

## Reply to Referee #2

Takahito Mitsui<sup>1</sup>, Shunji Kotsuki<sup>2,3</sup>, Naoya Fujiwara<sup>4</sup>, Astushi Okazaki<sup>2,3</sup>, and Keita Tokuda<sup>1</sup>

<sup>1</sup>Faculty of Health Data Science, Juntendo University, Urayasu, Japan

<sup>2</sup>Institute for Advanced Academic Research, Chiba University, Chiba, Japan

<sup>3</sup>Center for Environmental Remote Sensing, Chiba University, Chiba, Japan

<sup>4</sup>Graduate School of Information Sciences, Tohoku University, Sendai, Japan

**Correspondence:** Takahito Mitsui (takahito321@gmail.com)

Thank you very much for your detailed review of our manuscript and for providing valuable feedback. Below, we have listed your questions and comments (*italicized*), followed by our responses and proposed revisions to the manuscript (shown in **violet**). We believe that these changes improve the clarity and quality of our work.

5 *L115-120 The notation and explanation of the two intervention scenarios presented here are not so clear. In the “intervention-off” scenario, is there any intervention prior to time  $s$ ?*

Thank you for pointing this out. For the “intervention-off” scenario planned in  $s \leq t < s + 6$  h, there can be previously-planned interventions until  $t = s + 6$  h, but any intervention is turned off from that point onward. **We will add this explanation just after Eq. (2), which defines the intervention-off scenario.**

10 *In this scenario I assume  $u$  can be nonzero at different grid points. Is that correct?*

Yes, it is. More precisely, in the one-site intervention scenario, an intervention input  $u_i(t)$  can take the non-zero value  $u$  only at a single site, while in the two-site intervention scenario,  $u_i(t)$  can be non-zero at two different sites, at most.

15 *In the one-site intervention scenario,  $u$  is nonzero at one grid point only but can also be nonzero before time  $s$  and after time  $s+6$  hr. Is  $u_i(t)$  constant for  $t > s + 6$  hr?*

Thank you for pointing this out. A simple answer is yes: the intervention input  $u_i(t)$  is constant for  $t \geq s + 6$  h in a scenario planned in each 6-h cycle. However, we can update the scenario in the following 6-h cycles. Therefore, the actual sequence of  $u_i(t)$  can change in time for  $t \geq s + 6$  h.

20 *Is the intervention forcing  $u$  a constant (e.g., given that the forcing will be applied at grid point  $i$ , is the value of this forcing known a priori or is it something that will be optimized)?*

Yes, the intervention size,  $u$ , is a predetermined constant parameter here. The optimization of  $u$  is omitted in the present study to reduce computational cost. **For clarity, we will add the following sentence after Eq. (3): “where  $u$  is a constant parameter representing the actual intervention size.”** This assumption may be justified by the following consideration:  
25 since human influence is small compared to the dynamics of weather systems, it is reasonable to assume that the intervention

operates at its maximal feasible strength.

*Figure 2: What do the second panels of Figure a) i) and ii) mean? Also in Figure b)iii), ALERT instead of ALEAT.*

30 The first and the second panels in Fig. 2a (i) and (ii) show two different cases. In the second panels of Fig. 2a (i) and (ii), there is no intervention over  $s \leq t < s + 6$  h since it was not selected in the previous 6-h cycle. On the other hand, the first panels show the cases where there is an intervention over  $s \leq t < s + 6$  h. **We will add the following sentence in the caption: “The lower panels in both (a-i) and (a-ii) show the cases where no intervention over  $t \geq s$  was selected in the previous 6-hour cycle.”** Also the typo is corrected. Thank you.

35

*L137 Multi-scenario ensemble forecast and local intervention:*

Thank you: **We have changed from ‘milti’ to ‘Multi’.**

*L145 In the sentence starting with “Other criteria . . . “ I can not clearly follow the difference from the previous criterion.*

40 Sorry for this confusion. **We will replace ‘the expected maximum’ by ‘the ensemble-mean of the maximum value across all sites’.**

*L152 In this part the approach to evaluate sampling errors in the resulting scores is discussed. Why did the authors decide to work with smaller samples instead of applying bootstrap to the whole sample?*

45 We worked with the ten 100-y samples as in Sun et al. (2023). This allows us a clear comparison with that work. **However, in the revised manuscript, we will mention the dependence of the results on sampling methods.**

*L167 “This metric is particularly relevant . . . ” My impression is the opposite: when the intervention is static, scenarios can not be changed, and this metric is not relevant.*

50 Thank you for pointing this out. Indeed, if the intervention is static (i.e., time-invariant), the scenarios cannot be changed. We believe that this apparent contradiction is solved if ‘static’ is removed from the corresponding sentence.

*In the discussion of Figure 4, the method of Sun et al. is compared with the new method. In the first comparison the success rate is much higher for the new method. However, this is done for a larger intervention size and for a shorter forecast window than in Sun et al. Figures 5 and 6 show that under similar intervention energy and forecast length, Sun et al.’s method are closer to the results obtained with the proposed method. A similar comparison is presented in the abstract and in the conclusions; however, it is unclear if the numbers commented on in the abstract correspond to the numbers in this section. If so, the claim of the abstract and the conclusions does not seem to be a clear comparison with Sun et al.’s approach.*

60 Thank you for this comment. As you suggested, our comparison with Sun et al. (2023) was not entirely fair. **In the revised manuscript, we will modify the concluding statement to focus on our own achievement rather than making potentially debatable comparisons:**

(Before) **The success rate of our method is markedly higher than that of Sun et al.'s method**, reaching 94% even when applying interventions at one site per step, ...

(After) **Our method achieves a high success rate**, reaching 94% even when applying interventions at one site per step, ...

65 We also note your observation that our success rate is comparable with that of Sun et al. (2023) under similar intervention energy.

L190: "... necessary is not guaranteed". Can this be assessed from the previous experiment? The distribution of the distance of the optimal interventions with respect to the location of the extreme event can be obtained and analyzed to support this claim.

75 The original statement, "Whether interventions across all 40 sites are feasible or (if feasible) necessary is not guaranteed," was potentially misleading. What we actually intended was that examining all possible intervention combinations is infeasible in real-world operations. **We will revise the statement as follows: "Assuming real-world applications, interventions across all sites may be neither feasible nor necessary."**

75 Figure 7. Panel c describes the number of scenario changes. This metric seems to grow rapidly from 1 intervention-eligible site to 3. However, I wonder what the behavior would be if the distance associated with each change is also taken into account. It would make sense to distinguish between many small changes and few larger changes (also considering that sometimes the change needs to be done in a small time frame).

80 We assume that the reviewer is concerned with spatial distances between consecutive intervention sites. These distances may increase as the number of intervention-eligible sites increases. In the present study, this factor is not included in the cost estimation. **In the revised manuscript, we will mention that increasing the number of intervention-eligible sites may result in higher transportation costs.**

85 L211 complete instead of complate.

Thank you. We correct it.

L214 Figure 11?

Yes, we change it from Fig. 10 to Fig. 11. Thank you.

90 L215: Why is the ensemble size increased in this experiment? I understand that the localization scale has to be adjusted when the observation network is changed; however, increasing the ensemble size is assumed to always lead to a better performance of the filter (particularly at these relatively small ensemble sizes), but always limited by the available computational power.

We increased the number of ensemble size only a little from 10 to 11 in order to have a better performance. However, as you point out, it makes the interpretation of the results less clear. **Thus, in the revised manuscript, we will show also the cases with 10 ensemble members to clearly show the effect of partial observation.**

## References

Sun, Q., Miyoshi, T., and Richard, S.: Control simulation experiments of extreme events with the Lorenz-96 model, *Nonlinear Process. Geophys.*, 30, 117–128, <https://doi.org/10.5194/npg-30-117-2023>, 2023.