Supporting Information for "Singular Spectrum Analysis with Conditional Predictions for Real-Time State Estimation and Forecasting"

H. Reed Ogrosky[1], Samuel N. Stechmann[2,3], Nan Chen[2], and

Andrew J. Majda[4,5]

[1]Department of Mathematics and Applied Mathematics, Virginia Commonwealth University, Richmond, VA, USA

[2]Department of Mathematics, University of Wisconsin-Madison, Madison, WI, USA

[3]Department of Atmospheric and Oceanic Sciences, University of Wisconsin-Madison, Madison, WI, USA

[4]Department of Mathematics and Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, New York

[5]Center for Prototype Climate Modeling, NYU Abu Dhabi, Saadiyat Island, Abu Dhabi, United Arab Emirates

Contents of this file

Corresponding author: H. R. Ogrosky, Department of Mathematics and Applied Mathematics, Virginia Commonwealth University, 1015 Floyd Ave., Box 842014, Richmond, VA 23284-2014, USA. (hrogrosky@vcu.edu)

Introduction

This Supporting Information describes details of the datasets, methods, and models used in the main manuscript. The text sections are organized as follows:

Text S1. Calculation of pattern correlation and RMSE

Text S2. Statistics of precipitation anomalies

Text S3. Partially-observed multiscale model

Text S1. Calculation of pattern correlation and RMSE

To quantify the agreement between reconstructions, the root-mean-squared error (RMSE) and pattern correlation for spatial dimension $d = 1, ..., D$ are defined by

$$\text{RMSE}^{(r)}(j,d) = \left(\frac{1}{|I|} \sum_{i \in I} \left[(w_{r,d}(N_i + j) - u_{r,d}(N_i + j))^2 \right] \right), \tag{1a}$$

$$\text{Corr}^{(r)}(j,d) = \frac{\sum_{i \in I} \left(w_{r,d}(N_i + j) u_{r,d}(N_i + j) \right)}{\left[\sum_{i \in I} \left(w_{r,d}^2(N_i + j) \right) \times \sum_{i \in I} \left(u_{r,d}^2(N_i + j) \right) \right]^{1/2}}, \tag{1b}$$

where $\vec{w}_r(t) = (w_{r,1}(t), ..., w_{r,D}(t)) = \sum_{i=1}^{r} \vec{z}_i(t)$ is the reconstruction (using either traditional or SSA-CP method) of modes $1, ..., r$ without future knowledge; $\vec{u}_r(t) = (u_{r,1}(t), ..., u_{r,D}(t))$ is the reconstruction of modes $1, ..., r$ using future knowledge, i.e. the 'truth'; $|I|$ is the size of I; and where $j \in [-M, M]$ (and M is the embedding window).

For 2-dimensional data, a bivariate RMSE and pattern correlation defined by

$$\text{RMSE}^{(r)}(j) = \left(\frac{1}{|I|} \sum_{i \in I} \left[(w_{r,1}(N_i + j) - u_{r,1}(N_i + j))^2 + (w_{r,2}(N_i + j) - u_{r,2}(N_i + j))^2 \right] \right), \tag{2a}$$

$$\text{Corr}^{(r)}(j) = \frac{\sum_{i \in I} \left(w_{r,1}(N_i + j) u_{r,1}(N_i + j) + w_{r,2}(N_i + j) u_{r,2}(N_i + j) \right)}{\left[\sum_{i \in I} \left(w_{r,1}^2(N_i + j) + w_{r,2}^2(N_i + j) \right) \times \sum_{i \in I} \left(u_{r,1}^2(N_i + j) + u_{r,2}^2(N_i + j) \right) \right]^{1/2}}, \tag{2b}$$

are used.

Text S2. Statistics of precipitation anomalies

In this section, some statistics are presented for precipitation anomalies that are used in the second test discussed in the main text. These anomalies were computed from GPCP daily precipitation data (Huffman *et. al.*, 2012) which has a spatial resolution of $1^\circ \times 1^\circ$; the portion from 1 January 1997 through 31 December 2013 is used. Prior to applying SSA, a meridional mode truncation, the removal of annual mean and seasonal cycle, and interpolation to 64 equally-spaced zonal gridpoints were applied to the data; see, e.g., Ogrosky and Stechmann (2015); Stechmann and Majda (2015); Stechmann and Ogrosky (2014) for details of these steps.

Figure S1 shows the variance, skewness, and kurtosis of these anomalies as a function of longitude. At all longitudes the data has positive skewness and kurtosis greater than 3. A PDF of the precipitation anomalies at $x = 180$ is also shown along with its Gaussian fit.

Text S3. Partially-observed multiscale model

In this section, details of the multiscale model test in the main text are presented. The multiscale model used is

$$du_1 = (-\gamma_1 u_1 + F(t))\, dt + \sigma_1 dW_1, \tag{3a}$$

$$du_2 = (-\gamma_2 + i\omega_0/\epsilon + ia_0 u_1)\, u_2 dt + \sigma_2 dW_2, \tag{3b}$$

where $\gamma_1 = \gamma_2 = 0.2$, $\sigma_1 = \sigma_2 = 0.5$, $\omega_0 = a_0 = 1$, $\epsilon = 0.5$, and $F(t) = \sin(t/5)$; see Majda and Harlim (2012) for additional details. An approximate solution was calculated

numerically with the Euler-Maruyama method using $dt = 0.005$ and $t_{end} = 2000$. The real part of u_2 was then sampled every 0.5 time units to create a dataset with $D = 1$ and $N_{tot} = 4000$. A portion of the real part of $u_2(t)$ is shown in Figure S2. The pattern correlation and RMSE results using (i) the traditional reconstruction and (ii) SSA-CP on the model (3) are shown in Figure S3.

References

Huffman, G.J., Bolvin, D.T. & Adler, R.F., 2012: GPCP Version 2.2 SG Combined Precipitation Data Set. WDC-A, NCDC, Asheville, NC. Data set accessed 12 February 2014 at http://www.ncdc.noaa.gov/oa/wmo/wdcamet-ncdc.html.

Majda, A. J. and J. Harlim, 2012. *Filtering Complex Turbulent Systems*, Cambridge University Press.

Ogrosky, H. R. and S. N. Stechmann, 2015: Assessing the equatorial long-wave approximation: asymptotics and observational data analysis. *J. Atmos. Sci* **72**, 4821-4843.

Stechmann, S.N. and A.J. Majda, 2015: Identifying the skeleton of the Madden-Julian oscillation in observational data. *Mon. Wea. Rev.* **143**, 395-416.

Stechmann, S.N. and H.R. Ogrosky, 2014: The Walker circulation, diabatic heating, and outgoing longwave radiation. *Geophys. Res. Lett.* **41**, 9097-9105.

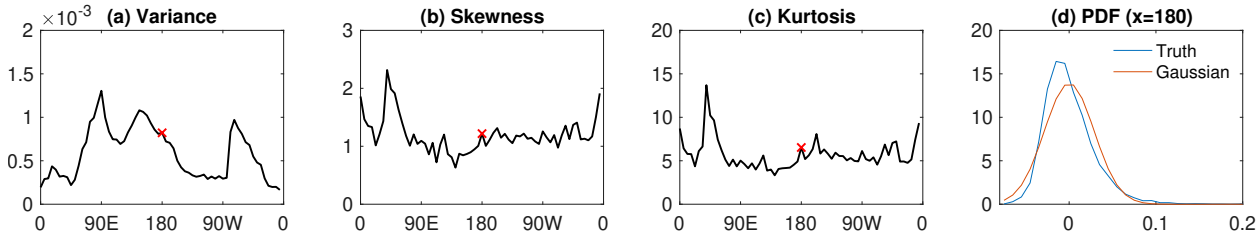

Figure S1. (a) Variance, (b) skewness, and (c) kurtosis of precipitation anomalies as a function of longitude. (d) PDF of precipitation anomalies at $x = 180$ and its Gaussian fit.

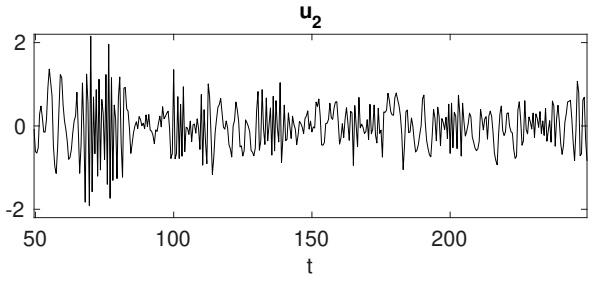

Figure S2. Portion of the real part of $u_2(t)$ from multiscale model (3).

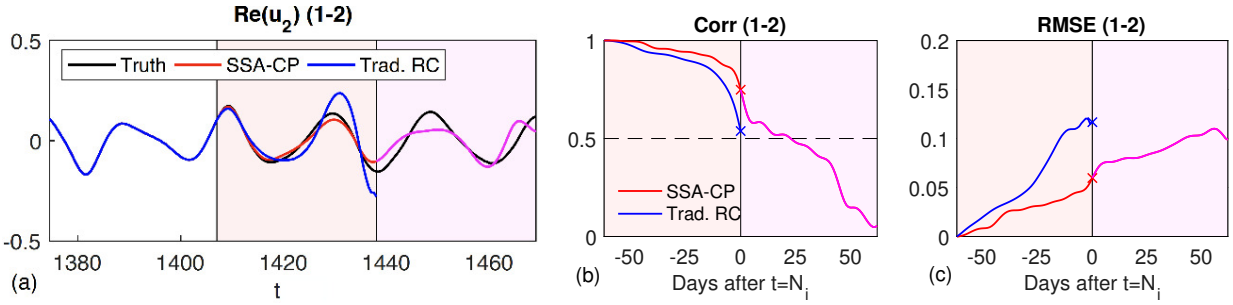

Figure S3. (a) Reconstructed real part of $u_2(t)$ from partially-observed multiscale model (3) with components 1-2, using (blue) traditional reconstruction, (red) SSA-CP, and (black) reconstruction using future information. (b) Pattern correlation and (c) RMSE for partially-observed multiscale model (3) with u_2 observed.

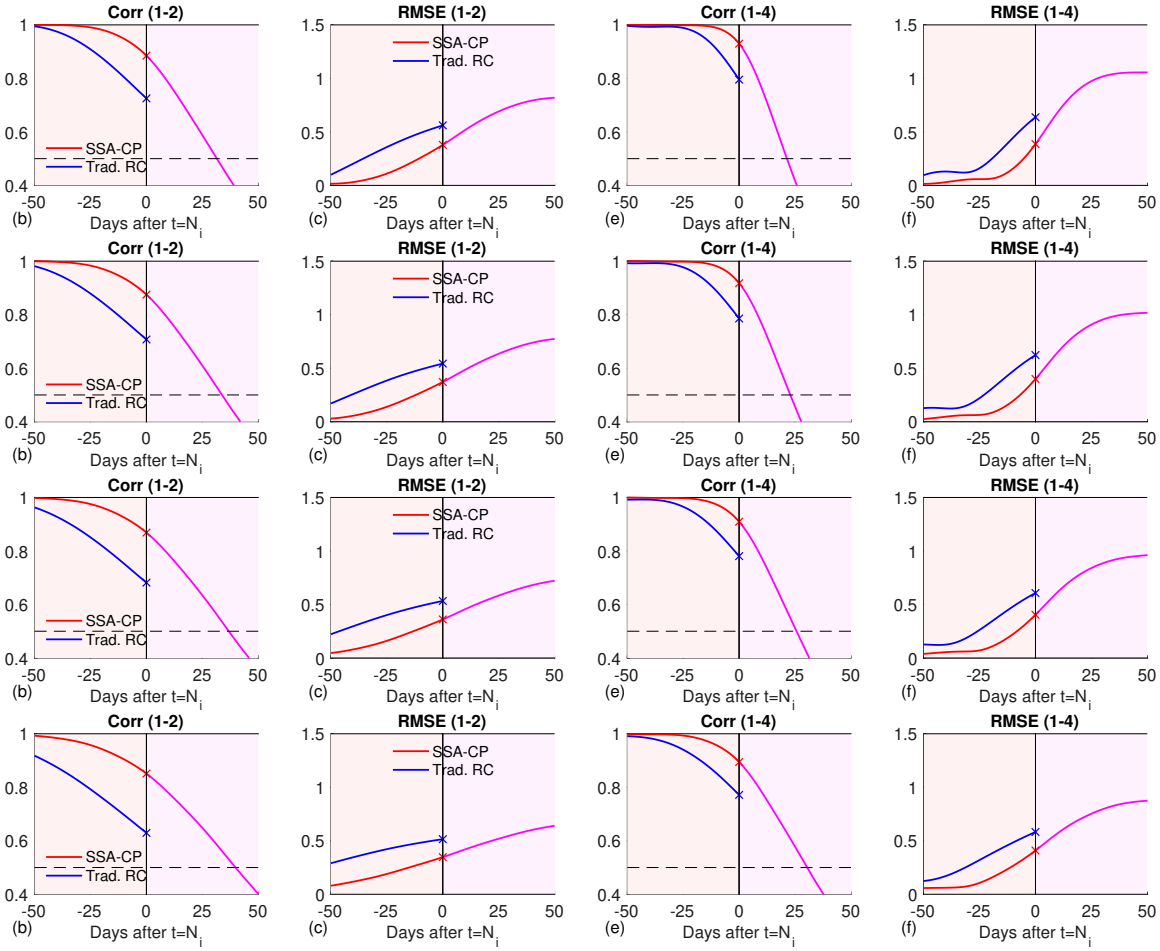

Figure S4. (Top row) Same as Figure 4(b,c,e,f) in the main text, but with $M = 61$. (Bottom three rows) Same as top row, but with $M = 71$, $M = 81$, and $M = 101$, respectively.

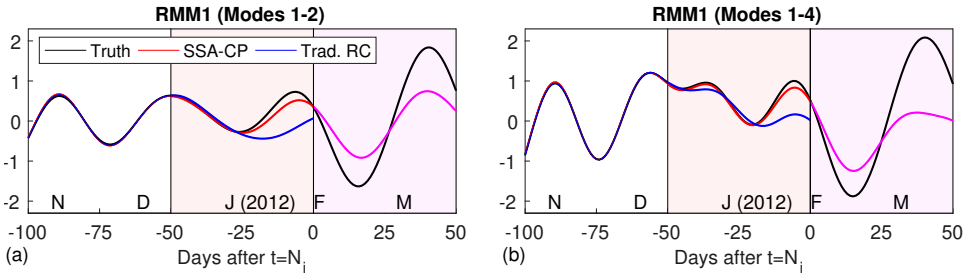

Figure S5. Same as Figure 1 in the main text, but with leading SSA modes computed using a training period of data consisting of 01-Jan-1999 through 13-Aug-2009.