Responses to Editor and Reviewers

Thank you, Editor (Dr. Olivier Talagrand), for overseeing the review of my manuscript, for providing invaluable feedback, and for granting my request for a deadline extension. I am also indebted to my Reviewers (Dr. Ian Grooms and Dr. Lili Lei) for their thorough review, commentary, feedback and kind words. This review process has substantially and obviously improved my manuscript's quality and clarity. I have expressed my gratitude to Dr. Talagrand, Dr. Grooms, and Dr. Lei in my acknowledgements.

1 Responses to Reviewer 1 (Ian Grooms)

Thank you, Ian, for taking the time to re-review my manuscript thoroughly and for sending me an email regarding my typographical errors. Because I could not locate your comments on EGUSphere, I am responding to your emailed comments below. I have made every effort to address your comments in the hopes that this will bring my manuscript to acceptance.

1.1 Minor Comments

1.1.1 page 8: "Note that this four-stage assumes" -> "Note that this four-stage procedure assumes"

I have corrected this error. Thank you for catching that.

1.2 Line 301ff: The Lorenz & Emmanuel 1998 paper on the L96 model identifies 0.05 model time units as comparable to 6 hours of synoptic-scale weather not arbitrarily but based on matching the predictability characteristics of the two systems. I searched the Anderson 2019 and 2023 papers and didn't see where Jeff says that 0.05 MTU equals 1 hour. This is not important, but I wanted to give you the feedback.

Thank you for the feedback. I have removed the offending paragraph and modified the previous paragraph slightly (see below).

"The L96 model uses 40 variables (i.e., 40 grid points in a ring), a forcing parameter value of 8 (i.e., $F = 8$), and a time-step of 0.05 L96 time units. The L96 time unit is henceforth referred to as τ. All results in this paper will be displayed and discussed in terms of τ. Forward time integration of the model is done via the Runge-Kutta fourth-order integration scheme. Every OSSE experiment runs for 5,500 cycles. Initial nature run states and the initial ensemble members are drawn from the L96's climatology. "

1.3 Thanks for the acknowledgment, but you acknowledge Groom not Grooms! Just a typo.

Thanks you for spotting that! The issue has been fixed.

2 Responses to Reviewer 2 (Lili Lei)

Thank you, Lili, for re-reviewing my manuscript, and for recommending acceptance. I am grateful for your thorough review and suggestions. I have acknowledged your contributions to this manuscript in my Acknowledgements section.

3 Responses to the Editor (Lili Lei)

3.1 General comments

Two referees have sent their evaluations. They are the same as the referees of the previous version (referee 1 is I. Grooms, and referee 2, who does not want to remain anonymous any more, is Lili Lei, from Nanjing University).
Both of them consider the author has satisfactorily responded to their concerns, and recommend acceptance of the paper. Referee 1 just mentions that there are typos to be corrected (the paper will in any case go through copy-editing).
I follow the referees' recommendation, and accept the paper. I however as editor still have a few suggestions for modifications.

Thank you for summarizing the two reviewers' comments, for your thorough review, and for overseeing the review of my manuscript. I have added an acknowledgement of Lili in my manuscript's Acknowledgements section. I have made every effort to incorporate your suggestions in the hopes of my manuscript's acceptance.

3.2 Editorial Suggestions

3.2.1 L. 268. It would be preferable to give a scale of comparison for the variances of the observation errors. My colleague Mohamed Jardak and myself (Jardak and Talagrand, 2018) found a 'climatological' variance of 10^3 for the Lorenz 96 model, to be compared with the value 1 used here by the author for the IDEN observations (but the numerical conditions of the experiments may not be the same).

Thank you for encouraging me to explore the climatological variances of my three observation operators. I have estimated those variances by constructing a 10,000-member climatological ensemble. This ensemble is constructed by initializing those members with noise samples drawn from $N(\mathbf{0}, \boldsymbol{I})$, and then integrating for 5000 time steps (i.e., 100 model time units).

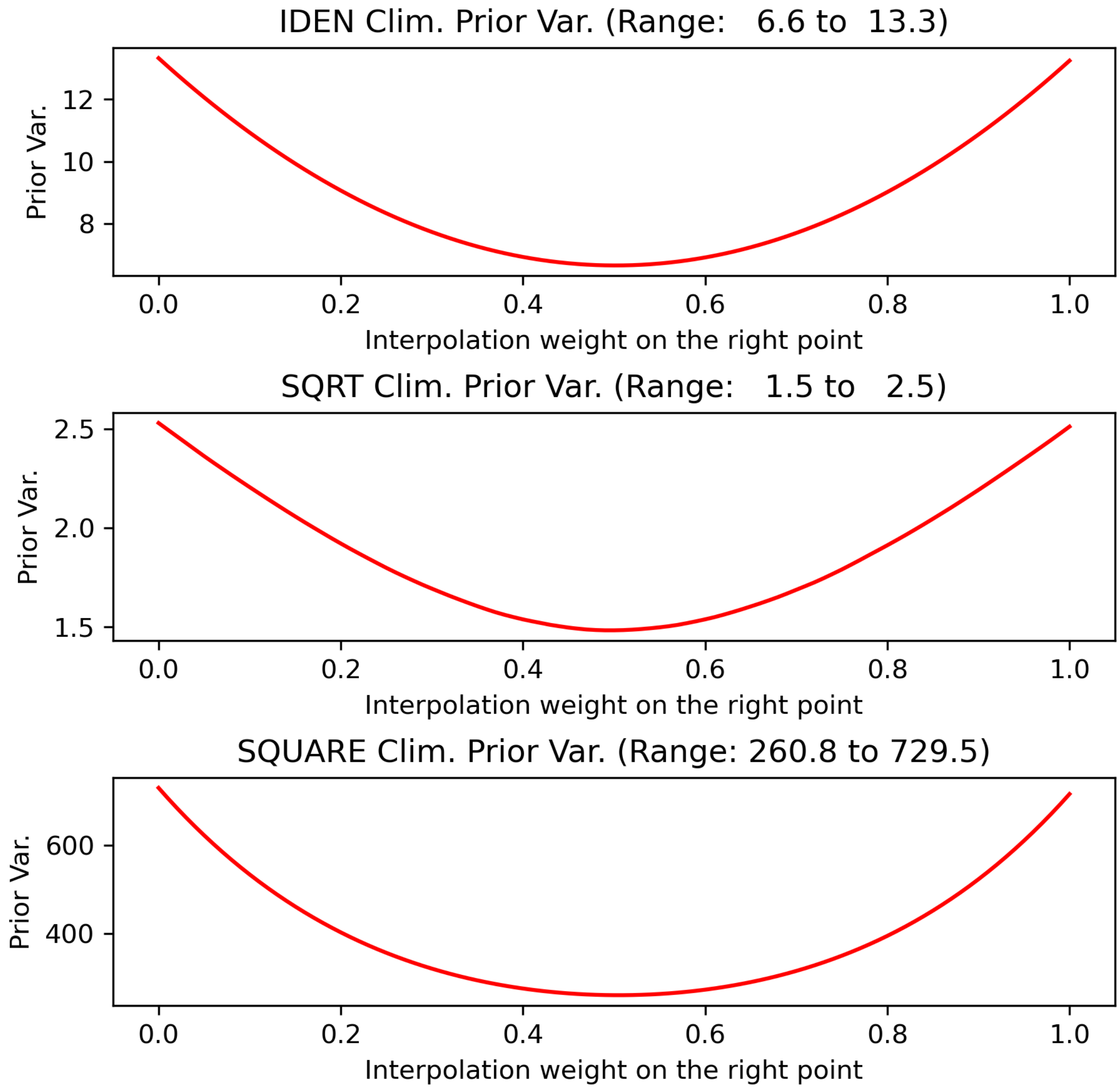

Figure 1: Ensemble-estimated climatological prior variances of (a) the IDEN observation operator, (b) the SQRT observation operator, and (c) the SQUARE observation operator. Because the observation sites are in-between Lorenz 1996 model grid points, the climatological variances vary depending on where the sites are between the two points. As such, the variances here are plotted as a function of the interpolation weight from the grid point to the right of the observation site.

Note that my variances are complicated by the fact that I am linearly interpolating from the 40 Lorenz model grid points to in-between grid point values. As such, I have computed the climatological variances as a function of the interpolation weight applied to the grid point to the right of the observation site

(Figure 1).

The IDEN observation's climatological variance is between 6.6 to 13.5. This means my observation error variance for the IDEN observations is between 7.4% to 15% of the climatological variance. As for the SQRT observations, my observation error variance of 0.25 is between 10% to 17% of the SQRT observation's climatological variance. Finally, the observation error variance for my SQUARE observations is 16, which is between 2% to 6% of the corresponding climatological error variance.

I have added the following sentence to the paragraph describing the observation error variance (Section 4.1):
"*... the chosen σ^2 for the IDEN, SQRT and SQUARE observations are 1.0, 0.25, and 16, respectively. Note that IDEN's σ^2 is between 7.4% to 15% of the climatological IDEN error variance, SQRT's σ^2 is between 10% to 17% of the climatological SQRT error variance, and SQUARE's σ^2 is between 2% to 6% of the climatological SQUARE error variance.*"

3.2.2 L. 217. Figure 5 indicates that the virtual members have better ensemble statistics than the forecast ensemble. What do you mean by better ? The CDFs shown in Fig. 5 from the virtual members are smoother than the CDFs obtained from the forecast members. Is that what you mean, or what ?

Thank you for encouraging me to be clearer on the matter. I have clarified my meaning in my manuscript (see below).
"*... Figure 5 indicates that the virtual members' CDFs and x-y relationship are closer to the true CDFs and relationship than those of the forecast members – the virtual members' curves have visibly less distances from the true curves than the forecast members' curves. In other words, the virtual members have better ensemble statistics than the forecast ensemble. This improvement in ensemble statistics ...*"

3.2.3 Ll. 73-74. I suggest to state more precisely what the notation Chol(C) exactly means. I understand it denotes the lower triangular matrix of the Cholesky decomposition C = U UT of the matrix C, but is it U or UT that is lower triangular (that may be irrelevant, but may nevertheless matter for a reader who wants to implement the algorithm)?

Thank you for encouraging me to be clearer on the subject. I have clarified that in my manuscript (see below).

*"The CAC2020 algorithm constructs N_V virtual members from N_E ensemble members using a three-step procedure. First, an $N_E \times N_V$ matrix of linear combination coefficients (**E**) is generated by evaluating*

$$\boldsymbol{E} \equiv \gamma \mathbf{1}_{\boldsymbol{N_E} \times \boldsymbol{N_V}} + \boldsymbol{L}_{\boldsymbol{C_E}} \left\{ \boldsymbol{L}_{\boldsymbol{W}\boldsymbol{W}^\top} \right\}^{-1} \boldsymbol{W}. \tag{1}$$

Here,

$$\gamma \equiv \frac{1}{N_V} \left(\sqrt{\frac{N_E + N_V - 1}{N_E - 1}} - 1 \right), \tag{2}$$

$\mathbf{1}_{\boldsymbol{N_E} \times \boldsymbol{N_V}}$ is an $N_E \times N_V$ matrix of ones, $\boldsymbol{L}_{\boldsymbol{C_E}}$ is a lower-triangular matrix obtained from the Cholesky decomposition of $\boldsymbol{C_E}$ (note that $\boldsymbol{C_E} = \boldsymbol{L}_{\boldsymbol{C_E}} \left(\boldsymbol{L}_{\boldsymbol{C_E}} \right)^\top$), $\boldsymbol{C_E}$ is an $N_E \times N_E$ matrix defined by

$$\boldsymbol{C_E} \equiv \frac{N_V}{N_E - 1} \boldsymbol{I}_{\boldsymbol{N_E}} - \gamma^2 N_V \mathbf{1}_{\boldsymbol{N_E} \times \boldsymbol{N_E}}, \tag{3}$$

*$\boldsymbol{I}_{\boldsymbol{N_E}}$ is the $N_E \times N_E$ identity matrix, $\mathbf{1}_{\boldsymbol{N_E} \times \boldsymbol{N_E}}$ is an $N_E \times N_E$ matrix where every element is one, $\boldsymbol{L}_{\boldsymbol{W}\boldsymbol{W}^\top}$ is a lower-triangular matrix obtained from the Cholesky decomposition of $\boldsymbol{W}\boldsymbol{W}^\top$ (note that $\boldsymbol{W}\boldsymbol{W}^\top = \boldsymbol{L}_{\boldsymbol{W}\boldsymbol{W}^\top} \left(\boldsymbol{L}_{\boldsymbol{W}\boldsymbol{W}^\top} \right)^\top$), and **W** is an $N_E \times N_V$ matrix whose (i, j)-th element is defined by...."*