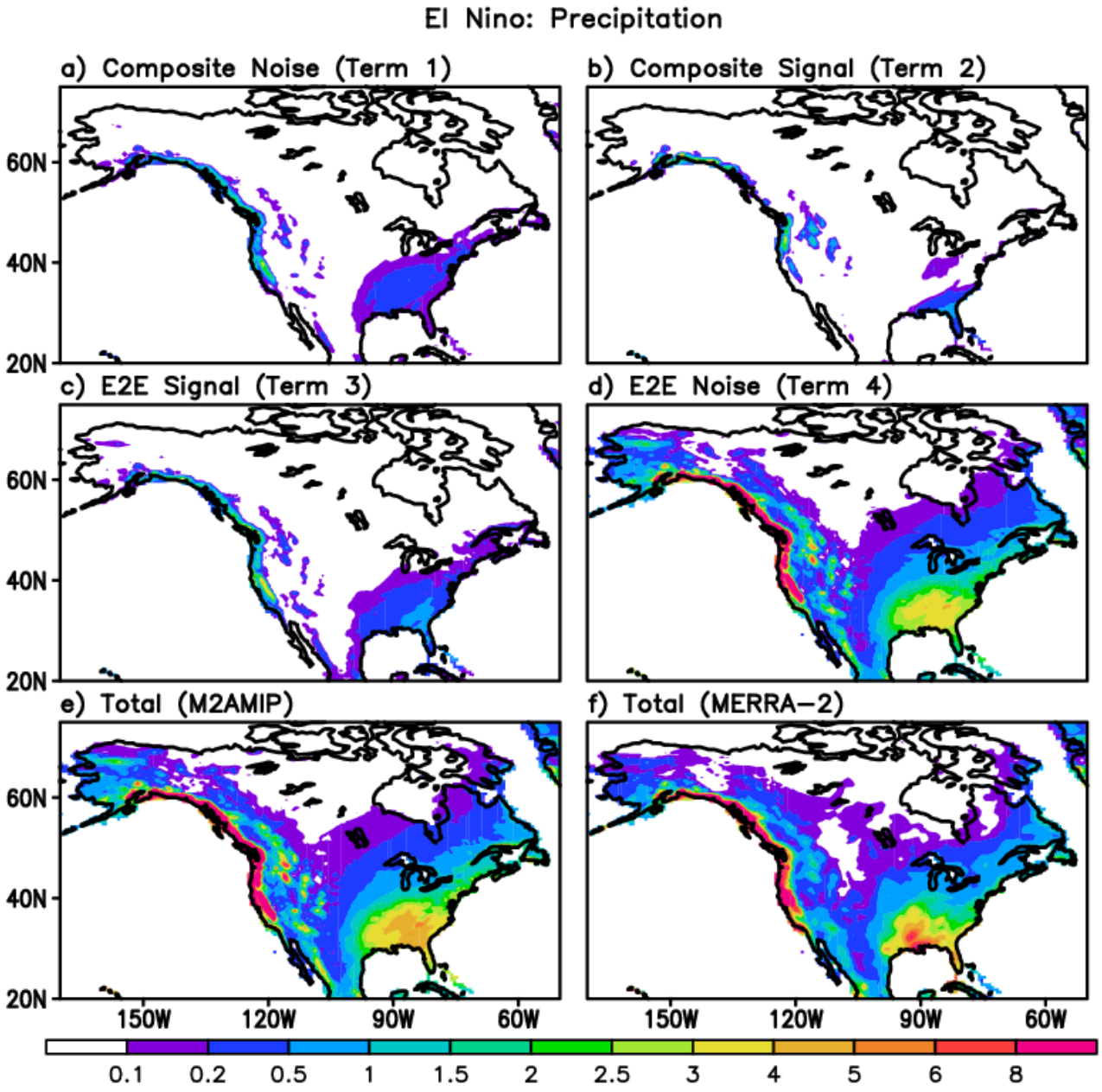

Figure 4a: The variance decomposition of the precipitation over North America for El Niño events that occurred during 1980-2016, averaged over the months December through March, with terms defined in eq. 2.1.2. See text for details. Units are mm^2/day^2.

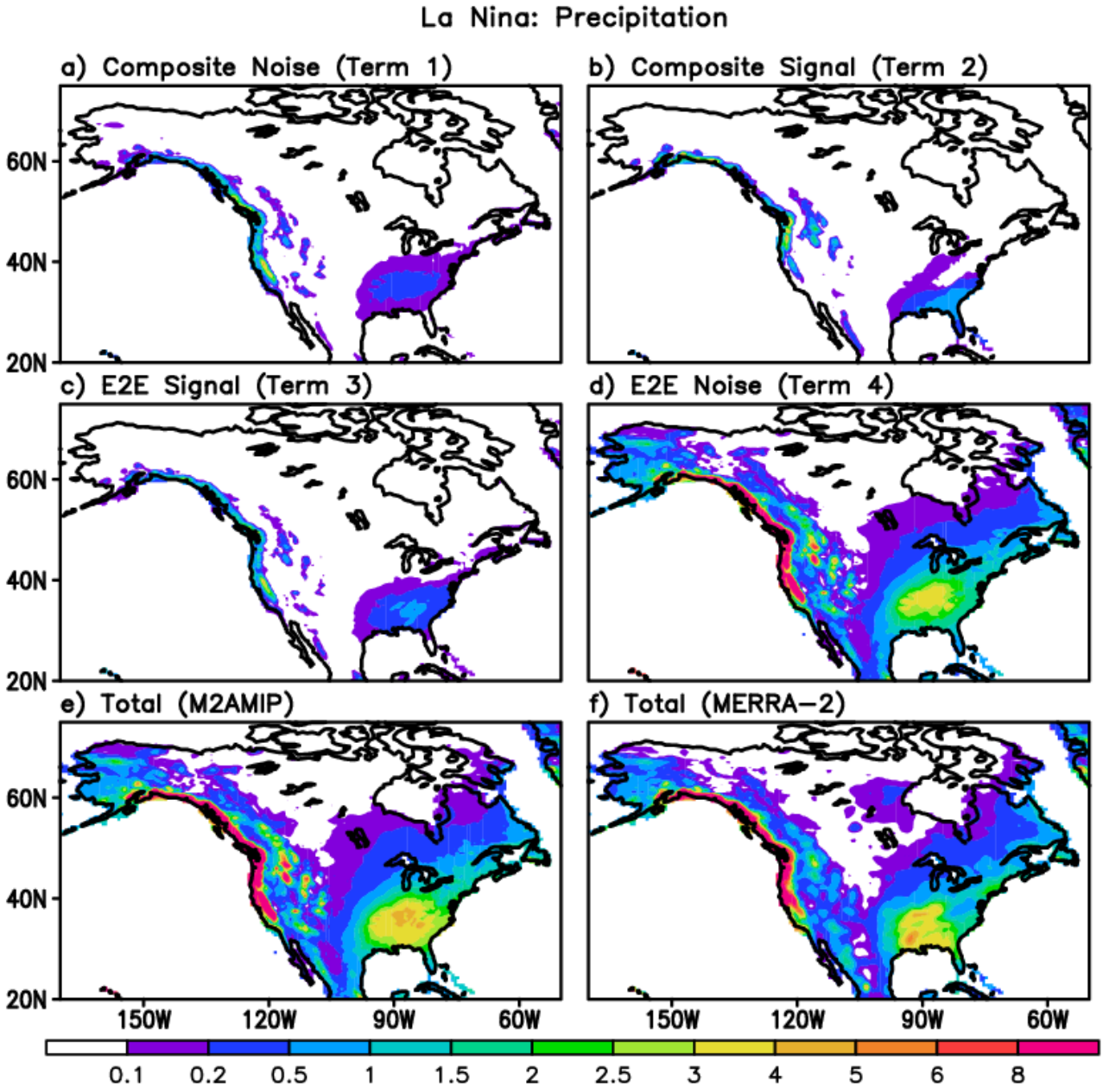

Figure 4b: Same as Fig. 4a, but for La Niña.

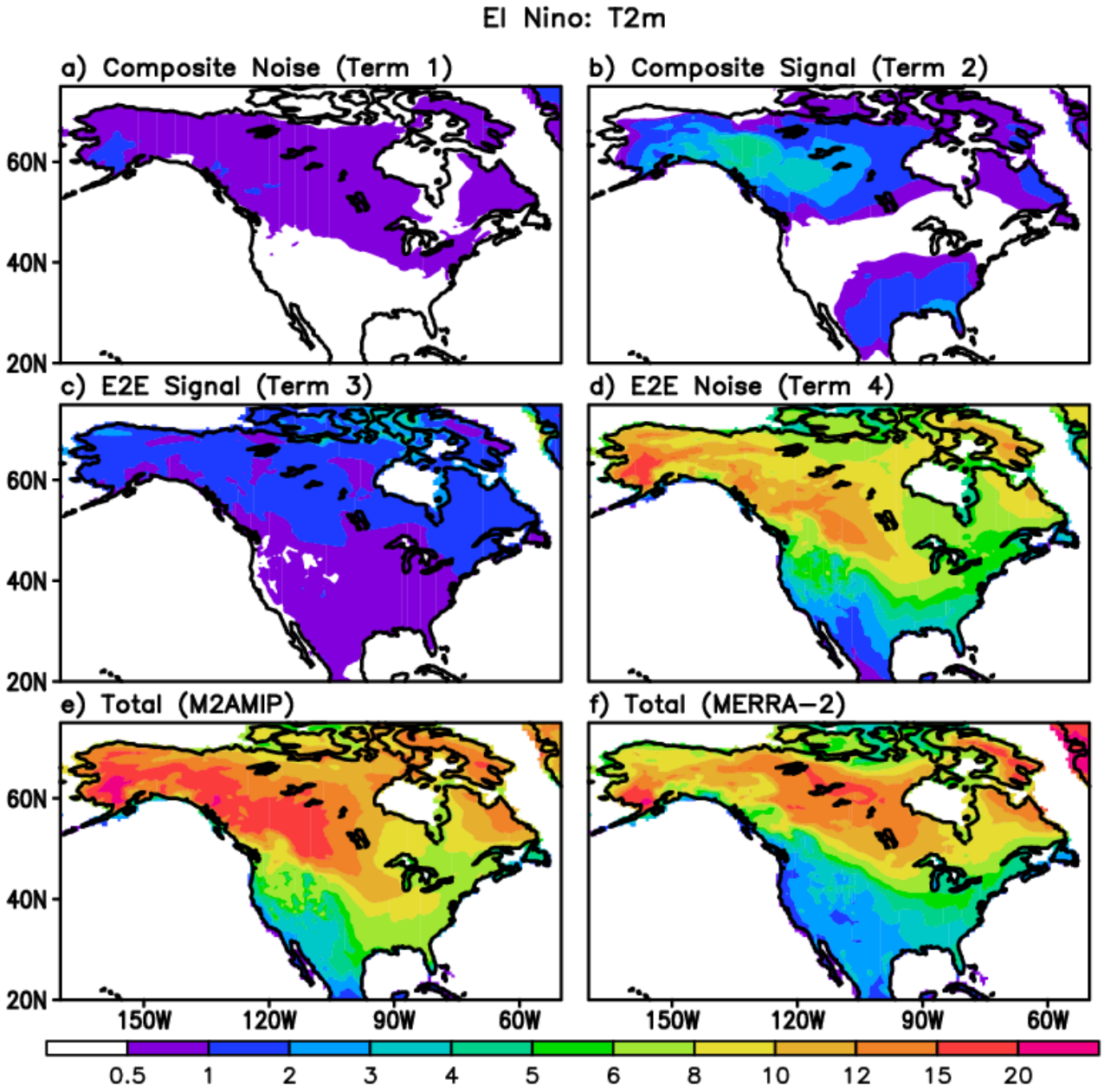

Figure 5a: The variance decomposition of T2m over North America for El Niño events that occurred during 1980-2016, averaged over the months December through March, with terms defined in eq. 2.1.2. See text for details. Units are $°K^2$.

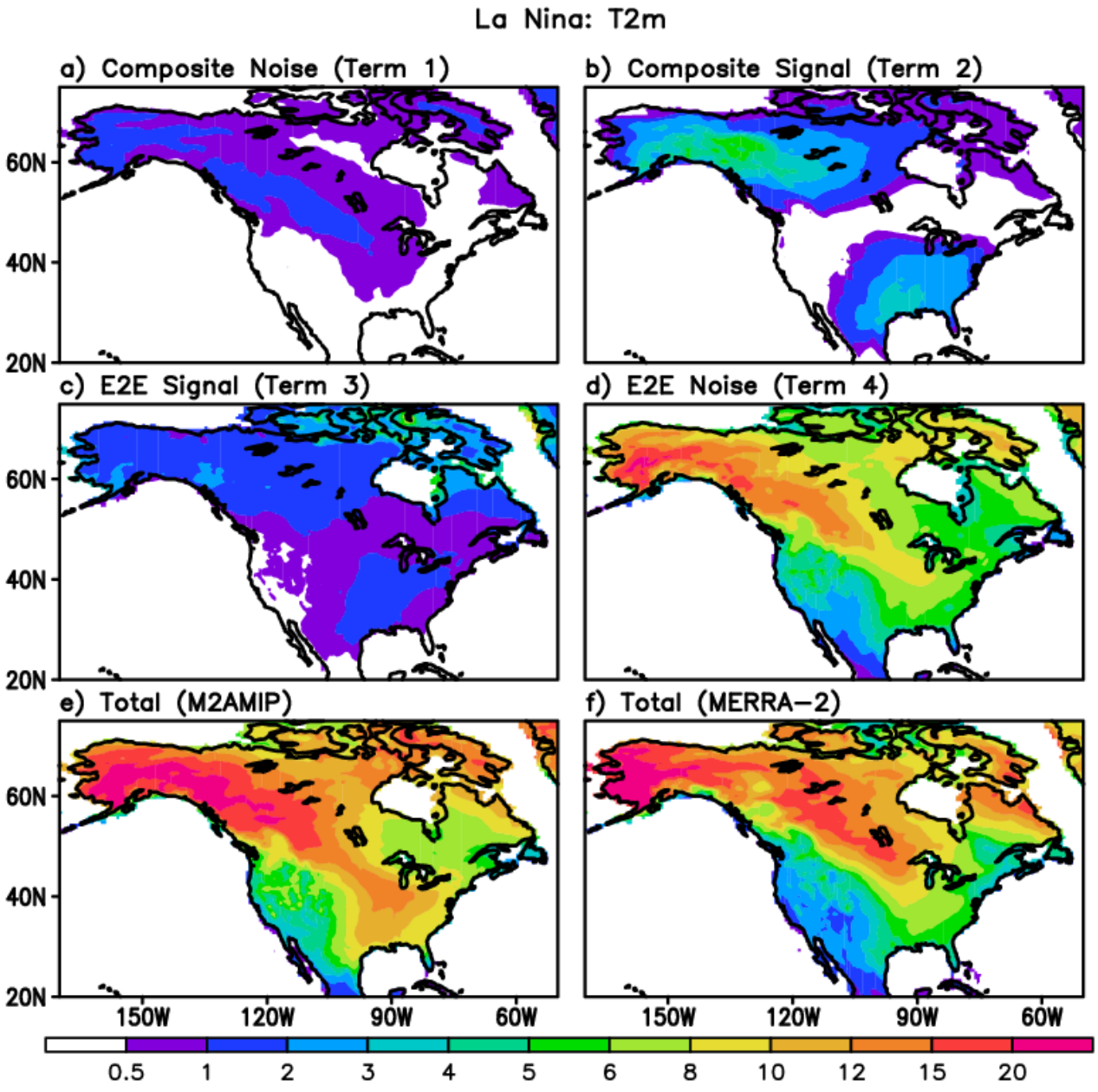

Figure 5b: Same as Fig. 5a, but for La Niña.

4.2 The composite variance

In section 2.1 we showed that the model's total variance of the composite mean is the sum the composite noise and the composite signal (2.1.6), and furthermore that it is this sum that is the appropriate quantity to compare with the composite variance estimated from MERRA-2. In the case of MERRA-2, this is just the square of the single composite mean anomaly, whereas for M2AMIP this is the ensemble mean of the square of the composite mean anomalies computed from the individual ensemble members (2.1.11).

Figure 6 shows the composite variance for the 250mb eddy height for both El Niño (top panels) and La Niña (bottom panels). The results for both the model (left panels) and MERRA-2 (right panels) reflect the well-known response to the tropical Pacific SST associated with ENSO, with the strongest anomalies occurring over the North Pacific and North America. There are however important differences between the model results and those based on MERRA-2. MERRA-2 shows generally stronger composite variance for El Niño compared with La Niña over the North Pacific and North America: the composite variance for La Niña is especially weak and disorganized over North America. This is in contrast with the model results (right panels), which shows remarkably similar variances for El Niño and La Niña, including over North America. As such, the model appears to produce a too strong wave response to ENSO over North America, especially for La Niña, though we must keep in mind that this conclusion is based on a limited number of ENSO events. We will come back to this in Section 4.5, where we show this leads to a mismatch between skill and signal strength in this region. Furthermore, the overall (southwest/northeast oriented) structure of the wave response over the North Pacific in the model (versus a more north/south orientation in MERRA-2) is suggestive of a too strong response to SST in the western tropical Pacific warm pool region. These results again reflect a level of linearity in the model's extratropical response to El Niño and La Niña that is inconsistent with the observations.

Fig. 7 is the same as Fig. 6, but for the 250mb zonal wind. We see that the model does an overall reasonable job of reproducing the main features of the composite zonal wind variance throughout the North Pacific and North America. Here we point out in particular the verisimilitude of the regions of relatively large variance over the Great Lakes region, and along northern Mexico and the southern tier of states. There is some evidence for somewhat larger composite zonal wind variance for El Niño than for La Niña over these regions for both MERRA-2 and the model, though the model results are again more linear than those based on MERRA-2.

Turning next to the composite precipitation variance (Fig. 8), we see that the model tends to produce excessive composite variance in the intertropical convergence zone (ITCZ) at the eastern edge of the tropical Pacific warm pool region (near the dateline just north of the equator), but fails to reproduce the MERRA-2 increased variance south of the equator—a variance that (as mentioned previously) is likely associated with changes in the SPCZ. It

appears plausible that both the model's more southwest/northeast oriented structure of the eddy height variance over the North Pacific (versus a more north/south structure for MERRA-2), and the excessive composite eddy height variance over North America mentioned previously, is a response to the excessive ITCZ composite precipitation variance in the western tropical Pacific warm pool region. Furthermore, it appears that the excessive total precipitation variance in the tropical Pacific discussed in Section 4.1 is at least in part the result of excessive composite variance, which we know from Fig. 3a is primarily made of the composite signal. We will look more into the contribution of the E2E variance to the excessive tropical Pacific total precipitation variance in the next section.

Focusing on the composite precipitation variance over North America (Fig. 9), we see that the model does a reasonable job of reproducing the composite variability for both El Niño (top panels) and La Niña (bottom panels), with the largest values occurring on the west coast and the southeast. In the southeast, the model tends to produce a somewhat more expansive region (compared with MERRA-2) of variability for both El Niño and La Niña, though much of that is associated with composite noise (cf. the top panels of Figs 4a and 4b).

In the case of the T2m over North America (Fig. 10), we see that the model substantially overestimates the composite variability. This is especially true for much of the eastern U.S., and in the case of La Niña, also much of Canada. Looking back at Figs. 5a and b (top panels) we see much of this overestimate in the T2m composite variance is tied to the composite signal. Furthermore, to the extent that the T2m composite variability is tied to the upper level circulation, this appears to reflect the unrealistically large composite 250mb eddy height variance over North America discussed above (also primarily consisting of the composite signal, cf. the top panels of Figs 1a and b). Recall that this overestimate in the composite eddy height variance is especially evident for La Niña (cf. the bottom panels of Fig. 6).

The above results provide some clues as to the aforementioned (section 4.1) apparently contradictory results concerning the relative skill of predicting precipitation and T2m over North America. The model clearly produces unrealistically large T2m composite signals, while the signals are more realistic for precipitation. We speculate that this is the result of the T2m variability being more strongly linked to the upper tropospheric eddy height (the large-scale waves), while the precipitation variability is more strongly tied to the zonal wind variability (e.g., the changes in weather transients linked to the extension/retraction of the jet along the southern tier of states). The model does a much better job in reproducing the composite zonal wind variability than the composite eddy height variability. As such, even though the T2m has a stronger signal relative to the noise compared to that for precipitation (more on that in Section 4.4), this is a false signal in that it reflects the model's excessive tropospheric wave response to ENSO SST over North America (see also Lim et al. 2020). The impact on actual prediction skill will be addressed in section 4.5.

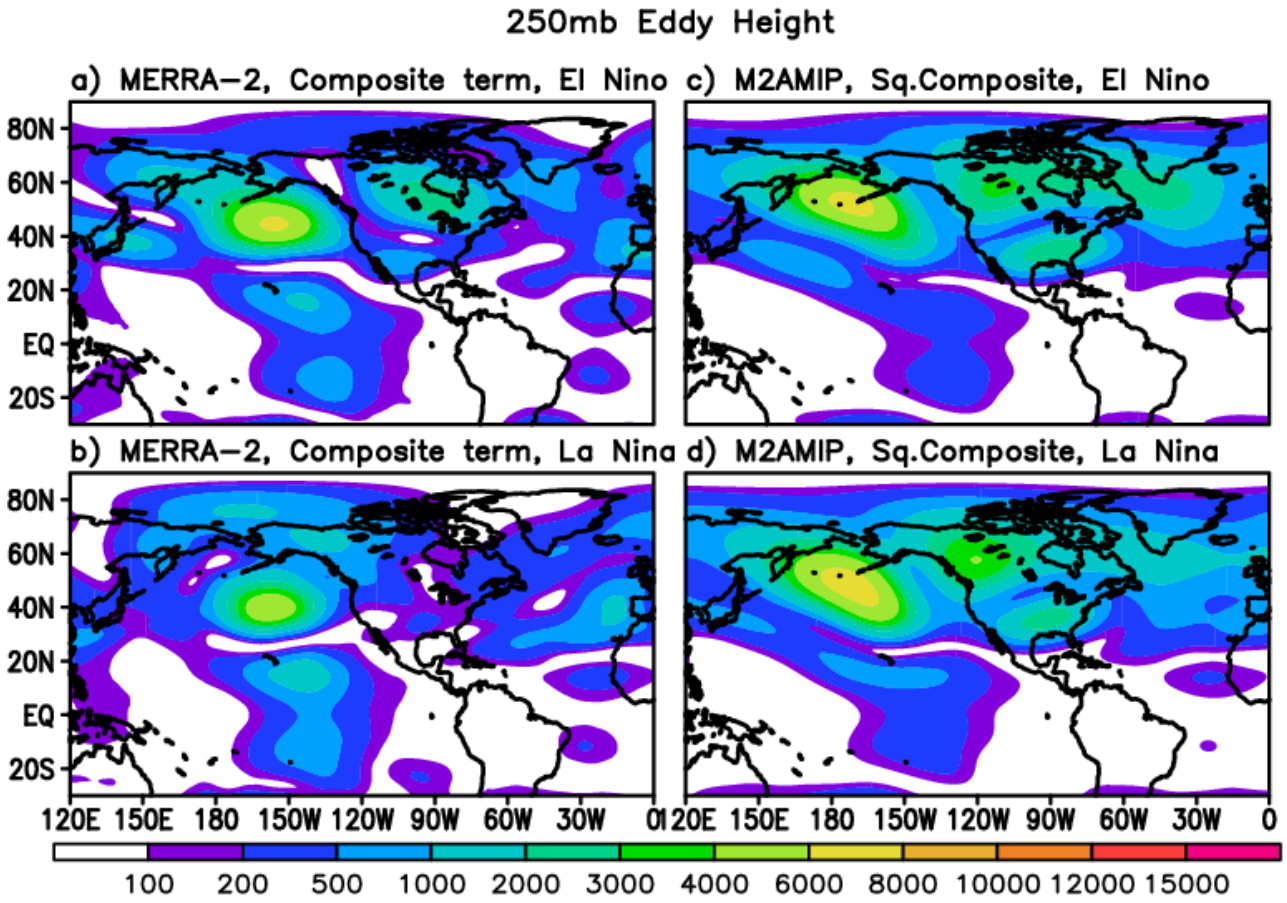

Figure 6: The variance of the 250mb eddy height composite mean based on MERRA-2 (Y'^2, left panels) and M2AMIP ($\{X'^2\}$, right panels). See 2.1.11 in the text. Top panels are for El Niño, and bottom panels are for La Niña events that occurred during 1980-2016. Units: m^2.

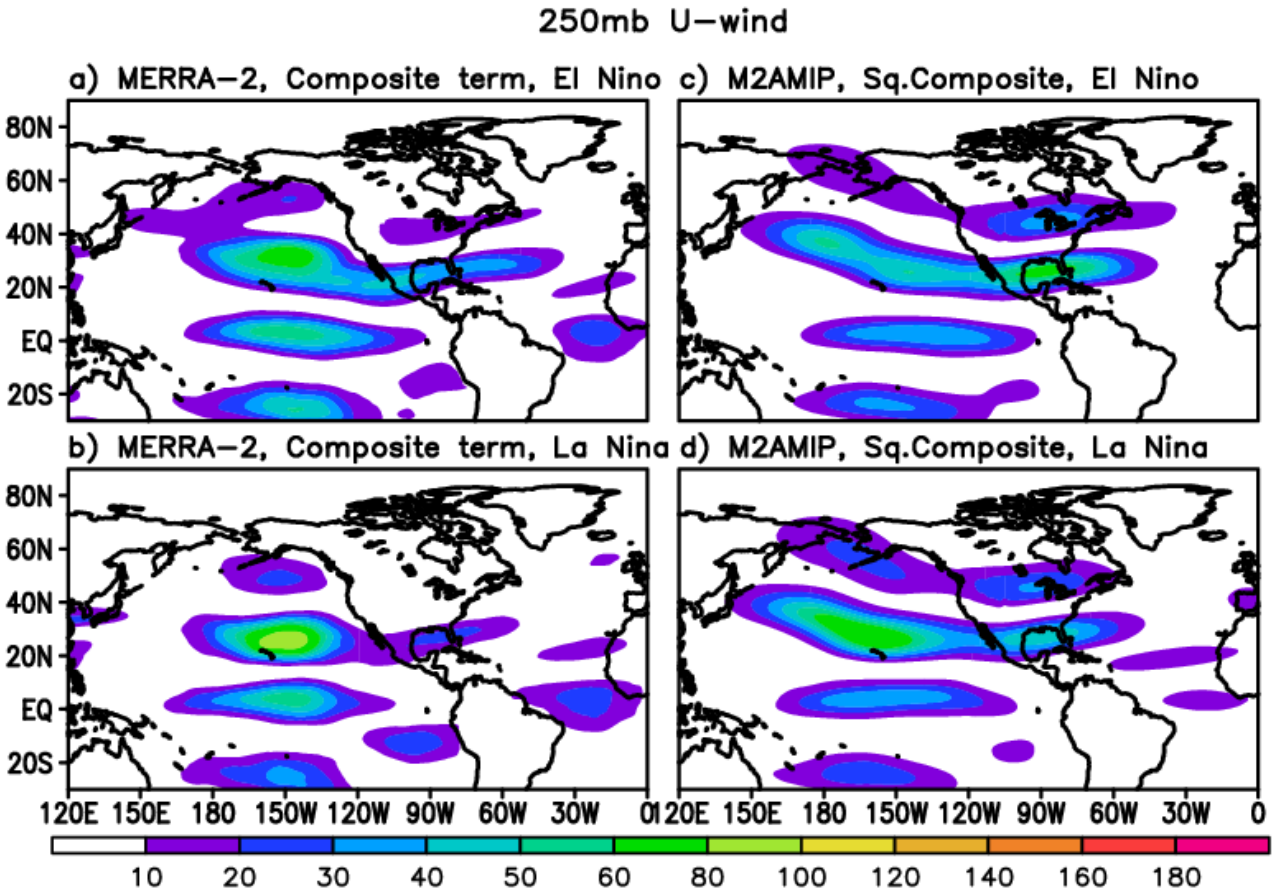

Figure 7: The variance of the 250mb U-wind composite mean based on MERRA-2 (Y'^2, left panels) and M2AMIP ($\{X'^2\}$, right panels). See 2.1.11 in the text. Top panels are for El Niño, and bottom panels are for La Niña events that occurred during 1980-2016. Units: m^2/s^2.

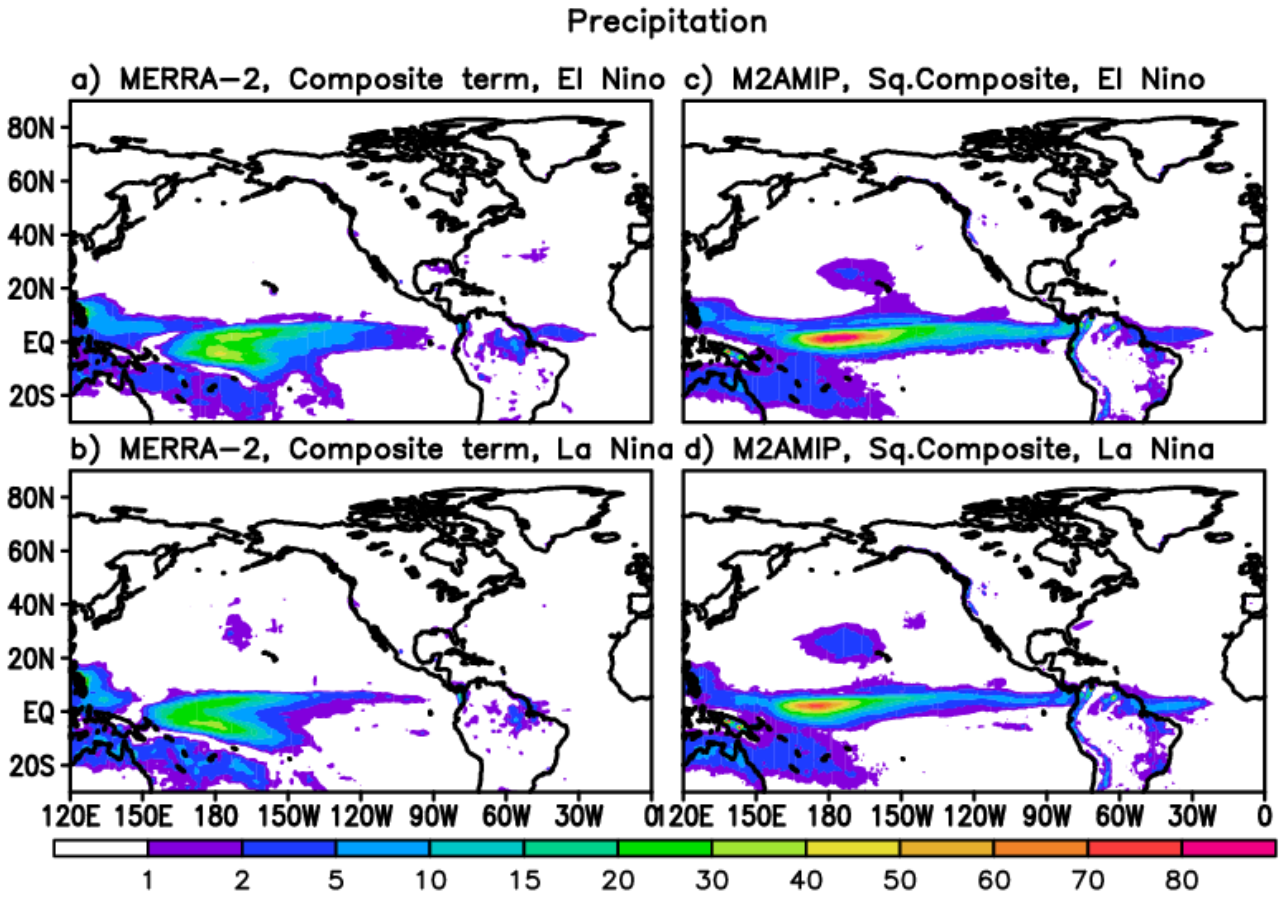

Figure 8: The variance of the precipitation composite mean based on MERRA-2 (Y'^2, left panels) and M2AMIP ($\{X'^2\}$, right panels). See 2.1.11 in the text. Top panels are for El Niño, and bottom panels are for La Niña events that occurred during 1980-2016. Units: mm^2/day^2.

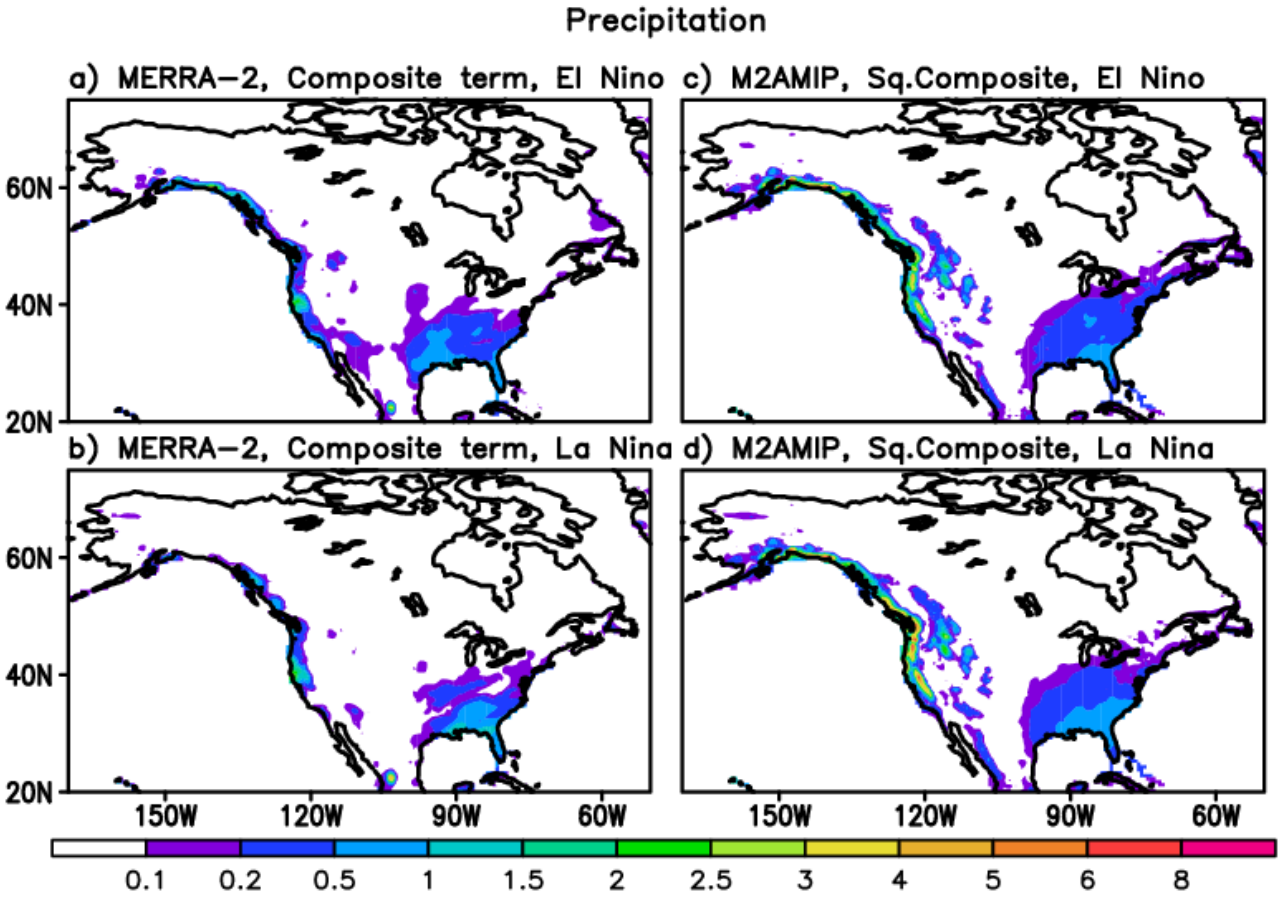

Figure 9: The variance of the North American precipitation composite mean based on MERRA-2 (Y'^2, left panels) and M2AMIP ($\{X'^2\}$, right panels). See 2.1.11 in the text. Top panels are for El Niño, and bottom panels are for La Niña events that occurred during 1980-2016. Units: mm^2/day^2.

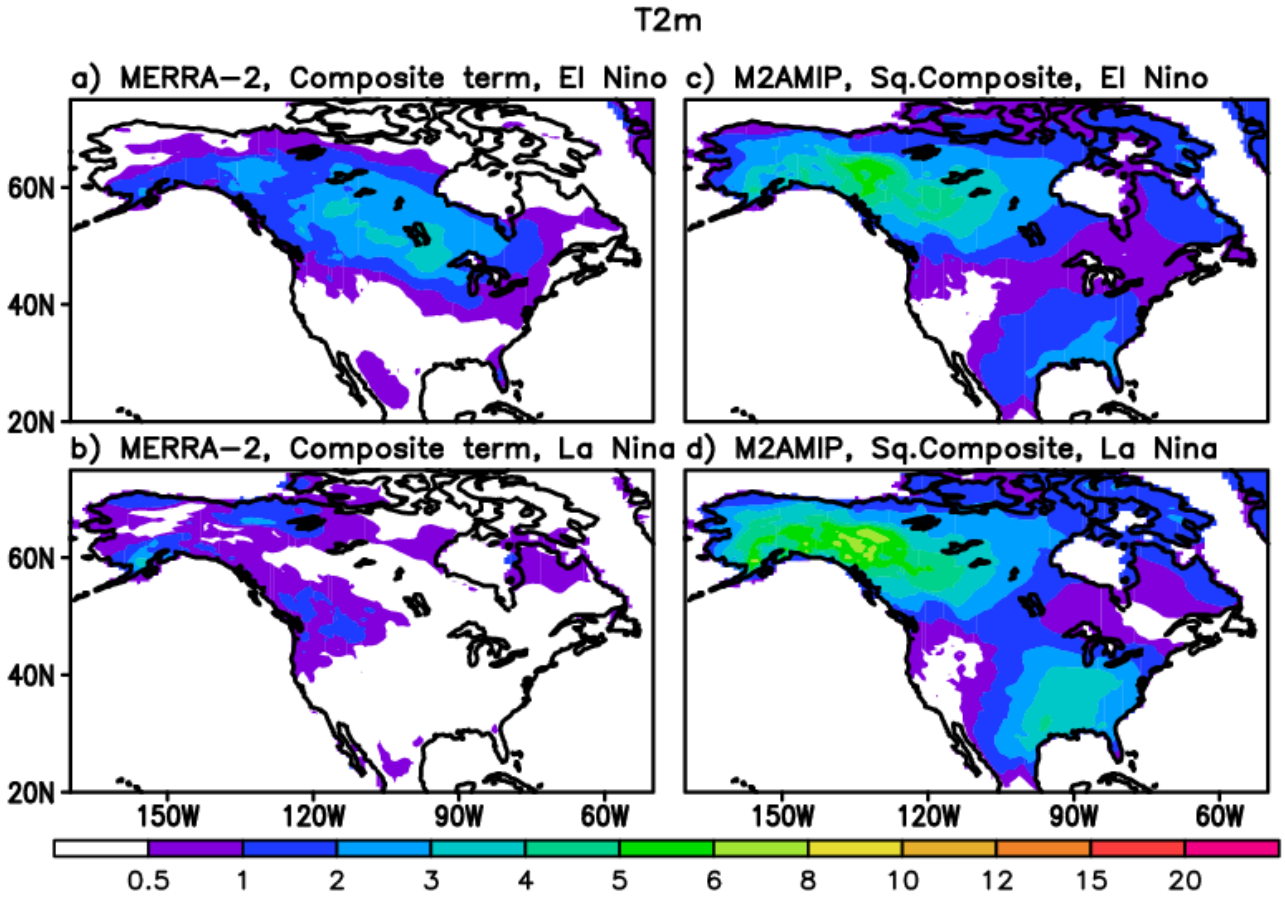

Figure 10: The variance of the North American T2m composite mean based on MERRA-2 (Y'^2, left panels) and M2AMIP ($\{X'^2\}$, right panels). See 2.1.11 in the text. Top panels are for El Niño, and bottom panels are for La Niña events that occurred during 1980-2016. Units: $°K^2$

4.3 The Event-to-Event (E2E) variance

We are especially interested in the E2E component of the variance budget since, as discussed in the Introduction, it is presumably the ability to predict the E2E signal that distinguishes the performance of sophisticated climate models from that of statistical models. We look at that component of the variance in more detail here. In particular, we compare in Figs. 11-15 the total E2E variability from the model with that estimated from MERRA-2 (eq. 2.1.12). Furthermore, the model ensemble allows us to distinguish between the signal (term 3) and noise (term 4) components of the total E2E variability (eq. 2.1.5). The hope is that comparisons with the observations (MERRA-2) for the total E2E variability can provide some insights into whether the model is providing realistic estimates of the individual components (signal and noise) of the E2E variability (note that these are already shown in the middle panels of Figs 1-5). As such, to better quantify the relative contributions of the signal and noise to the E2E variance, we include in Figs. 11-15 the ratio of the E2E signal to the E2E noise, defined as (see eq. 2.1.2):

$$\text{S/N} = \langle \{\hat{X}\}^2 \rangle \; / \{ \langle (X^* - \langle X^* \rangle)^2 \rangle \} \qquad 4.2.1$$

Figure 11 shows the E2E variance for the 250mb eddy height field for El Niño (left panels) and La Niña (right panels). Comparing the M2AMIP total E2E variance with the MERRA-2 values for El Niño (top left and middle left panels of Fig. 11), we see that the model tends to overestimate the total E2E variance, especially in the North Pacific and over North America. This would indicate (given the dominance of the noise term in this region, see Fig. 1) that the model is very likely substantially overestimating the E2E noise during El Niño in the North Pacific. In addition, (as we saw for the total variance, bottom panels of Fig. 1) the model also places the maximum E2E variance in the North Pacific to the west of the observed maximum. In the North Atlantic, the model also appears to overestimate the total E2E variance just south of Greenland, with MERRA-2 showing the largest variance further to the northeast, roughly centered over Iceland. In contrast, the model appears to do a much better job of reproducing the total E2E variance during La Niña events (cf. the top right and middle right panels). This reflects a substantial increase in the variance (compared with El Niño) in the MERRA-2 values, as the model's variance differs little between El Niño and La Niña events. As such, it would appear that the model underestimates the potential predictability associated with E2E variability over much of the Northern Hemisphere for El Niño events (E2E noise is too large), though it likely provides a more realistic estimate of the predictability for La Niña events.

The bottom panels of Fig. 11 show that the E2E S/N (4.2.1) of the 250mb eddy height is largest (values greater than 2) in the tropics between about 150W and 120W, with the largest

values for El Niño (bottom left panel of Fig. 11) straddling the equator, and with the La Niña maximum values (bottom right panel of Fig. 11) primarily located south of the equator. It is noteworthy that for the spatial domain shown here, the tropical S/N values outside the tropical Pacific are generally larger for La Niña than for El Niño. In the NH extratropics, the values are considerably smaller than in the tropics (less than 1). For El Niño, values greater than 0.3 occur over the Gulf of Alaska, the southern U.S. and northern Mexico (where values reach 0.6) and northeast Canada. In contrast, for La Niña there is only a relatively small region over the southcentral US and northern Mexico where values exceed 0.3. The fact that the model overestimates the total E2E 250mb eddy height variability for El Niño (and assuming much of that is from the noise), it is likely we are underestimating the E2E S/N for El Niño over much of the North Pacific and North America.

Figure 12 is the same as Figure 11, but for the 250mb u-wind. The left panels (top and middle) highlight the disparity for El Niño between the total E2E variance from M2AMIP and that from MERRA-2 in the North Pacific. MERRA-2 shows that the region of maximum variance is just off the west coast of North America, while the M2AMIP results have the maximum variance both larger and shifted further to the west, approximately centered on the dateline. This appears to reflect both the E2E signal and E2E noise terms for El Niño being located too far west, though the noise is considerably larger than the signal in that region (cf. the middle panels of Fig. 2a). For El Niño, the model also produces excessive E2E zonal wind variance over North America. In the North Atlantic, the model tends to underestimate the E2E variance for both El Niño and La Niña (cf. the middle and top panels of Fig. 12). Again (as we saw for the 250mb eddy height) , the model seems to produce more realistic E2E variability in the North Pacific during La Niña events than for El Niño events, with MERRA-2 showing substantial differences between El Niño and La Niña, and the model essentially producing the same variability during both warm and cold events.

The E2E S/N for the 250mb zonal wind (bottom panels of Fig. 12) is again (as we saw for the 250mb eddy height) largest in the tropics, with the values for El Niño exceeding 3 over much of the tropical eastern Pacific. In contrast, the values for La Niña over the tropical Pacific are considerably smaller, with peak values (<2) occurring south of the equator near 130W. In the NH extra-tropics, the El Niño values (bottom left panel of Fig. 12) are largest along the southern tier of states and northern Mexico (> 0.6), and over the Great Lakes region (>0.4). The S/N in the NH extra-tropics is considerably smaller for La Niña than for El Niño, with values exceeding 0.3 only over northern Mexico and along coast of the Gulf states. Again, we need to keep in mind that (as was the case for the 250mb eddy height), we are likely underestimating the true 250mb u-wind E2E S/N for El Niño over the North Pacific and North America, given that the model is likely producing excessive E2E noise variance over much of that region.

Figure 13 shows the E2E variability for precipitation. The results from MERRA-2 highlight the remarkable differences between El Niño and La Niña in the tropical Pacific E2E

variability (top panels), with relatively large variability during El Niño events extending across the Pacific, while during La Niña events the largest E2E variability is largely confined to the western Pacific, presumably reflecting the underlying cold SST in the eastern Pacific during La Niña events. While M2AMIP generally shows similar differences between warm and cold events (middle panels of Fig. 13), the model clearly overestimates the tropical Pacific E2E variability associated with El Niño events, while it does a better job of reproducing the variability tied to La Niña events (though this variability is still excessive in the western tropical Pacific). Coming back to the question posed earlier (what is the source of the large El Niño precipitation variability in the tropical Pacific?), it appears that the much of the excessive precipitation variance is tied to excessive E2E variability (more so than the excessive composite variability—Fig. 8), though it is unclear how much of that is signal and how much is noise since both contribute substantially to the E2E variance (see middle panels of Fig. 3a). Turning now to the E2E S/N (bottom panels of Fig. 13), the ratio is largest in the eastern tropical Pacific (values > 2) for El Niño, while for La Niña the largest values occur over a smaller region of the tropical Pacific centered near 130W—though that is a region with overall small E2E variance (see middle right panel of Fig. 13). During El Niño, there is also a band of relatively large S/N values (>0.6) centered between about 5-10°N throughout the tropical Pacific.

Focusing on North America, Fig. 14 shows that M2AMIP generally does a good job of reproducing the overall geographical distribution of E2E precipitation variability, with the maximum variability occurring along the west coast and the southeastern U.S. for both warm and cold events (cf. the top and middle panels of Fig. 14). There is even some evidence that the model reproduces some of the more subtle aspects of the geographical distribution of the E2E variability over the southeast, with warm events showing somewhat greater variability along the Gulf coast (including Florida) compared with cold events, and the largest variances in the southeast shifted further to the west for La Niña (cf. the middle left and middle right panels of Fig. 14). On the other hand, M2AMIP appears to produce excessive variance over the mid-Atlantic states for both the warm and cold events. The E2E S/N (bottom panels of Fig. 14) is overall quite small over North America, with values exceed 0.2 occurring mostly over the SE and primarily for La Niña, with also some values exceeding 0.2 along the US west coast for El Niño. The somewhat larger E2E S/N for La Niña compared with El Niño in the southeast is surprising in light of the larger E2E S/N found for El Niño over North America for the 250mb eddy height and u-wind (cf. the bottom panels of Figs. 11 and 12). On the other hand, there is a suggestion of somewhat larger overall S/N over North America for El Niño (values greater than 0.1), though these are likely not significantly different from the La Niña values. As such, it is likely that the differences in the E2E S/N values for precipitation in the southeast between El Niño and La Niña are driven in part by factors other than the upper level circulation (e.g., low level circulation), though that will not be pursued further here.

Turning next to T2m over North America, the results for MERRA-2 in Fig. 15 (top panels) show that there are substantial differences in the E2E variability between warm and cold

events. The E2E variability during La Niña (top right panel) shows a broad swath of relatively large variability extending from Alaska southeastward through central Canada into the Great Lakes region. That basic pattern is reproduced by M2AMIP (middle right panel). In contrast, during warm events the variability (based on MERRA-2) is more confined to northern and central Canada (top left panel of Fig. 15). This is not well reproduced by M2AMIP, which has a variance distribution during El Niño that is not very different from that of La Niña (cf. middle panels of Fig. 15). The model tends to overestimate the E2E T2m variance for El Niño, including a tendency to produce an excessive variance over western Canada and much of the U.S.. The model does an overall better job in reproducing the E2E variance for La Niña (cf. the top right and top middle panels of Fig. 15), though it produces excessive variance over the western U.S.. The E2E T2m S/N (bottom panels of Fig. 15) is overall larger than that for precipitation (cf. the bottom panels of Fig. 14), with the largest values (>0.3) occurring for El Niño in the SW U.S. and northwestern Mexico and in northeast Canada (bottom left panel of Fig. 15). A similar distribution occurs for La Niña (bottom right panel of Fig. 15) though with somewhat smaller values, and with the values in the U.S. more confined to the southcentral states.

A recurring theme so far has been the inability of the model to reproduce the differences in the variances between cold and warm events. As such, the model appears to be considerably more linear in its atmospheric response to ENSO SST anomalies than is warranted by the observations. In addition, the model's excessive precipitation variability in the tropical Pacific during warm events suggests that the model's precipitation response is too sensitive to the ENSO-related SST anomalies in that region (likely resulting in excessive noise and perhaps excessive signal in the atmospheric response, especially for El Niño). The other rather sobering result is the overall small E2E S/N ratios over North America for T2m and especially for precipitation, though these values are more than likely underestimates of the true values given the model's excessive E2E variability, which we suggest is largely composed of noise.

We note that while the above E2E S/N ratio is instructive in that it gives a clear indication of the fraction of the E2E variability that is potentially predictable (for this model), it is not a practical measure of the overall potential predictability—that requires measuring the strength of the signal with respect to the total variance, as we do next.

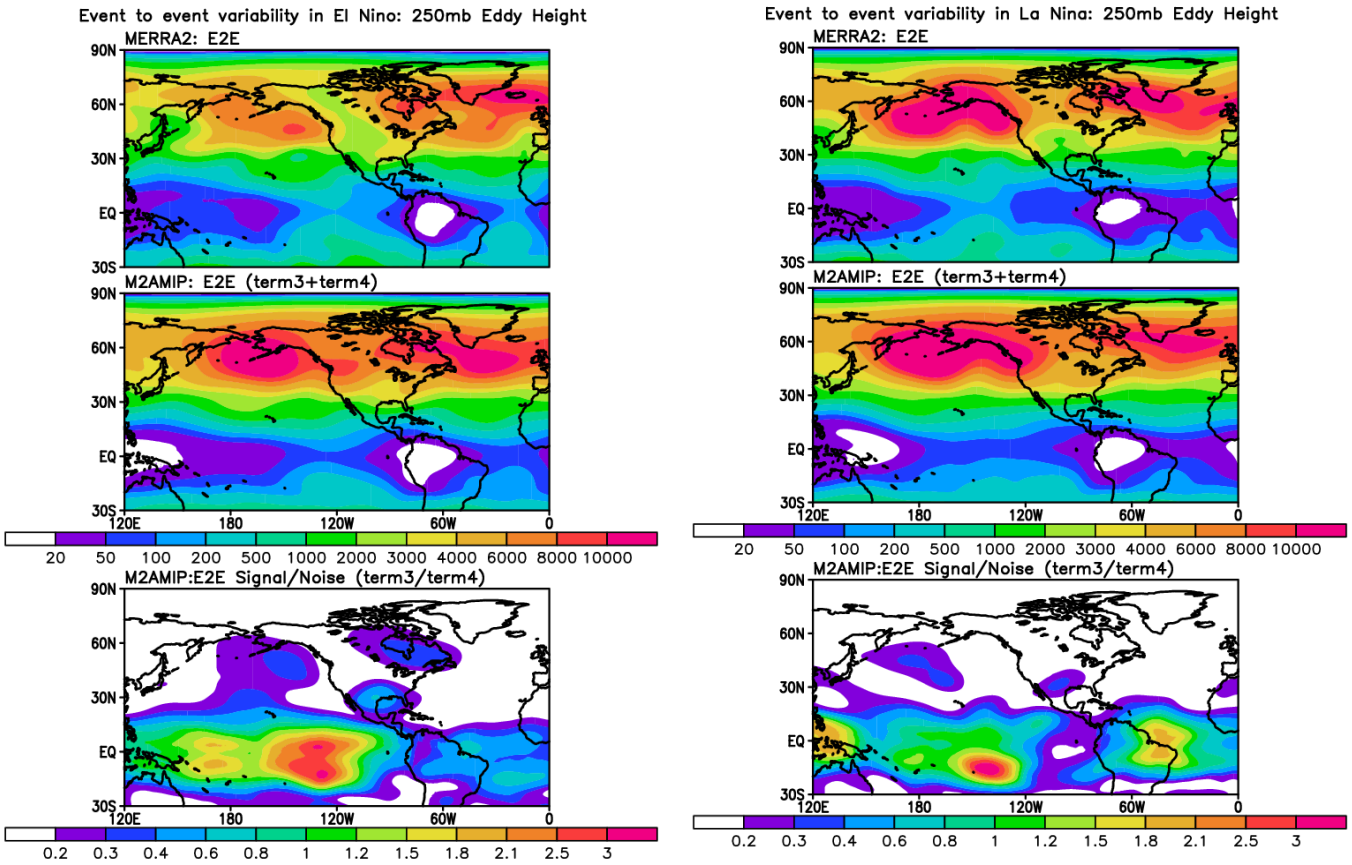

Figure 11: E2E results for 250mb eddy height for El Niño (left panels) and La Niña (right panels) averaged over the months December through March for the events that occurred during 1980-2016. Top panel: E2E variance based on MERRA-2. Middle panel: E2E (Term 3 +Term 4) variance based on M2AMIP. Units: m^2. Bottom panel: E2E signal (Term 3) / E2E noise (Term 4) based on M2AMIP.

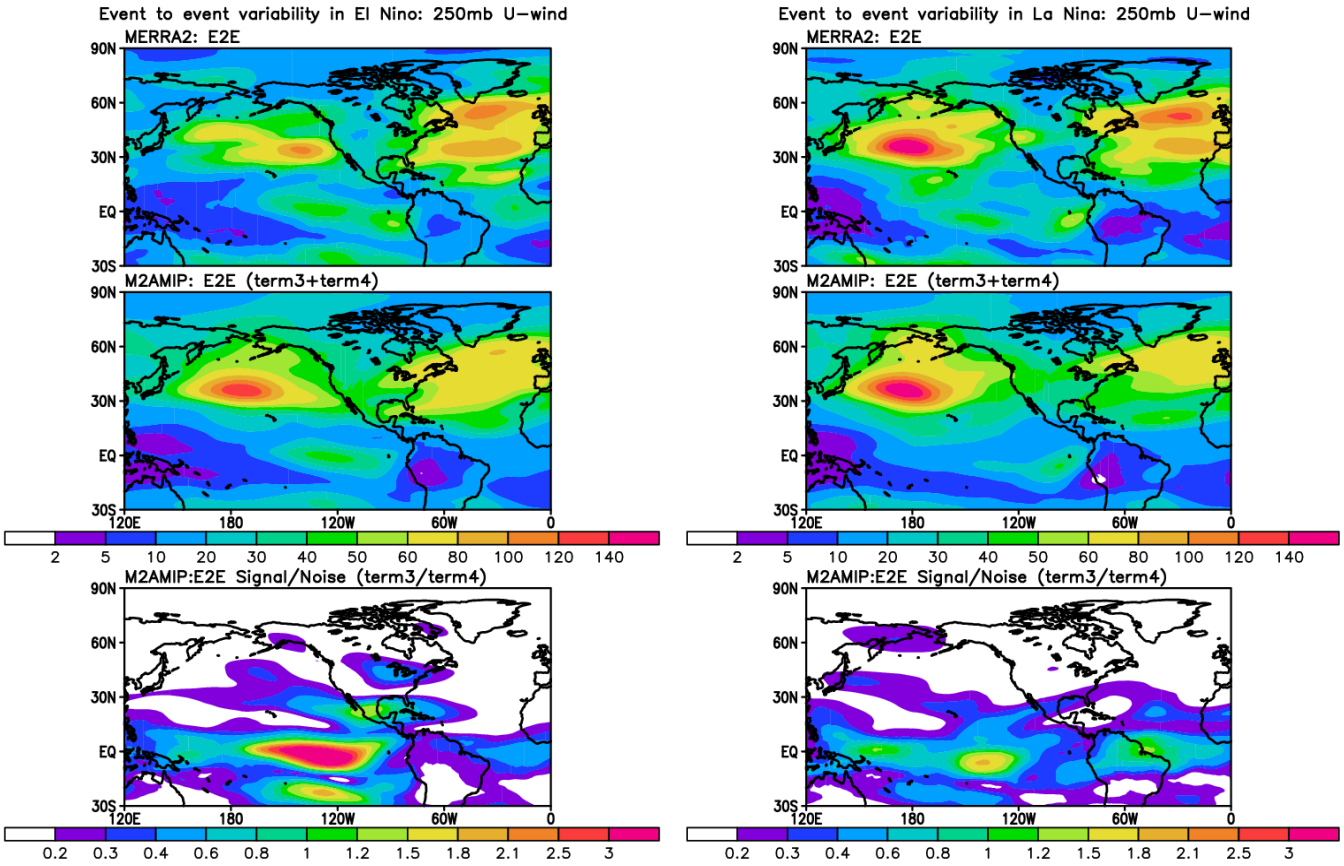

Figure 12: E2E results for 250mb u-wind for El Niño (left panels) and La Niña (right panels) averaged over the months December through March for the events that occurred during 1980-2016. Top panel: E2E variance based on MERRA-2. Middle panel: E2E (Term 3 +Term 4) variance based on M2AMIP. Units: m^2/s^2. Bottom panel: E2E signal (Term 3) / E2E noise (Term 4) based on M2AMIP.

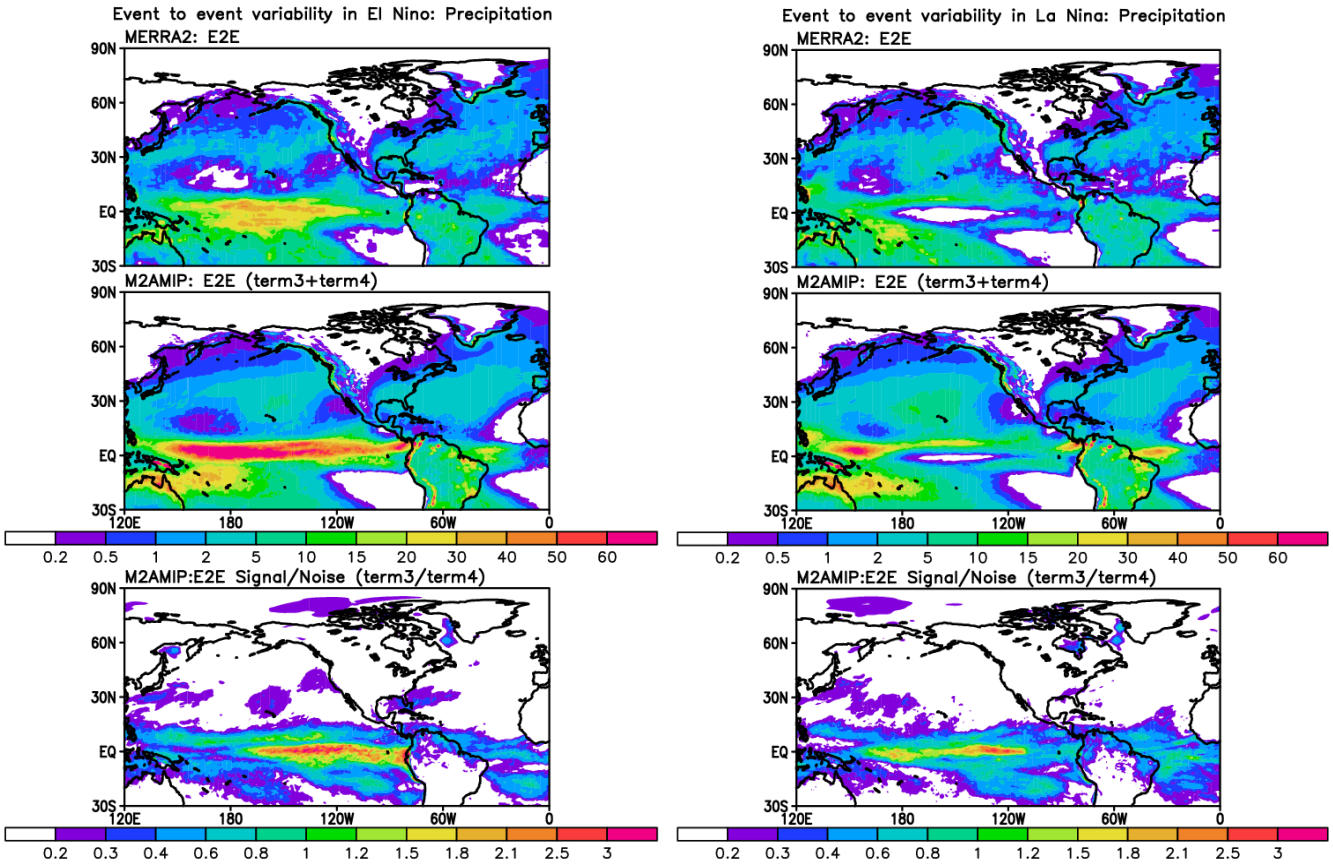

Figure 13: E2E results for precipitation for El Niño (left panels) and La Niña (right panels) averaged over the months December through March for the events that occurred during 1980-2016. Top panel: E2E variance based on MERRA-2. Middle panel: E2E (Term 3 +Term 4) variance based on M2AMIP. Units: mm^2/day^2. Bottom panel: E2E signal (Term 3) / E2E noise (Term 4) based on M2AMIP.

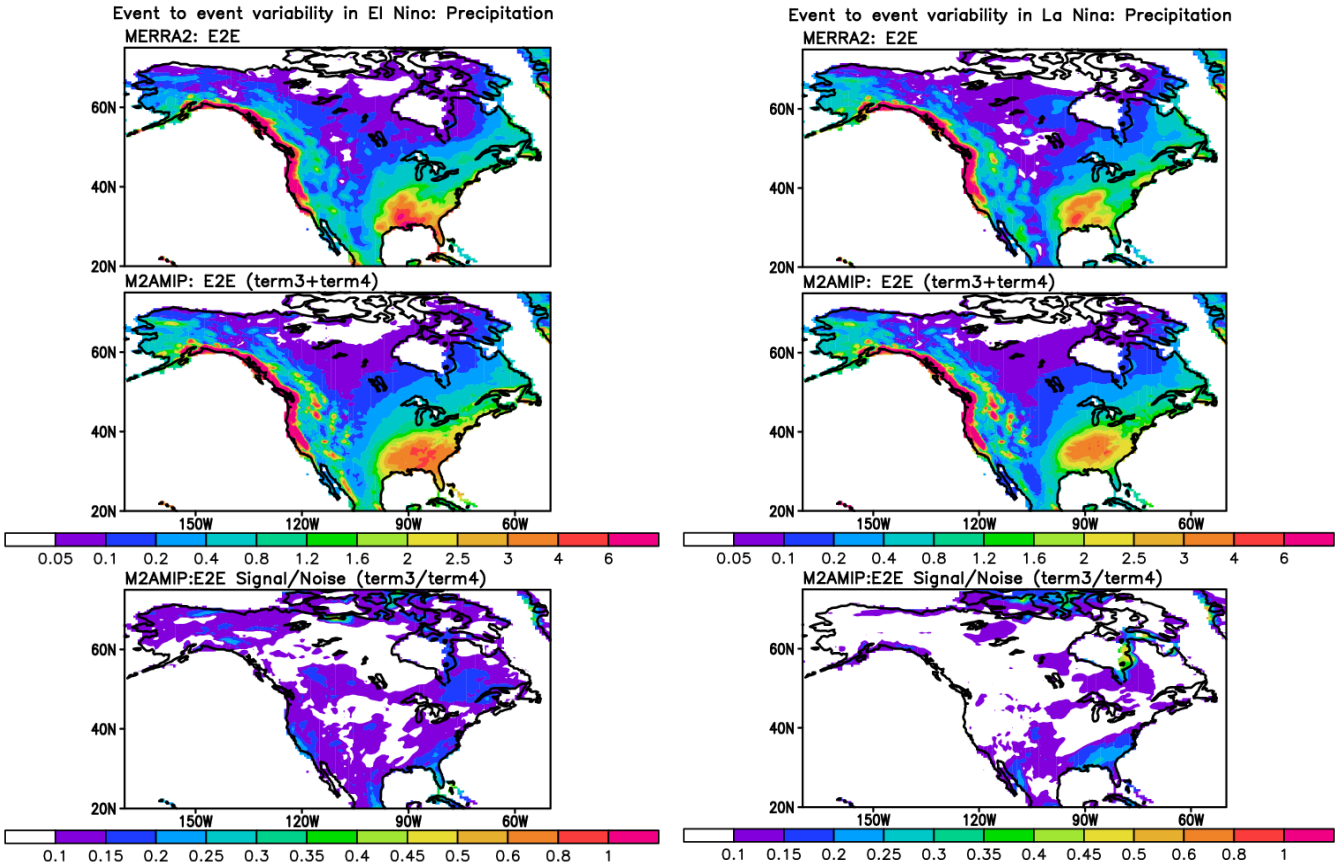

Figure 14: E2E results for precipitation over North America for El Niño (left panels) and La Niña (right panels) averaged over the months December through March for the events that occurred during 1980-2016. Top panel: E2E variance based on MERRA-2. Middle panel: E2E (Term 3 +Term 4) variance based on M2AMIP. Units: mm^2/day^2. Bottom panel: E2E signal (Term 3) / E2E noise (Term 4) based on M2AMIP.

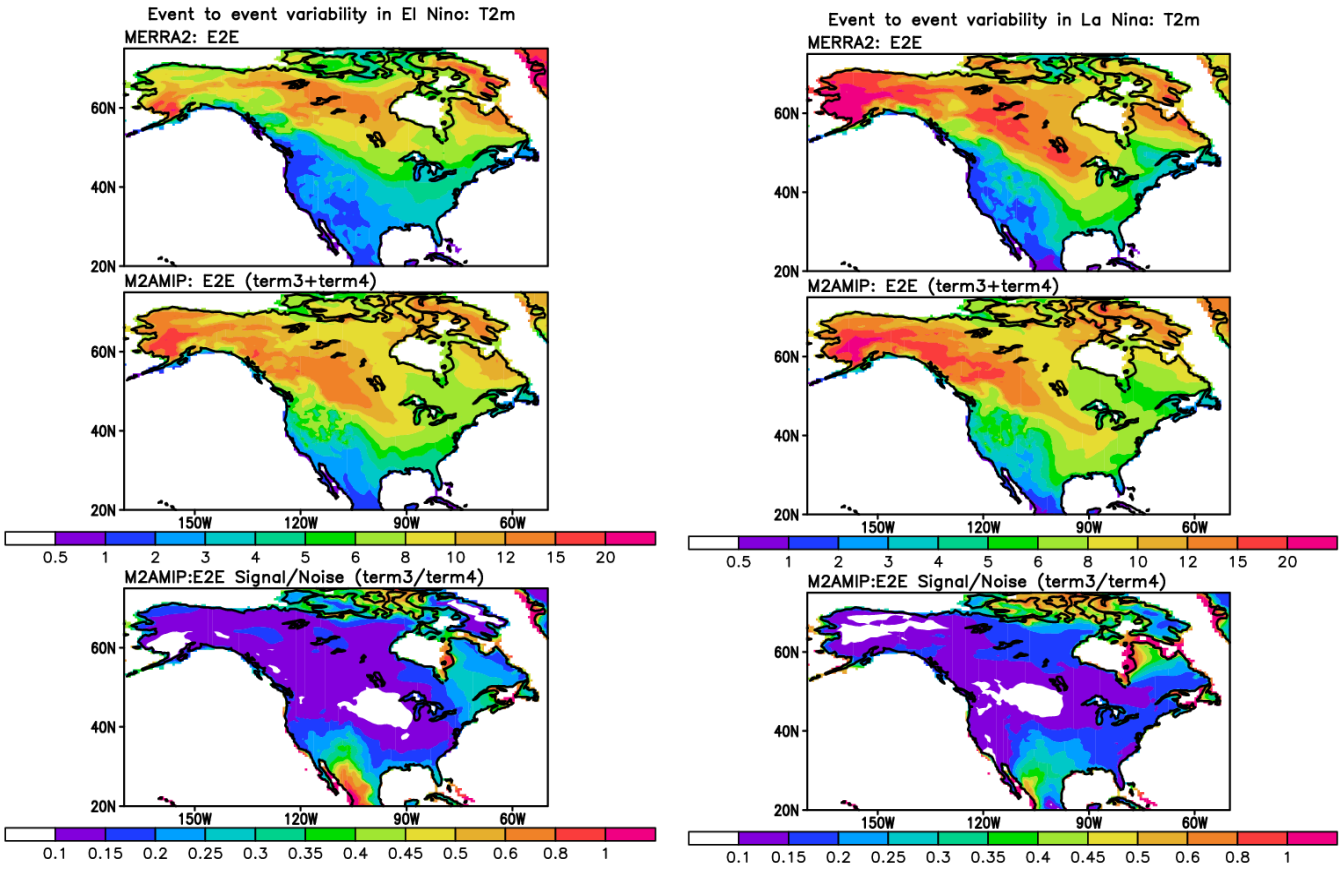

Figure 15: E2E results for T2m over North America for El Niño (left panels) and La Niña (right panels) averaged over the months December through March for the events that occurred during 1980-2016. Top panel: E2E variance based on MERRA-2. Middle panel: E2E (Term 3 +Term 4) variance based on M2AMIP. Units: $°K^2$. Bottom panel: E2E signal (Term 3) / E2E noise (Term 4) based on M2AMIP.

4.4 The signal-to-total (S/T) variance ratio

Here (Figs. 16-19) we examine the two different signal-to-total (S/T) variance ratios defined in section 2. Recall that the S/T associated with individual ENSO events (the E2E variance) is given by 2.1.7, while that of the composite means is given by 2.1.8. To be clear, the denominator is the same in both 2.1.7 and 2.1.8 (the total variance); only the numerator (the signal) differs depending on whether we are considering the signal associated with E2E ($S/T|_{E2E}$) or the composite ($S/T|_{composite}$).

Focusing first on the precipitation (Fig. 16), there are several features of the S/T values worth noting. Particularly evident is the generally broader spatial extent of the $S/T|_{E2E}$ coverage compared to the $S/T|_{composite}$ coverage (cf. left and right panels), with some evidence for a tendency for spatially complementary coverage, especially evident in the tropical Pacific for La Niña events ($S/T|_{composite}$ is largest where $S/T|_{E2E}$ is a minimum). Also, the stronger $S/T|_{composite}$ for La Niña in the central/eastern tropical Pacific (compared with El Niño) is striking (cf. left panels of Fig. 16)– reflecting in part the weak total variability (especially E2E) in that region during La Niña events (see middle panels of Fig. 3b). On the other hand, the tropical/subtropical Pacific $S/T|_{E2E}$ is larger and covers a broader region during El Niño events compared with La Niña events (cf. right panels of Fig. 16).

Turning next to the 250mb eddy height (Fig. 17), we see a somewhat similar tendency for a spatially complementary coverage of the $S/T|_{composite}$ and $S/T|_{E2E}$ values. The somewhat larger $S/T|_{composite}$ values in the central/eastern tropical Pacific for La Niña compared with El Niño (cf. left panels of Fig. 17) is consistent with the stronger $S/T|_{composite}$ values for the precipitation noted above. Also noteworthy are the larger $S/T|_{composite}$ values (compared with those for El Niño events) over the tropical Atlantic. In the NH extratropics, there is remarkably little difference between El Niño and La Niña in the $S/T|_{composite}$ values, though they are somewhat larger over the southeast U.S. for El Niño, and larger for La Niña just north of Hawaii. There is a tendency for larger $S/T|_{E2E}$ values for El Niño (compared with La Niña) over the Gulf of Alaska, the southeast U.S., and northeastern Canada, though the values are relatively small (on the order of 0.2, cf. right panels of Fig. 17) compared to those in the tropics where they exceed 0.6. In fact, the $S/T|_{E2E}$ values are generally quite small (< 0.1) for La Niña over most of North America. This again indicates that the circulation impacts of La Niña events over North America are on average not predictable beyond what can be achieved from predicting the canonical (composite) La Niña response, while there is some hope that we can predict event-to-event differences in the responses over North America for El Niño (cf. the right panels of Fig. 17).

The results for 250mb u-wind (Fig. 18) are overall reflective of the results for the 250mb eddy height, with the $S/T|_{composite}$ maximum values occurring in zonal bands in the central/eastern tropical Pacific, in the NH subtropics extending from Hawaii to the southeastern U.S., over the northern U.S./southern Canada, and in the SH (near 25°S) extending from Australia eastward across the South Pacific. The $S/T|_{E2E}$ values over North

America are again (as we saw for the 250mb eddy height) very small (0.1 or less) for La Niña events, while they are somewhat larger for El Niño events (greater than 0.2 over much of the eastern U.S.).

We next focus on the S/T over North America for precipitation (Fig. 19, top 4 panels) and T2m (Fig. 19, bottom 4 panels). Perhaps the most striking aspect of the $S/T|_{composite}$ for precipitation is that the values are very small (less than 0.1) over most of North America for both El Niño and La Niña (top two left panels). The main exception is the La Niña composite in the S.E. U.S., especially Florida, where the relatively small noise (Fig. 4b) boosts the $S/T|_{composite}$ values. The distribution of the precipitation $S/T|_{E2E}$ for El Niño (top right panel of Fig. 19) is, in contrast, characterized by somewhat broader extent (along the west coast, northeastern Mexico, the U.S. southeast, north east Canada), though the values are overall quite small (0.1 – 0.2). The $S/T|_{E2E}$ for La Niña is for the most part even smaller (less than 0.1), with somewhat larger values (greater than 0.15) largely confined to the southeastern U.S..

In contrast with the overall weak S/T for precipitation, the $S/T|_{composite}$ for T2m (bottom two left panels of Fig. 19) shows values exceeding 0.3 over much of the southern U.S./northern Mexico and parts of northwest Canada for both El Niño and La Niña, though the values are somewhat larger and cover a greater area for the La Niña composite. In the case of $S/T|_{E2E}$ (bottom two right panels of Fig. 19), El Niño shows the largest values (greater than 0.2) over the southwest U.S. and western Mexico, over the northeastern U.S., and over northeastern and northern Canada. In the case of La Niña (bottom right panel of Fig. 19), the T2m $S/T|_{E2E}$ is for the most part less than 0.15 except over northern Mexico and northern Canada.

In summary, the S/T values over North America are overall quite small for precipitation, the main exception being over the extreme southeast U.S. (especially for La Niña), where the largest contribution to the signal comes from the composite variance. In contrast, the S/T values are substantially larger and more widespread for T2m, with contributions from both $S/T|_{composite}$ and $S/T|_{E2E}$. Values tend to be larger and spatially more extensive for El Niño than for La Niña, with spatial distributions (and relative magnitudes) overall consistent with the $S/T|_{composite}$ and $S/T|_{E2E}$ values of the 250mb eddy height and zonal wind shown earlier. As such, we would expect the skill of ENSO-related T2m forecasts to be larger than the skill of precipitation forecasts. Here, however, we need to keep in mind that (as discussed earlier) the model's tropospheric wave response to the ENSO SST over North America is too strong, producing an unrealistic strong (false) signal in T2m. On the other hand, the precipitation signal over North America (though smaller) is more realistic, as it is more strongly tied to the response of the upper tropospheric zonal wind to the ENSO SST – something the model does relatively well. The actual impact on forecast skill (correlations with the observations) is addressed in the next section.

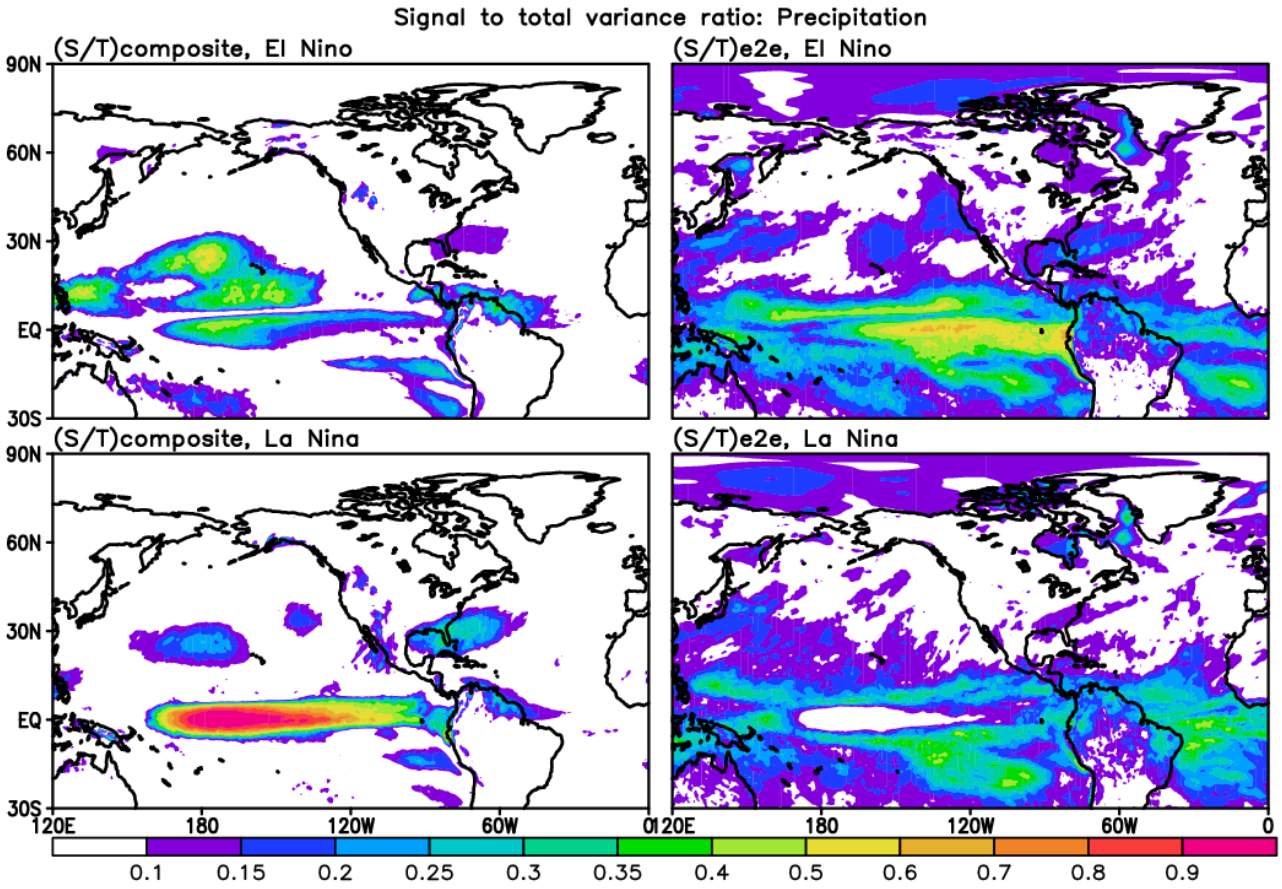

Figure 16: The signal-to-total variance ratios for precipitation for the composite mean (S/T $|_{composite}$, left panels), and event-to-event variability (S/T $|_{E2E}$, right panels). Top row is for El Niño and the bottom row is for La Niña for the events that occurred during 1980-2016. See Section 2 for details.

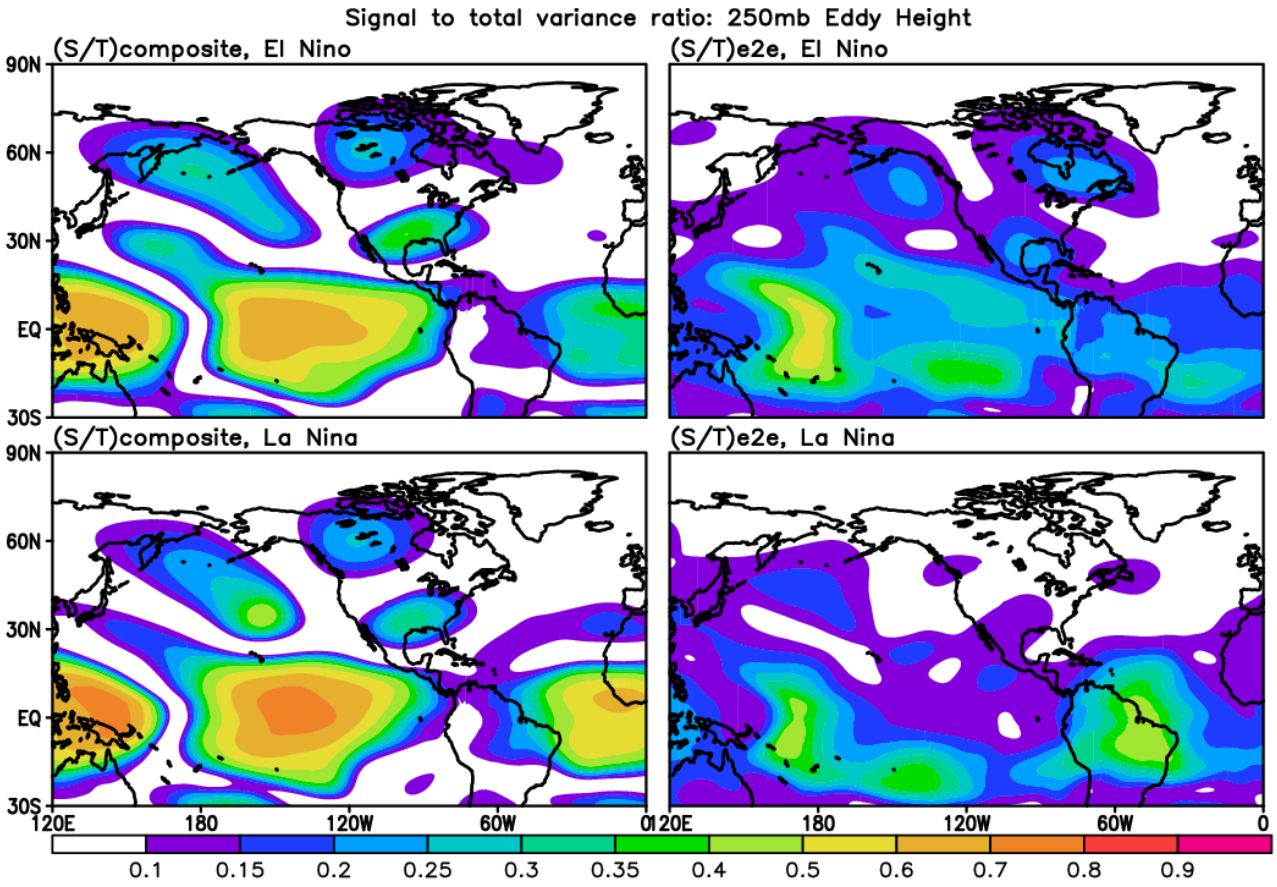

Figure 17: Same as Fig. 16, except for 250mb eddy height.

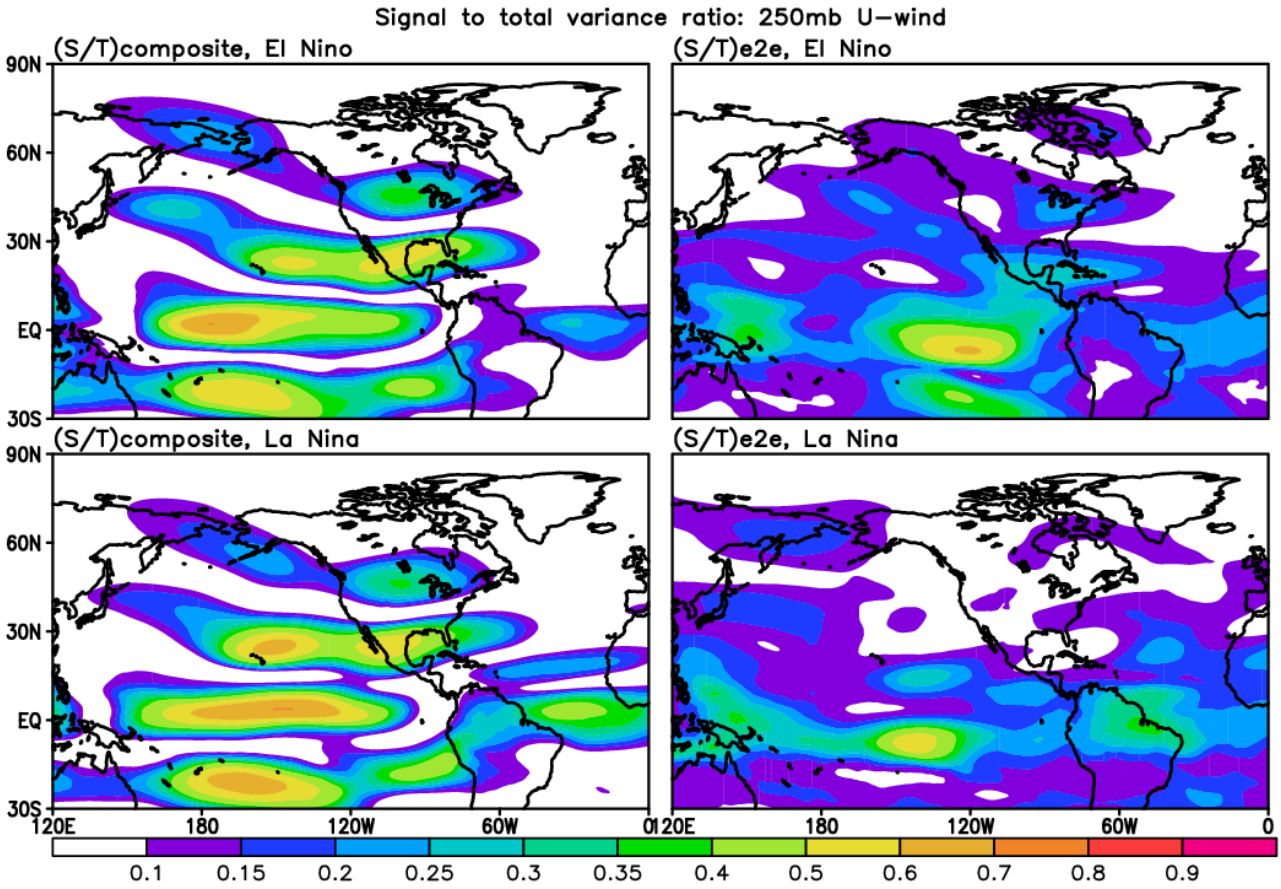

Figure 18: Same as Fig. 16, except for 250mb u-wind.

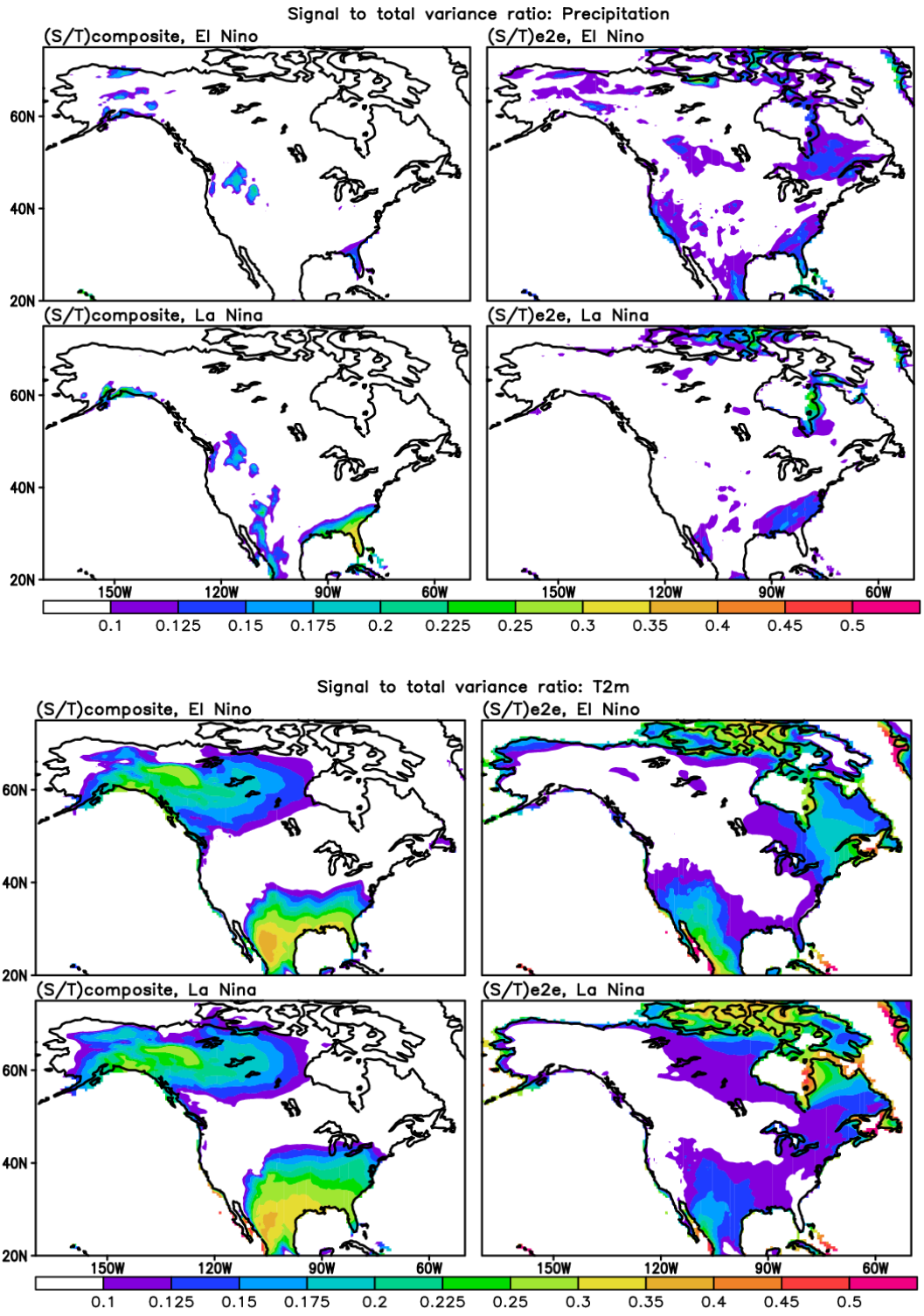

Figure 19: The signal-to-total variance ratios for precipitation (top four panels) and T2m (bottom four panels) over North America for the composite mean ($S/T|_{composite}$, left panels), and event-to-event variability ($S/T|_{E2E}$, right panels). Top row in each set of four is for El Niño and the bottom row is for La Niña for the events that occurred during 1980-2016. See Section 2 for details.

4.5 Correlations with observations

We next examine a more direct correspondence (the covariability) with the observations (MERRA-2) as defined in section 2.2. We focus in particular on the decomposition of the conditional correlation into the terms associated with the composite mean and the E2E variability (see eq. 2.2.2). Following Lim et al. (2020), we focus on the months of January and February for which the GEOS coupled model exhibits the greatest differences in skill over North America, with January characterized by having the poorest, and February the best skill in reproducing the observed El Niño precipitation response over North America. In fact, as mentioned in the Introduction, Chen et al. (2017) showed that the tendency for better ENSO-related skill over North America in February compared to the other winter months is a feature common to many coupled models and holds for both El Niño and La Niña, and for both precipitation and surface air temperature. In light of those results, we show in Figs. 20-24 the correlations separately for January and February and, in order to allow a more direct comparison with the results of the previous sections, we also present the correlations averaged over the four months (December through March). In the following we shall use the words skill and correlation interchangeably.

Before showing the results, we think it is useful to briefly discuss how to interpret the two components of the correlations (eq. 2.2.2). To do that, it is helpful to consider two different idealized cases. The first case is one in which the ENSO response is constant so that there is no E2E variability ($\hat{Y}$=$\hat{X} = 0$). Then the first term on the RHS of eq. 2.2.2 is zero, and the second term would take on a value of either 1 or -1, depending on whether the observed and simulated composite anomalies are of the same or of the opposite sign. Note that in this case the variances are $\sigma^2_{\{X\}}$= $\{X'\}^2$, and $\sigma^2_Y = Y'^2$. The other case to consider is one in which the composite mean ENSO anomaly is zero (Y'= $X' = 0$), so that $\langle X \rangle = [X]$, and $\langle Y \rangle = [Y]$. Then the second term on the RHS of eq. 2.2.2 is zero, and the first term on the RHS is simply the usual definition of a correlation between the two quantities $(\{\hat{X}\}, \hat{Y})$, where the anomalies in the numerator and denominator are all computed with respect to the long-term mean, though of course the averaging is still done conditional on ENSO. So, this hopefully makes it clear that the second term on the RHS of eq. 2.2.2 is simply providing information about similarity of the observed and simulated phases of the ENSO composites, while the first term on the RHS of eq. 2.2.2 is more of a traditional correlation term between two variates (involving the variations about the composite mean), though it differs from the usual definition in that the normalization terms in the denominator (the variances, $\sigma^2_{\{X\}}$ and σ^2_Y) are computed with respect to means that are different from those used in computing the numerator, $\langle \{\hat{X}\} \hat{Y} \rangle$: the long term means versus the ensemble means.

One other key issue to consider is the statistical significance of the correlations. There are basically two sources of uncertainty. The first involves the uncertainty in the model estimates arising from the fact that we do not have an infinite number of ensemble members

(we only have 15). The second is due to the fact that we do not have an infinitely long climate record (we only have the years 1980-2016). We deal with the first source of uncertainty by taking a Monte Carlo approach, where we take a large number of subsets of 10 ensemble members (in this case 400), and repeat the calculations for each to determine whether we can discount the null hypothesis (at 1% significance) that the correlations are in fact zero. We note that while there is considerable overlap in the 10-member subsets (and we may therefore be underestimating that uncertainty), we are limited in reducing that number much further by fact that the ensemble mean would likely begin to diverge considerably from the 15-member mean for subsets much smaller than 10. In any event, the second source of uncertainty is likely greater, and the more difficult problem to address. Here we have to acknowledge that having only on the order of 10 events (9 for La Nina and 11 for El Nino) necessitates that the correlations we produce here be viewed with caution, in particular for the E2E-related correlations (the first term on the RHS of eq. 2.2.2), and especially those shown for the individual months of January and February. Note that with 10 events and employing a t-test with a significance level of 5%, the critical values are +/- 0.63, though these values are for the total correlation in eq. 2.2.2 rather than for the individual terms. On the other hand, the statistical significance of the second term on the RHS of eq. 2.2.2 is really a question about the statistical significance of the composite means and could be addressed with a separate t-test of those means. We, however, do not pursue such tests here, as our focus in this report is on highlighting the basic approach rather than on making definitive statements about the various statistics presented here. In fact, it is likely that significance issues can only be resolved with longer climate/model prediction records. We do, however, also show the average correlations (averaged over the 4 months), which should decrease the uncertainty, though it is not clear how independent the results from the individual months are. In fact, there is evidence (as mentioned above and as we show next) that the correlations do evolve during the course of the winter.

With those caveats in mind, we show in Figure 20 the correlations for the 250mb eddy height field. Focusing on North America, for El Niño (Fig. 20a), we see that indeed the skill tends to be higher during February (middle panels of Fig. 20a) than during January (top panels of Fig. 20a). This is primarily due to the differences in the E2E correlations (cf. the upper left and middle left panels of Fig. 20a). The correlation associated with the composite mean in fact shows little difference between January and February (cf. the upper right and middle right panels of Fig. 20a). For La Niña (Fig. 20b), we see that the skill also tends to be higher over North America during February (compared with January). In this case, the higher February skill is primarily coming from the composite mean (over much of the eastern U.S. and the northwest), with also some contribution (over the Great Lakes region) from E2E variability (middle panels of Fig. 20b). This is in contrast with January, where there is essentially no skill (with northeast Canada being the only exception) tied to E2E variability and only a modest amount of skill (mostly less than 0.3) coming from the composite mean centered over northern Mexico and northwest Canada. Considering the Dec – Mar average correlations we see overall weak correlations over North America (bottom panels of Figs 20a and 20b). For El Niño (bottom panels of Fig. 20a) we find that

the skill over North America comes from both E2E and the composite variance. For La Niña (bottom panels of Fig. 20b), the Dec – Mar average correlations are similar to those for El Niño, though the correlations (especially for E2E) are generally weaker over much North America. For both El Niño and La Niña we see a tendency for the E2E and composite skill to have spatially complementary coverage (the maxima in one tending to be located in the minima of the other) over both North America and the tropics (bottom panels of Fig. 20a and 20b). Overall, the spatial distributions of the correlations are (not surprisingly) quite similar to those of the S/T ratios (cf. bottom panels of Figs. 20a and 20b with Fig. 17). That is, the correlations are largest in the regions with the largest S/T, and this is true for both El Niño and La Niña, and for both E2E and the composite correlations. A key difference is the weaker composite correlations over the southeast U.S. than one would be expect based on the relatively strong composite S/T values in that region (cf. the bottom right panels of Figs. 20a and b with the left panels of Fig. 17). This presumably reflects the fact that the model produces an unrealistically strong eddy height composite signal in the southeast, as discussed earlier (Fig. 6).

It is interesting that in the tropical eastern Pacific the differences in the composite skill are such that the January skill tends to be larger than the February skill (especially for La Niña, cf. top right and top middle panels of Fig. 20b). Also, regarding the strongest composite correlations in the tropical Pacific, the maximum for El Niño is centered near 160°W (bottom right panel of Fig. 20a), whereas that for La Niña is centered further east near 140°W (bottom right panel of Fig. 20b). This is surprising given that the location of the maximum composite precipitation variance in the tropical Pacific for La Niña is to the west of that for El Niño (cf. the left panels of Fig. 8).

The correlations for the 250mb zonal wind are shown for El Niño in Fig. 21a and for La Niña in Fig. 21b. Perhaps the most striking aspect of these results is the overall considerably stronger correlations over North America for the composite mean compared with those from the E2E variance (cf. the left and right panels of Figs. 21a and 21b). This is especially true for La Niña, which, overall, has little correlation tied to E2E variance over North America (left panels of Fig. 21b) - a result that is consistent with the weak E2E S/T ratios there (bottom right panel of Fig. 18). The largest correlations associated with the composites occur (for both El Niño and La Niña) in zonally oriented bands with (in the Northern Hemisphere) one stretching from Hawaii eastward across Mexico and across the southern tier of the U.S., and the other stretching from the Aleutians across the northern U.S. and southern Canada. This is again consistent with the regions of strongest composite S/T ratios (cf. the left panels of Fig. 18). There is again, for both El Niño and La Niña, a tendency for the E2E and composite skill to have spatially complementary coverage, over both North America and the tropics (bottom panels of Fig. 21a and 21b). There is also again a tendency for February to have higher skill than January over North America. For La Niña this is primarily tied to the differences in the correlations tied to the composite mean (cf. the top right and middle right panels of Fig. 21b), whereas for El Niño, it appears to be due to both

the E2E and the composite having somewhat larger correlations in February (cf. the top and middle panels of Fig. 21a).

Figure 22 shows the correlations for precipitation. For El Niño (Fig. 22a), the average correlations (bottom panels) show that, not surprisingly, the highest correlations tend to occur in the tropics, although in the case of the composite some of the largest correlations in the Pacific occur well north of the equator in a broad swath from Indonesia to Hawaii, with another swath to the south of the equator in the region of the SPCZ. There is also again a tendency for the patterns of higher correlations of the E2E and composite to be complementary throughout much of the tropics and subtropics. In the extratropics, the correlations are generally small, though over the North Pacific and North America there is a general agreement between the regions of relatively higher E2E correlations (greater than 0.2) and the regions of relatively higher E2E S/T (cf. bottom left panel of Fig. 22a and the top right panel of Fig. 16). Comparing January and February (top and middle panels of Fig. 22a), we see generally very similar results, though over the North Pacific and over North America the correlations are higher for February, consistent with the findings of Lim et al. (2020) and Chen et al. (2017). In fact, it appears that the February values are the main contributors to the higher correlations seen in the Dec -Mar averages in the North Pacific and North America region (cf. middle left and bottom left panels of Fig. 22a). For La Niña, the average precipitation correlations (bottom panels of Fig. 22b) are by far the largest in the central and eastern tropical Pacific for the composite (greater than 0.9) – though east of about 160W, the large precipitation correlations occur in a region of overall small variability (bottom panels of Fig. 3b). The E2E correlations (bottom right panel of Fig. 22b) tend to be rather noisy, though again there is a tendency for the regions of relatively high correlations for E2E and the composite to be complementary.

Focusing on the precipitation over North America (Fig. 23), we see overall small and spatially scattered correlations for both El Niño and La Niña. Nevertheless, there are some larger regions of higher correlations. In general, the Dec-Mar mean values (bottom panels of Fig. 23a and 23b) tend to be largest along the southern tier of the U.S., in northern Mexico, and along the west coast. Also evident are the relatively higher February correlations for E2E throughout much of North America for El Niño (left middle panel of Fig. 23a), with some of the highest correlations (greater than 0.6) occurring over the southern tier of the U.S. and along the east and west coasts. In contrast, La Niña tends to have the highest correlations associated with the composite, with the largest values confined to northern Mexico and the southern tier of the U.S., especially the southeast (right panels of Fig. 23b). We note that the relatively strong composite-related correlations in the southeast U.S. for La Niña and the strongest E2E correlations occurring on the west coast and southeast for El Niño are generally consistent with the distributions of the S/T (cf. bottom left panel of 23a and top right panel of Fig. 19; and cf. the bottom right panel of Fig.23b and the second row left panel of Fig. 19).

Finally, we turn to an examination of the correlations for T2m over North America (Fig. 24). Looking first at the results for El Niño (Fig. 24a), we see that the average correlations (bottom panels) are largest (greater than 0.3) for the composite over northern Mexico, along the southern tier of the U.S., and over northwestern Canada. For the E2E variance, the largest correlations (though weaker than for the composite) occur primarily over the western U.S., western Mexico, and northeast Canada. These regions of relatively high E2E correlations are remarkably consistent with the regions of relatively high E2E signal to total (S/T) variance (cf. right panels of Fig 19). Comparing the mean composite correlations (bottom right panels of Figs. 24a and 24b) with the corresponding composite S/T ratios (left panels of Fig. 19) we do see a mismatch, with the correlations being much more confined to northern Mexico and the southern tier of the U.S., whereas the largest S/T values encompass much of the eastern U.S. Similar to what we found for the eddy height (see discussion of Figs. 20a and b above), this mismatch presumably reflects a model bias in the T2m composite (a false signal) over North America (see Fig. 10).

Overall, the El Niño correlations (both composite and E2E) for T2m tend to be higher for February than for January (cf. top and middle panels of Fig. 24a). For La Niña, we find somewhat larger average correlations for E2E (bottom left panel of Fig. 24b) compared with El Niño, though these correlations are overall quite small (~0.1) and are likely not statistically significant. The largest values (> 0.4) occur in northeastern Canada. The somewhat larger correlations for La Niña compared with El Niño are surprising given the overall weak S/T values for the upper level circulation during La Niña events (e.g., bottom right panels of Fig. 17 and 18). For La Niña, the largest average correlations for the composite are confined to the eastern tier of U.S. states, northern Mexico and the west coast (bottom right panel of Fig. 24b), with less overall spatial coverage than found for El Niño (cf. bottom right panel of Fig. 24a). These again appear to reflect the larger February correlations (cf. the top and middle right panels of Fig. 24b). The La Niña E2E correlations for January and February (top and middle left panels of Fig. 24b) are unusual in that, while they show a monthly distinction in the spatial coverage of the largest values, there is no clear difference in the overall strength of the correlations. In fact, the areas with the largest mean correlations for E2E are not very consistent with the areas of the largest S/T values (cf. bottom left panels of Fig. 24b with the bottom right panel of Fig. 19), though both metrics tend to be small over much of North America.

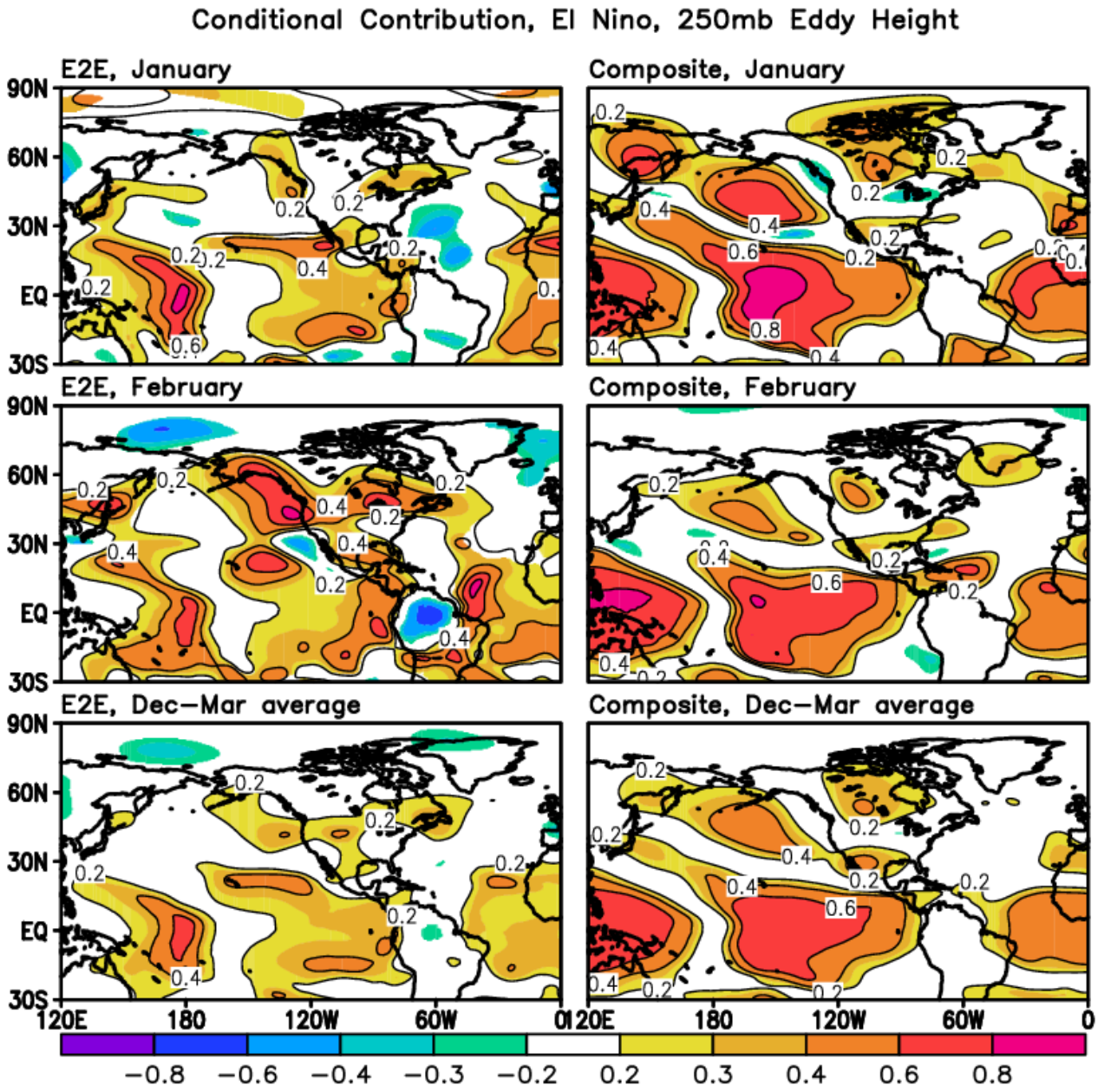

Figure 20a: The El Niño January (top panels), February (middle panels) and average of January through March (bottom panels) conditional correlations between MERRA-2 and M2AMIP for the 250mb eddy height field for the events that occurred during 1980-2016. The correlations are decomposed into the terms associated with event-to-event (E2E) variability (left panels) and the composite mean (right panels). Values not significant at the 1% level based on a Monte Carlo approach to assess ensemble uncertainty and/or correlations with absolute values less than 0.2 are masked out. Positive values are contoured.

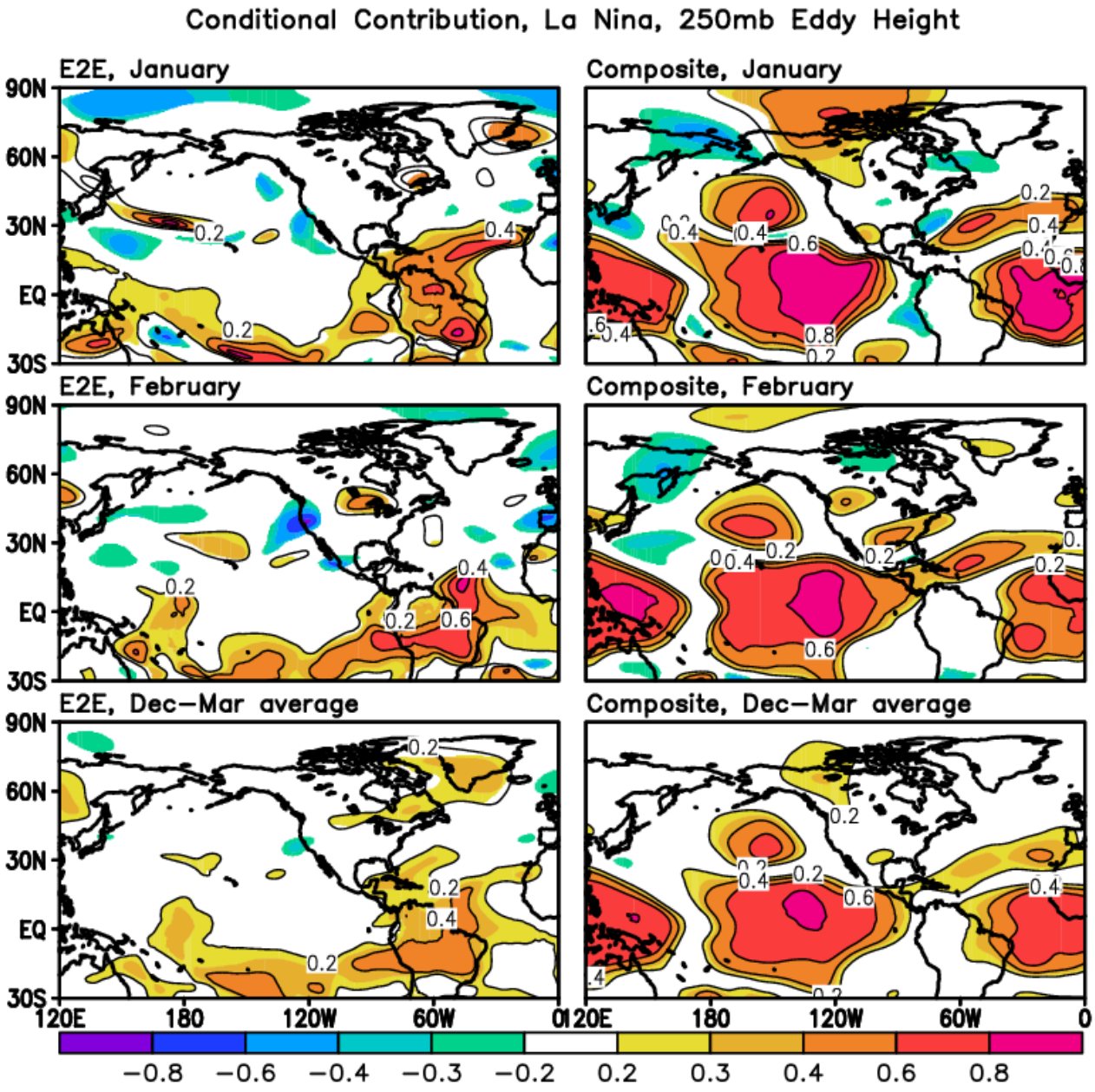

Figure 20b: Same as Fig. 20a except for La Niña.

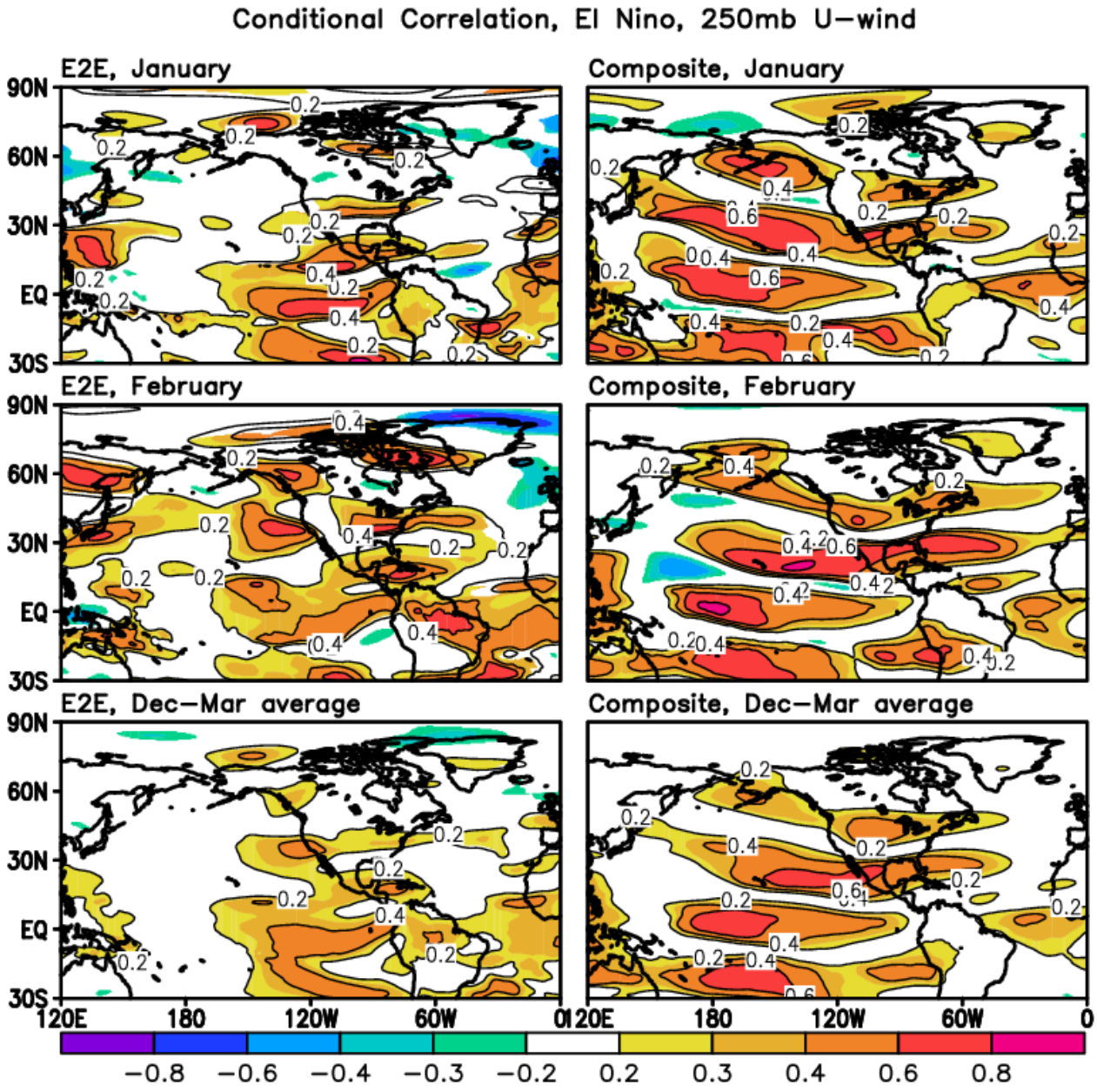

Figure 21a: The El Niño January (top panels), February (middle panels) and average of January through March (bottom panels) conditional correlations between MERRA-2 and M2AMIP for the 250mb u-wind field for the events that occurred during 1980-2016. The correlations are decomposed into the terms associated with event-to-event (E2E) variability (left panels) and the composite mean (right panels). Values not significant at the 1% level based on a Monte Carlo approach to assess ensemble uncertainty and/or correlations with absolute values less than 0.2 are masked out. Positive values are contoured.

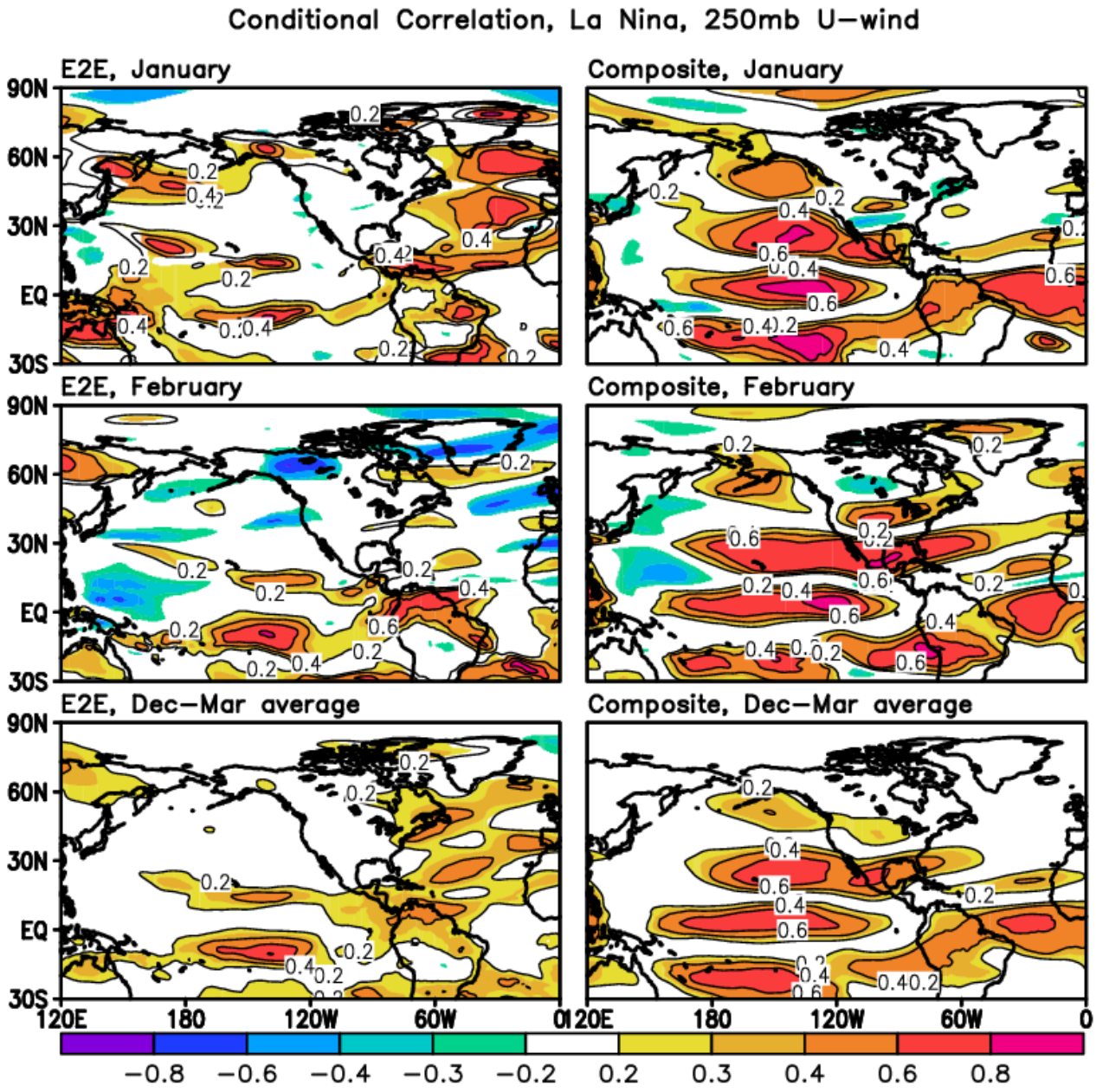

Figure 21b: Same as Fig. 21a except for La Niña.

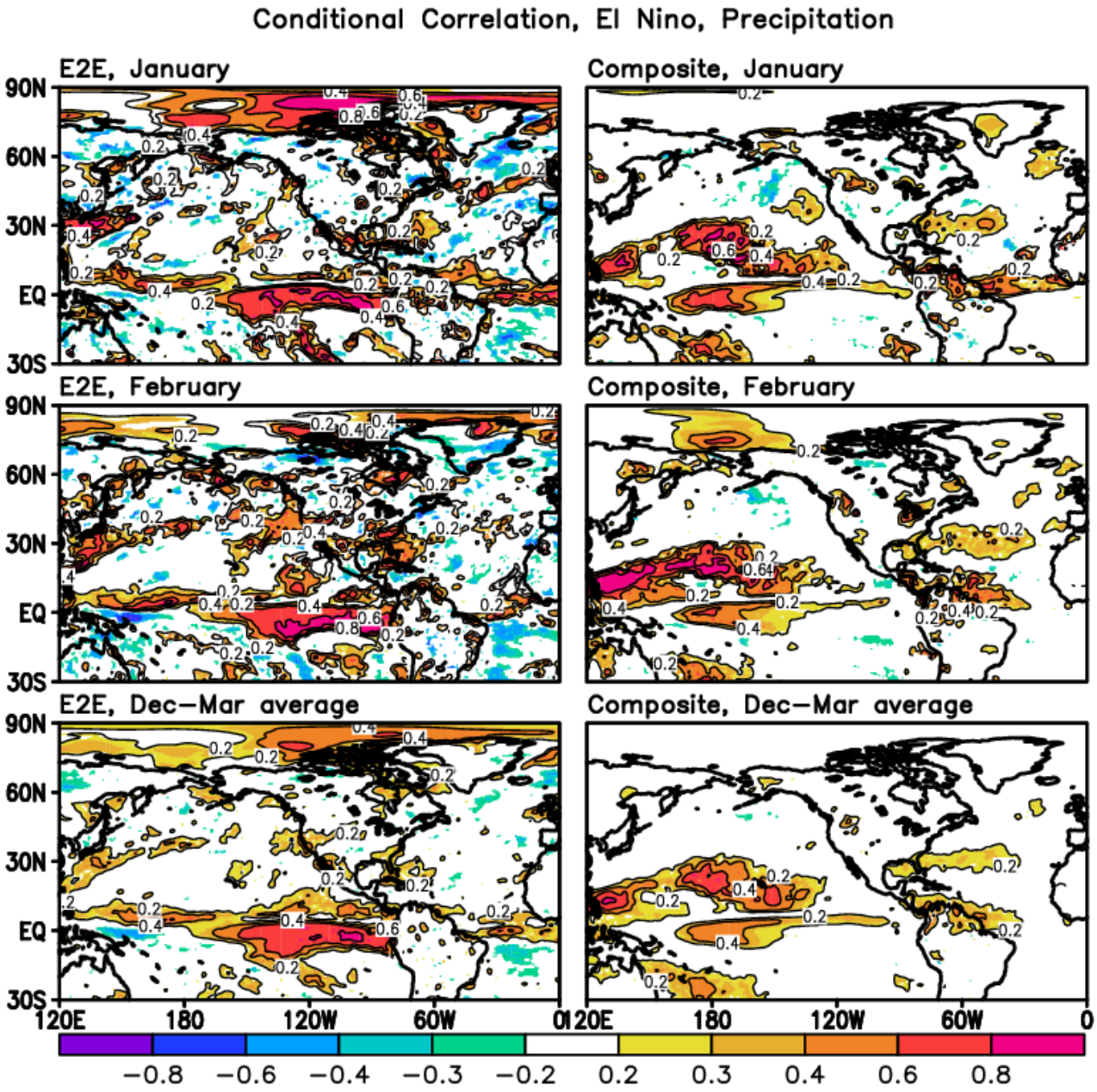

Figure 22a: The El Niño January (top panels), February (middle panels) and average of January through March (bottom panels) conditional correlations between MERRA-2 and M2AMIP for precipitation for the events that occurred during 1980-2016. The correlations are decomposed into the terms associated with event-to-event (E2E) variability (left panels) and the composite mean (right panels). Values not significant at the 1% level based on a Monte Carlo approach to assess ensemble uncertainty and/or correlations with absolute values less than 0.2 are masked out. Positive values are contoured.

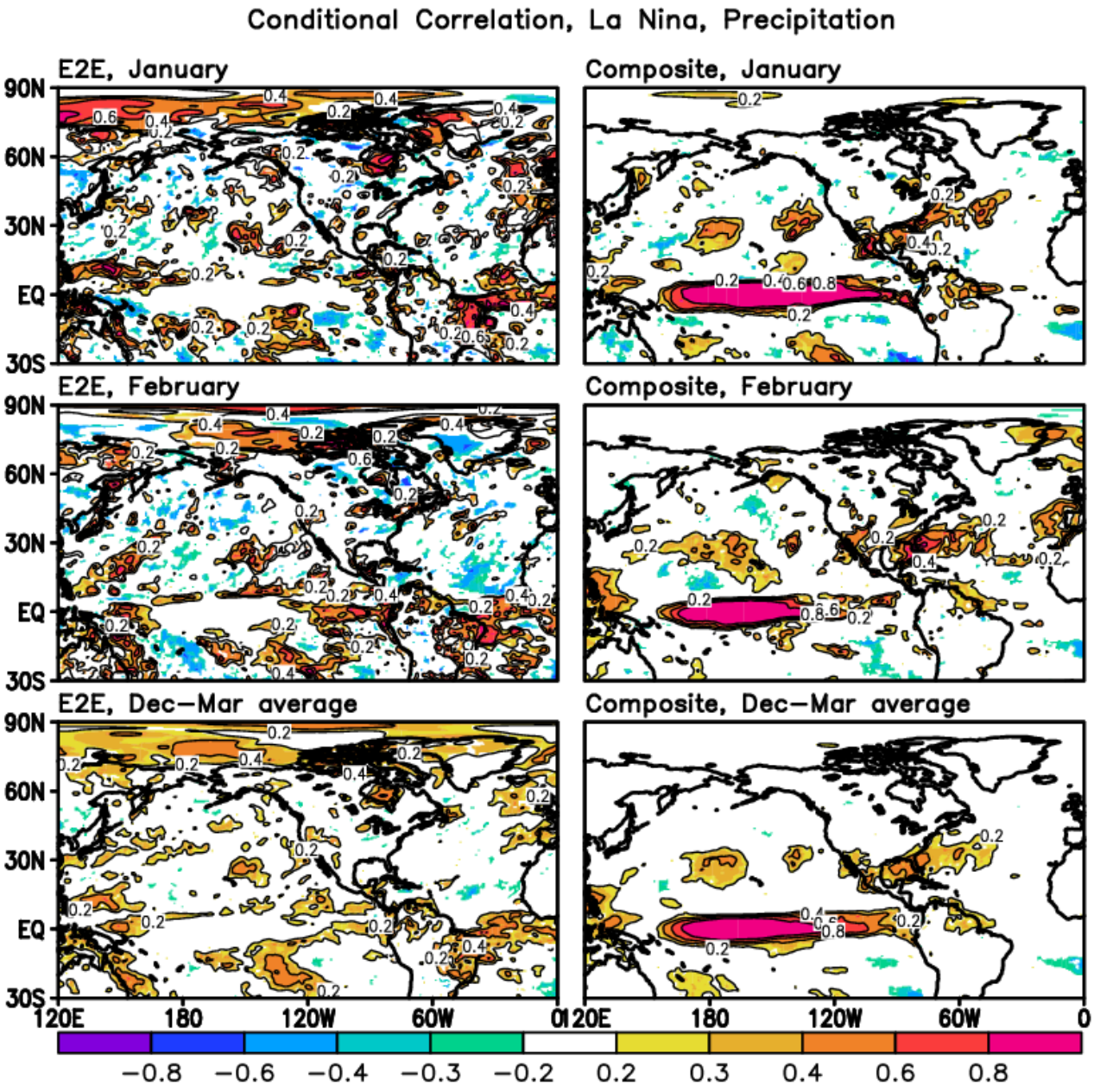

Figure 22b: Same as Fig. 22a except for La Niña.

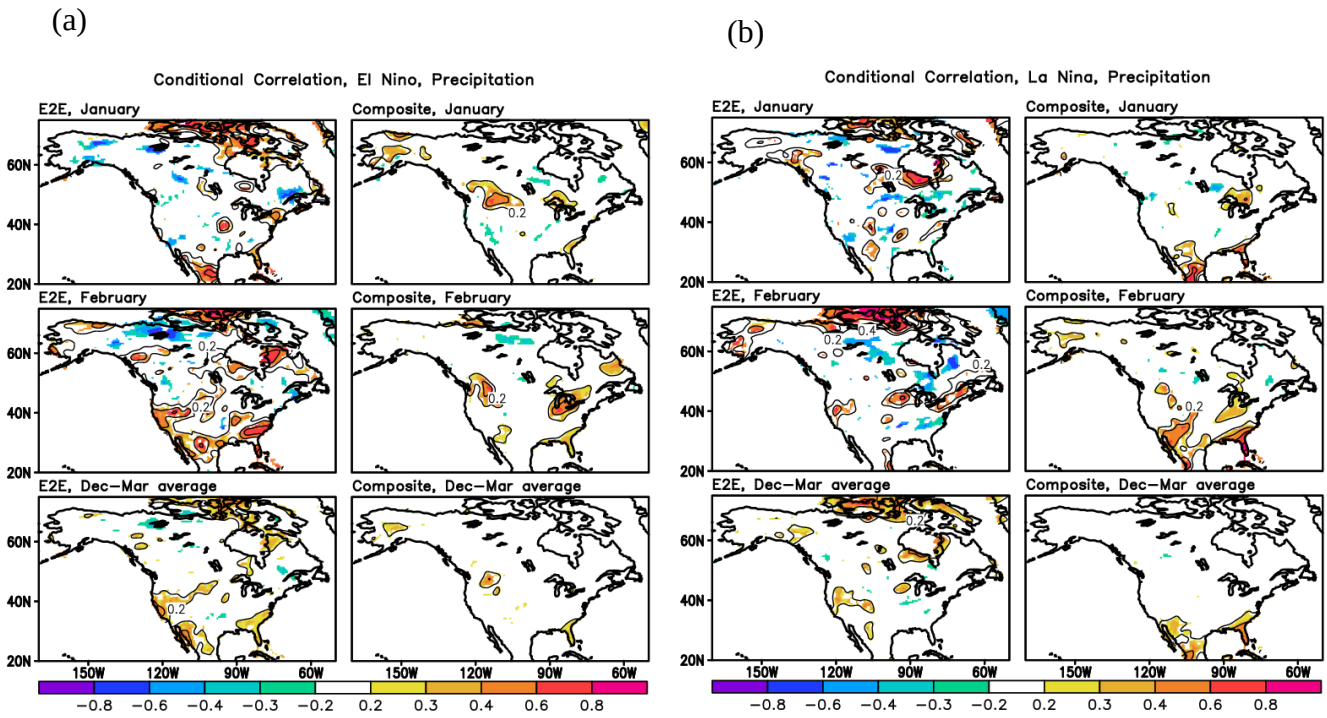

Figure 23: The conditional correlations between MERRA-2 and M2AMIP for precipitation over North America for El Niño (a) and La Niña (b) events that occurred during 1980-2016. In each set of six panels, the results are shown for January (top panels), February (middle panels) and average of January through March (bottom panels). The correlations are decomposed into the terms associated with event-to-event (E2E) variability (left panels) and the composite mean (right panels). Values not significant at the 1% level based on a Monte Carlo approach to assess ensemble uncertainty and/or values with absolute correlations less than 0.2 are masked out. Positive values are contoured.

(a)

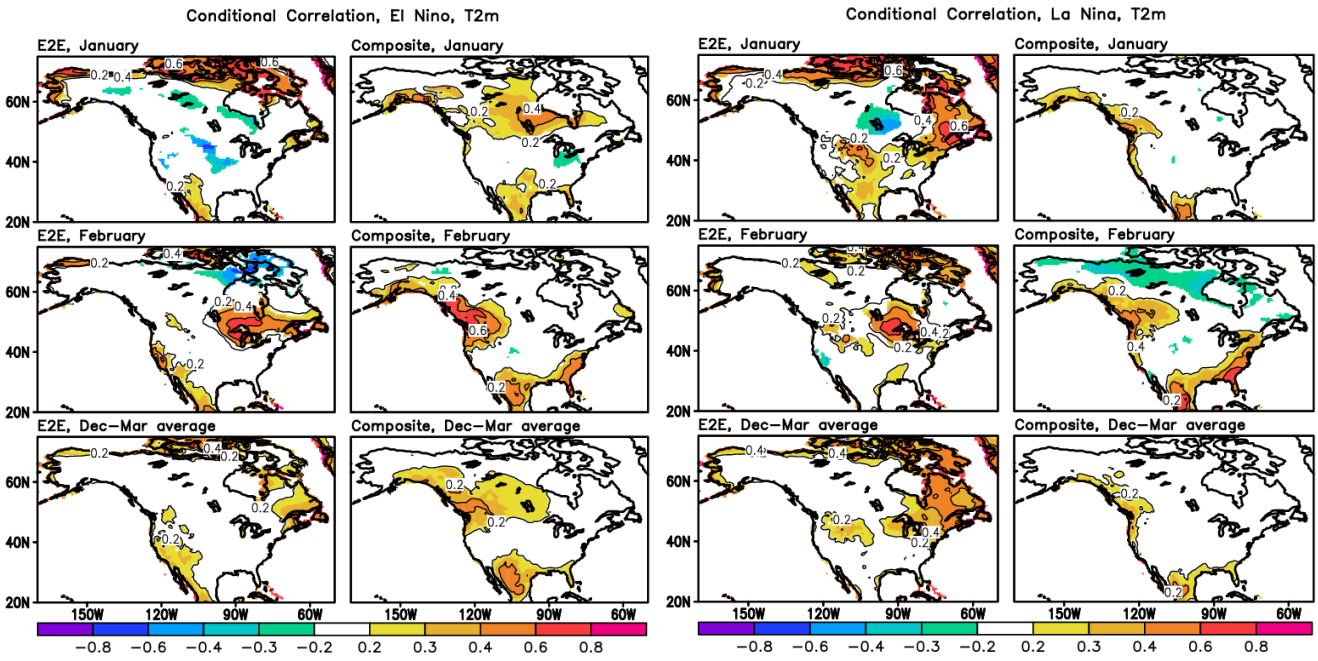

Figure 24: The conditional correlations between MERRA-2 and M2AMIP for T2m over North America for El Niño (a) and La Niña (b) events that occurred during 1980-2016. In each set of six panels, the results are shown for January (top panels), February (middle panels) and average of January through March (bottom panels). The correlations are decomposed into the terms associated with event-to-event (E2E) variability (left panels) and the composite mean (right panels). Values not significant at the 1% level based on a Monte Carlo approach to assess ensemble uncertainty and/or correlations with absolute values less than 0.2 are masked out. Positive values are contoured.

5. Summary and Discussion

We have presented here a decomposition of the variance and covariance of a recurring climate phenomenon that isolates the contributions from the composite mean and the variability about that composite mean (the event-to-event or E2E variability). The decomposition is further tailored to isolate the signal and noise contributions of each (both the composite mean and the E2E variability), taking advantage of the existence of large ensembles of climate model simulations or predictions. As such, the decomposition provides an important metric of model quality, one that focuses on the quality of the model's simulation/prediction of second moments associated with specific recurring climate phenomena. While the decomposition has its roots in "analysis of variance", we believe that there are several important differences from how that is usually applied in the context of statistical regression (where the variance is separated into the explained and unexplained components). First, here we deal with two sources of variability: climate and ensemble. Second, the statistics are conditional on the occurrence of a specific phenomenon. Third, we also decompose the covariances, allowing the separation of the skill of a model prediction into that coming from the composite mean and that coming from the E2E variability.

As an example of the decomposition, we have focused here on ENSO as represented in an ensemble of 15 climate simulations produced with the NASA GEOS AGCM forced with observed SST for the period 1980-2016. Results are presented separately for El Niño and La Niña, focusing on the extended boreal cold season consisting of the months December - March. Results are presented for 250mb eddy height, 250mb zonal wind, and precipitation, covering a large part of the globe where ENSO impacts are known to be important, together with a more detailed look at North America (focusing on precipitation and T2m). In the following, we summarize our findings.

The decomposition separates the total variance about the long-term mean into four contributions consisting of the composite signal, the composite noise, the E2E signal and the E2E noise. Here, the noise contributions are associated with the (unpredictable) intra-ensemble variability, and the signals are associated with the ensemble mean – the part of the variance that is presumably predictable if the SSTs are known (specified as in this case, or more generally if they are predicted). In general, we find that for both El Niño and La Niña the largest contribution to the variance in the NH extra-tropics is the E2E noise, highlighting the challenge of predicting E2E differences in extra-tropical ENSO response. We found that the E2E noise is smaller for El Niño than for La Niña, especially for the 250mb eddy height over the North Pacific and parts of North America, indicating that E2E variability may be more predictable for warm events in those regions. Overall, the E2E signal tends to be somewhat weaker than the composite signal, though it tends to be spatially more extensive. The contribution to the variance from the composite noise tends to be small, reflecting the overall stability of the ensemble mean ENSO response (i.e., the composite mean ENSO response differs little between individual ensemble members).

Focusing on the fraction of the total variability associated with the composite and E2E signals (S/T), we found that the circulation impacts for La Niña events over North America appear not to be predictable beyond that which can be achieved from predicting the canonical (composite) La Niña response, while there is some hope that we can predict event-to-event differences in the responses over North America for El Niño. Here, the larger E2E (S/T) values for El Niño reflect the smaller E2E noise (mentioned above) rather than a larger E2E signal (compared with La Niña) for warm events. We also find little difference between El Niño and La Niña in the magnitude of the composite mean variance, which, if anything, is slightly larger for La Niña over North America. This appears to be inconsistent with Frauen et al. (2014), who showed that the higher signal-to-noise ratios for El Niño events are associated with a stronger El Niño response (signal) compared with the La Niña response. The lack of a stronger signal for El Niño in the current results appears to reflect a bias in this model's ENSO response over North America, which exhibits a remarkable (but unrealistic) degree of linearity (we find little difference between El Niño and La Niña in the strength of the response).

The decomposition of the covariances (specifically the correlations with the observations) into the contributions from E2E and the composite mean allows a direct assessment of the extent to which a model is capable of making predictions that are more skillful than simply predicting the composite mean response of a particular phenomenon. In fact, this assessment was one of the main motivations for developing the decomposition approach. Specifically, as highlighted in the Introduction, even for a much-studied and understood phenomenon such as ENSO, there are still concerns about the ability of current state-of-the-art climate models to predict event-to-event differences in the responses (e.g., WMO, 2015). Here, in addition to an assessment of the Dec-Mar average results, we also looked separately at the results for January and February—motivated by studies of Chen et al (2017) and Lim et al. (2020), who found that ENSO impacts over North America tend to be more skillful for February than for the other winter months (especially January).

The results of our analysis show, first, that we indeed do find evidence of greater skill over North America for February compared with January. This is especially clear for the 250mb eddy height and zonal wind, for which, for El Niño, the higher February skill is primarily the result of the higher skill associated with E2E variability (rather than the composite). In contrast, for La Niña (which exhibits little skill over North America associated with E2E variability) the greater skill in February (compared with January) is tied to differences in the skill of the composite responses. These results are also reflected in the skill of the precipitation responses over North America, though less so for T2m. Another interesting and somewhat unexpected result is the complementary nature of the spatial coverage of the correlations (or skill) for the two signals—the E2E impact tends to be larger in regions where the composite impact is smaller, and vice versa. This complementarity presumably reflects the fact that the E2E variability is in part tied to spatial shifts in the underlying composite ENSO response—something that would presumably focus the more predictable components of the E2E impacts on the edges of the composite responses where the gradients

are largest. In the tropical Pacific, we found considerably higher skill (correlations greater than 0.9) in reproducing the composite precipitation response for La Niña, compared with El Niño. In contrast, during El Niño events there is greater skill in reproducing the precipitation associated with E2E variability; indeed, for La Niña there is very little skill overall in reproducing E2E precipitation variability throughout the tropics and extra-tropics.

Turning next to an assessment of the quality and generality of the model results, we highlight a number of potential problems that must be kept in mind. First, it is clear that the model produces excessive circulation (250mb eddy height and 250mb zonal wind) variability over much of the North Pacific and North America, and this is very likely primarily associated with excessive E2E noise. This is especially true for El Niño, for which the observations show much reduced variability (compared with La Niña) over the North Pacific; the model, in contrast, shows roughly the same level of variance for both El Niño and La Niña. As noted above, this seems to reflect a greater linearity of the model's response to El Niño and La Niña than is warranted by the observations. Another important model deficiency is the excessive composite signal over North America for both El Niño and La Niña. This deficiency, especially evident in the 250mb eddy height and T2m fields, results in unrealistically large signal-to-total variance ratios in these quantities over the U.S. that are inconsistent with the relatively low skill (correlations with observations) there. In the tropics, the model produces excessive tropical Pacific precipitation variability, especially for El Niño. In fact, it is quite likely that this is a key factor in producing the excessive E2E noise variability in the extratropical circulation over the North Pacific and North America.

Finally, we turn to possible future work regarding the decomposition described here. The statistical significance of the results is clearly a concern. Given the somewhat complicated nature of the decomposition, a Monte Carlo approach is likely the most straightforward approach to assessing the uncertainty associated with the finite ensemble size, rather than attempting a more theoretical (parametric) assessment that would involve making various distributional assumptions. We indeed employ such a Monte Carlo approach in this study. Of course, with larger ensembles such concerns will become less of an issue. It is also important to note that the estimates of the various quantities involving the ensemble mean (the signal) presented here are not unbiased estimates. We have not attempted to derive unbiased estimates here, and it would indeed be a useful thing to pursue, though again that problem would also be alleviated as we increased the ensemble size, since the bias in the estimated signal is itself reduced with increasing ensemble size (e.g., Rowell et al. 1995).

Perhaps a more difficult issue to address concerns the estimates of the observed E2E variability, as such estimates are more fundamentally limited by (and presumably sensitive to) the length of the observational record. It would certainly be helpful to have more ENSO realizations than are currently available for the satellite era (since about 1980 or so). Model simulations/predictions that start in the middle of the 20th century would help, but there are of course concerns about the quality of the SST observations during that earlier period, and

(for initialized predictions) the quality of the atmospheric and land observations. Finally, it is not clear that ENSO statistics are stationary in a warming world. Accordingly, nature itself might put a limit on the robustness of the statistics involved in the decomposition described here.

Acknowledgements

This study was supported by NASA MAP funding under NNG17HP01C and WBS 802678.02.17.01.33. MERRA-2 data were developed by the Global Modeling and Assimilation Office (GMAO) at NASA GSFC under funding by the NASA MAP program and are disseminated through the Goddard Earth Science Data and Information Services Center (GES DISC). The GEOS model M2AMIP simulations were produced by the GMAO.

6. References

Bacmeister, J. T., and Stephens, G. L., 2011: Spatial statistics of likely convective clouds in CloudSat data, *J. Geophys. Res.*, 116, D04104, doi:10.1029/2010JD014444.

Barnston, A. G., Glantz, M. H., and He, Y., 1999: Predictive skill of statistical and dynamical climate models in forecasts of SST during the 1997−98 El Niño episode and the 1998 La Niña onset, Bull. Amer. Meteorol. Soc., 80, 217−244.

Bosilovich, M. G., and coauthors, 2015: MERRA-2: Initial evaluation of the climate**.** ***NASA/TM–2015–104606***, Vol. **43**, 139 pp.

Capotondi, A., and coauthors, 2015: Understanding ENSO diversity. *Bull. Amer. Meteor. Soc.*, **96**, 921-938, doi:10.1175/BAMS-D-13-00117.1.

Chen, L., H. van den Dool, E. Becker, and Q. Zhang, 2017: ENSO Precipitation and Temperature Forecasts in the North American Multimodel Ensemble: Composite Analysis and Validation. *J. Climate*, **30**, 1103–1125, https://doi.org/10.1175/JCLI-D-15-0903.1

Collow, Allison B. Marquardt, Sarith P. Mahanama, Michael G. Bosilovich, Randal D. Koster, and Siegfried D. Schubert, 2017**.** An Evaluation of Teleconnections Over the United States in an Ensemble of AMIP Simulations with the MERRA-2 Configuration of the GEOS Atmospheric Model. ***NASA Technical Report Series on Global Modeling and Data Assimilation, NASA/TM-2016-104606***, Vol. **47**, 68 pp. Document (5623 kB).

Cullather, R. I., S. M. J. Nowicki, B. Zhao, and M. J. Suarez, 2014: Evaluation of the Surface Representation of the Greenland Ice Sheet in a General Circulation Model. ***J. Climate***, **27**, 4835–4856, https://doi.org/10.1175/JCLI-D-13-00635.1.

Frauen, C., D. Dommenget, N. Tyrrell, M. Rezny, and S. Wales, 2014: Analysis of the nonlinearity of El Niño-Southern Oscillation teleconnections. *J. Climate*, **27**, 6225-6244, doi:10.1175/JCLI-D-13-00757.1.

Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). ***J. Climate***, **30**, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1.

Goddard, L., and M. Dilley, 2005: El Niño: Catastrophe or Opportunity. ***J. Climate***, **18**, 651–665, https://doi.org/10.1175/JCLI-3277.1.

Goddard, L., David G. DeWitt and David Legler, 2005: Seeking Progress in El Niño

Prediction. IRI report.

Hoerling, M. P., and A. Kumar, 1997: Why do North American climate anomalies differ from one El Niño event to another? *Geophys. Res. Lett*., **24**, 1059-1062.

Hoerling, M. P., A. Kumar, and M. Zhong, 1997: El Niño, La Niña, and the Nonlinearity of Their Teleconnections. *J. Climate*, **10**, 1769–1786, https://doi.org/10.1175/1520-0442(1997)010<1769:ENOLNA>2.0.CO;2.

Hoerling, M. P., A. Kumar, and T. Xu, 2001: Robustness of the Nonlinear Climate Response to ENSO's Extreme Phases. ***J. Climate***, **14**, 1277–1293, https://doi.org/10.1175/1520-0442(2001)014<1277:ROTNCR>2.0.CO;2.

Imada, Y., H. Tatebe, M. Ishii, Y. Chikamoto, M. Mori, M. Arai, M. Watanabe, and M. Kimoto, 2015: Predictability of Two Types of El Niño Assessed Using an Extended Seasonal Prediction System by MIROC. *Mon. Wea. Rev.*, **143**, 4597–4617, https://doi.org/10.1175/MWR-D-15-0007.1.

Kao, H.-Y., and J.-Y. Yu, 2009: Contrasting Eastern-Pacific and Central-Pacific types of ENSO. *J. Climate*, **22**, 615-632, doi:10.1175/2008JCLI2309.1.

Kirtman, B. P., J. Shukla, M. Balmaseda, N. Graham, C. Penland, Y. Xue, and S. Zebiak, 2000. Current status of ENSO forecast skill. A Report to the CLIVAR Working Group on Seasonal to Interannual Prediction, 26pp.

Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, 2000: A catchment-based approach to modeling land surface processes in a general model: 1. Model structure. *J. Geophys. Res*., **105(D20)**, 24809-24822, doi:10.1029/2000JD900327.

Kug, J.-S., F.-F. Jin, and S.-I. An, 2009: Two-types of El Niño events: Cold tongue El Niño and warm pool El Niño. *J. Climate*, **22**, 1499–1515, doi:10.1175/2008JCLI2624.1.

Kumar, A., M. Hoerling, M. Ji, A. Leetmaa, and P. Sardeshmukh, 1996: Assessing a GCM's Suitability for Making Seasonal Predictions. ***J. Climate***, **9**, 115–129, https://doi.org/10.1175/1520-0442(1996)009<0115:AAGSFM>2.0.CO;2.

Kumar, A., and M. P. Hoerling, 1997: Interpretation and Implications of the Observed Inter–El Niño Variability. ***J. Climate***, **10**, 83–91, https://doi.org/10.1175/1520-0442(1997)010<0083:IAIOTO>2.0.CO;2.

Lim, Y.-K., S.D. Schubert, Y. Chang, and H. Wang, 2020: The Boreal Winter El Niño Precipitation Response over North America: Insights into why January is more difficult to predict than February. *Accepted in J. Climate*.

Marquardt Collow, A. B., R. I. Cullather, and M. G. Bosilovich, 2020: Recent Arctic Ocean Surface Air Temperatures in Atmospheric Reanalyses and Numerical Simulations. ***J. Climate***, **33**, 4347–4367, https://doi.org/10.1175/JCLI-D-19-0703.1.

Molod, A. M., L. Takacs, M. Suarez, and J. Bacmeister, 2015: Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2. *Geosci. Model Dev.*, **8**, 1339-1356, doi:10.5194/gmd-8-1339-2015.

Moorthi, S., and M. J. Suarez, 1992: Relaxed Arakawa-Schubert: A parameterization of moist convection for general circulation models. *Mon. Wea. Rev.*, **120**, 978-1002.

Reichle, R. H., Q. Liu, R. D. Koster, C. S. Draper, S. P. P. Mahanama, and G. S. Partyka, 2017: Land Surface Precipitation in MERRA-2. *Journal of Climate*, **30**, 1643–1664, doi:10.1175/jcli-d-16-0570.1.

Rienecker, M.M., and Coauthors, 2011: MERRA - NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624-3648. doi: 10.1175/JCLI-D-11-00015.1.

Rowell, D. P., Folland, C. K., Maskell, K. and Ward, M. N., 1995: Variability of summer rainfall over tropical north Africa (1906–1992): observations and modelling. *Q. J. R. Meteorol. Soc.*, **121**, 669–704.

Sohn, S., Tam, C. & Kug, J. How does ENSO diversity limit the skill of tropical Pacific precipitation forecasts in dynamical seasonal predictions?. *Clim Dyn* **53,** 5815–5831 (2019). https://doi.org/10.1007/s00382-019-04901-2.

Wang, L., Robertson, A.W. Week 3–4 predictability over the United States assessed from two operational ensemble prediction systems. *Clim Dyn* **52,** 5861–5875 (2019). https://doi.org/10.1007/s00382-018-4484-9

WMO, 2015: Progress in Observing and Predicting ENSO, WMO Bulletin, vol 64 (1), March 18, 2015. https://public.wmo.int/en/resources/bulletin/progress-observing-and-predicting-enso-0

Zhang T, Perlwitz J, Hoerling MP (2014) What is responsible for the strong observed asymmetry in teleconnections between El Niño and La Niña? Geophys Res Lett 41:1019–1025. https://doi.org/10.1002/2013GL058964

Previous Volumes in This Series

Volume 1 September 1994	Documentation of the Goddard Earth Observing System (GEOS) general circulation model - Version 1 **L.L. Takacs, A. Molod, and T. Wang**
Volume 2 October 1994	Direct solution of the implicit formulation of fourth order horizontal diffusion for gridpoint models on the sphere **Y. Li, S. Moorthi, and J.R. Bates**
Volume 3 December 1994	An efficient thermal infrared radiation parameterization for use in general circulation models **M.-D. Chou and M.J. Suarez**
Volume 4 January 1995	Documentation of the Goddard Earth Observing System (GEOS) Data Assimilation System - Version 1 **James Pfaendtner, Stephen Bloom, David Lamich, Michael Seablom, Meta Sienkiewicz, James Stobie, and Arlindo da Silva**
Volume 5 April 1995	Documentation of the Aries-GEOS dynamical core: Version 2 **Max J. Suarez and Lawrence L. Takacs**
Volume 6 April 1995	A Multiyear Assimilation with the GEOS-1 System: Overview and Results **Siegfried Schubert, Chung-Kyu Park, Chung-Yu Wu, Wayne Higgins, Yelena Kondratyeva, Andrea Molod, Lawrence Takacs, Michael Seablom, and Richard Rood**
Volume 7 September 1995	Proceedings of the Workshop on the GEOS-1 Five-Year Assimilation **Siegfried D. Schubert and Richard B. Rood**
Volume 8 March 1996	Documentation of the Tangent Linear Model and Its Adjoint of the Adiabatic Version of the NASA GEOS-1 C-Grid GCM: Version 5.2 **Weiyu Yang and I. Michael Navon**
Volume 9 March 1996	Energy and Water Balance Calculations in the Mosaic LSM **Randal D. Koster and Max J. Suarez**

Volume 10 April 1996	Dynamical Aspects of Climate Simulations Using the GEOS General Circulation Model **Lawrence L. Takacs and Max J. Suarez**
Volume 11 May 1997	Documentation of the Tangent Linear and Adjoint Models of the Relaxed Arakawa-Schubert Moisture Parameterization Package of the NASA GEOS-1 GCM (Version 5.2) **Weiyu Yang, I. Michael Navon, and Ricardo Todling**
Volume 12 August 1997	Comparison of Satellite Global Rainfall Algorithms **Alfred T.C. Chang and Long S. Chiu**
Volume 13 December 1997	Interannual Variability and Potential Predictability in Reanalysis Products **Wie Ming and Siegfried D. Schubert**
Volume 14 August 1998	A Comparison of GEOS Assimilated Data with FIFE Observations **Michael G. Bosilovich and Siegfried D. Schubert**
Volume 15 June 1999	A Solar Radiation Parameterization for Atmospheric Studies **Ming-Dah Chou and Max J. Suarez**
Volume 16 November 1999	Filtering Techniques on a Stretched Grid General Circulation Model **Lawrence Takacs, William Sawyer, Max J. Suarez, and Michael S. Fox-Rabinowitz**
Volume 17 July 2000	Atlas of Seasonal Means Simulated by the NSIPP-1 Atmospheric GCM **Julio T. Bacmeister, Philip J. Pegion, Siegfried D. Schubert, and Max J. Suarez**
Volume 18 December 2000	An Assessment of the Predictability of Northern Winter Seasonal Means with the NSIPP1 AGCM **Philip J. Pegion, Siegfried D. Schubert, and Max J. Suarez**
Volume 19 July 2001	A Thermal Infrared Radiation Parameterization for Atmospheric Studies **Ming-Dah Chou, Max J. Suarez, Xin-Zhong Liang, and Michael M.-H. Yan**

Volume 20 August 2001	The Climate of the FVCCM-3 Model **Yehui Chang, Siegfried D. Schubert, Shian-Jiann Lin, Sharon Nebuda, and Bo-Wen Shen**
Volume 21 September 2001	Design and Implementation of a Parallel Multivariate Ensemble Kalman Filter for the Poseidon Ocean General Circulation Model **Christian L. Keppenne and Michele M. Rienecker**
Volume 22 August 2002	A Coupled Ocean-Atmosphere Radiative Model for Global Ocean Biogeochemical Models **Watson W. Gregg**
Volume 23 November 2002	Prospects for Improved Forecasts of Weather and Short-term Climate Variability on Subseasonal (2-Week to 2-Month) Time Scales **Siegfried D. Schubert, Randall Dole, Huang van den Dool, Max J. Suarez, and Duane Waliser**
Volume 24 July 2003	Temperature Data Assimilation with Salinity Corrections: Validation for the NSIPP Ocean Data Assimilation System in the Tropical Pacific Ocean, 1993–1998 **Alberto Troccoli, Michele M. Rienecker, Christian L. Keppenne, and Gregory C. Johnson**
Volume 25 December 2003	Modeling, Simulation, and Forecasting of Subseasonal Variability **Duane Waliser, Siegfried D. Schubert, Arun Kumar, Klaus Weickmann, and Randall Dole**
Volume 26 April 2005	Documentation and Validation of the Goddard Earth Observing System (GEOS) Data Assimilation System – Version 4 **Senior Authors: S. Bloom, A. da Silva and D. Dee** **Contributing Authors: M. Bosilovich, J-D. Chern, S. Pawson, S. Schubert, M. Sienkiewicz, I. Stajner, W-W. Tan, and M-L. Wu**
Volume 27 December 2008	The GEOS-5 Data Assimilation System - Documentation of Versions 5.0.1, 5.1.0, and 5.2.0. **M.M. Rienecker, M.J. Suarez, R. Todling, J. Bacmeister, L. Takacs, H.-C. Liu, W. Gu, M. Sienkiewicz, R.D. Koster, R. Gelaro, I. Stajner, and J.E. Nielsen**

Volume 28 April 2012	The GEOS-5 Atmospheric General Circulation Model: Mean Climate and Development from MERRA to Fortuna **Andrea Molod, Lawrence Takacs, Max Suarez, Julio Bacmeister, In-Sun Song, and Andrew Eichmann**
Volume 29 June 2012	Atmospheric Reanalyses – Recent Progress and Prospects for the Future. A Report from a Technical Workshop, April 2010 **Michele M. Rienecker, Dick Dee, Jack Woollen, Gilbert P. Compo, Kazutoshi Onogi, Ron Gelaro, Michael G. Bosilovich, Arlindo da Silva, Steven Pawson, Siegfried Schubert, Max Suarez, Dale Barker, Hirotaka Kamahori, Robert Kistler, and Suranjana Saha**
Volume 30 December 2012	The GEOS-iODAS: Description and Evaluation **Guillaume Vernieres, Michele M. Rienecker, Robin Kovach and Christian L. Keppenne**
Volume 31 March 2013	Global Surface Ocean Carbon Estimates in a Model Forced by MERRA **Watson W. Gregg, Nancy W. Casey and Cecile S. Rousseaux**
Volume 32 March 2014	Estimates of AOD Trends (2002-2012) over the World's Major Cities based on the MERRA Aerosol Reanalysis **Simon Provencal, Pavel Kishcha, Emily Elhacham, Arlindo M. da Silva, and Pinhas Alpert**
Volume 33 August 2014	The Effects of Chlorophyll Assimilation on Carbon Fluxes in a Global Biogeochemical Model **Cécile S. Rousseaux and Watson W. Gregg**
Volume 34 September 2014	Background Error Covariance Estimation using Information from a Single Model Trajectory with Application to Ocean Data Assimilation into the GEOS-5 Coupled Model **Christian L. Keppenne, Michele M. Rienecker, Robin M. Kovach, and Guillaume Vernieres**
Volume 35 December 2014	Observation-Corrected Precipitation Estimates in GEOS-5 **Rolf H. Reichle and Qing Liu**

Volume 36
March 2015

Evaluation of the 7-km GEOS-5 Nature Run

Ronald Gelaro, William M. Putman, Steven Pawson, Clara Draper, Andrea Molod, Peter M. Norris, Lesley Ott, Nikki Prive, Oreste Reale, Deepthi Achuthavarier, Michael Bosilovich, Virginie Buchard, Winston Chao, Lawrence Coy, Richard Cullather, Arlindo da Silva, Anton Darmenov, Ronald M. Errico, Marangelly Fuentes, Min-Jeong Kim, Randal Koster, Will McCarty, Jyothi Nattala, Gary Partyka, Siegfried Schubert, Guillaume Vernieres, Yuri Vikhliaev, and Krzysztof Wargan

Volume 37
March 2015

Maintaining Atmospheric Mass and Water Balance within Reanalysis

Lawrence L. Takacs, Max Suarez, and Ricardo Todling

Volume 38
September 2015

The Quick Fire Emissions Dataset (QFED) – Documentation of versions 2.1, 2.2 and 2.4

Anton S. Darmenov and Arlindo da Silva

Volume 39
September 2015

Land Boundary Conditions for the Goddard Earth Observing System Model Version 5 (GEOS-5) Climate Modeling System - Recent Updates and Data File Descriptions

Sarith Mahanama, Randal Koster, Gregory Walker, Lawrence Takacs, **Rolf**

Reichle, Gabrielle De Lannoy, Qing Liu, Bin Zhao, and Max **Suarez**

Volume 40
October 2015

Soil Moisture Active Passive (SMAP) Project Assessment Report for the Beta-Release L4_SM Data Product

Rolf H. Reichle, Gabrielle J. M. De Lannoy, Qing Liu, Andreas Colliander,

Austin Conaty, Thomas Jackson, John Kimball, and Randal D. Koster

Volume 41
October 2015

GDIS Workshop Report

Siegfried Schubert, Will Pozzi, Kingtse Mo, Eric Wood, Kerstin Stahl, Mike Hayes, Juergen Vogt, Sonia Seneviratne, Ron Stewart, Roger Pulwarty, and Robert Stefanski

Volume 42
November 2015

Soil Moisture Active Passive (SMAP) Project Calibration and Validation for the L4_C Beta-Release Data Product

John Kimball, Lucas Jones, Joseph Glassy, E. Natasha Stavros, Nima Madani, Rolf Reichle, Thomas Jackson, and Andreas Colliander

Volume 43 September 2015	MERRA-2: Initial Evaluation of the Climate **Michael G. Bosilovich, Santha Akella, Lawrence Coy, Richard Cullather, Clara Draper, Ronald Gelaro, Robin Kovach, Qing Liu, Andrea Molod, Peter Norris, Krzysztof Wargan, Winston Chao, Rolf Reichle, Lawrence Takacs, Yury Vikhliaev, Steve Bloom, Allison Collow, Stacey Firth, Gordon Labow, Gary Partyka, Steven Pawson, Oreste Reale, Siegfried Schuber**t, **and Max Suarez**
Volume 44 February 2016	Estimation of the Ocean Skin Temperature using the NASA GEOS Atmospheric Data Assimilation System **Santha Akella, Ricardo Todling, Max Suarez**
Volume 45 October 2016	The MERRA-2 Aerosol Assimilation **C. A. Randles, A. M. da Silva, V. Buchard, A. Darmenov, P. R. Colarco, V. Aquila, H. Bian, E. P. Nowottnick, X. Pan, A. Smirnov, H. Yu, and R. Govindaraju**
Volume 46 October 2016	The MERRA-2 Input Observations: Summary and Assessment **Will McCarty, Lawrence Coy, Ronald Gelaro, Albert Huang, Dagmar Merkova, Edmond B. Smith, Meta Sienkiewicz, and Krzysztof Wargan**
Volume 47 **May 2017**	An Evaluation of Teleconnections Over the United States in an Ensemble of AMIP Simulations with the MERRA-2 Configuration of the GEOS Atmospheric Model. **Allison B. Marquardt Collow, Sarith P. Mahanama, Michael G. Bosilovich, Randal D. Koster, and Siegfried D. Schubert**
Volume 48 **July 2017**	Description of the GMAO OSSE for Weather Analysis Software Package: Version 3 **Ronald M. Errico, Nikki C. Prive, David Carvalho, Meta Sienkiewicz, Amal El Akkraoui, Jing Guo, Ricardo Todling, Will McCarty, William M. Putman, Arlindo da Silva, Ronald Gelaro, and Isaac Moradi**
Volume 49 **March 2018**	Preliminary Evaluation of Influence of Aerosols on the Simulation of Brightness Temperature in the NASA Goddard Earth Observing System Atmospheric Data Assimilation System **Jong Kim, Santha Akella, Will McCarty, Ricardo Todling, and Arlindo M. da Silva**

Volume 50 **March 2018**	The GMAO Hybrid Ensemble-Variational Atmospheric Data Assimilation System: Version 2.0 **Ricardo Todling and Amal El Akkraoui**
Volume 51 **July 2018**	The Atmosphere-Ocean Interface Layer of the NASA Goddard Earth Observing System Model and Data Assimilation System **Santha Akella and Max Suarez**
Volume 52 **July 2018**	Soil Moisture Active Passive (SMAP) Project Assessment Report for Version 4 of the L4_SM Data Product **Rolf H. Reichle, Qing Liu, Randal D. Koster, Joe Ardizzone, Andreas Colliander, Wade Crow, Gabrielle J. M. De Lannoy, and John Kimball**
Volume 53 **October 2019**	Ensemble Generation Strategies Employed in the GMAO GEOS-S2S Forecast System **Siegfried Schubert, Anna Borovikov, Young-Kwon Lim, and Andrea Molod**
Volume 54 **August 2020**	Position Estimation of Atmospheric Motion Vectors for Observation System Simulation Experiments **David Carvalho, Will McCarty**