List of Responses

Responds to the Anonymous Referee #2's comments:

Special thanks for your good comments which are very useful for us to improve the paper.

1. Response to comment: In order to find viable alternatives for using an adjoint, the authors test a combination of two other search algorithms, "particle swarm optimisation" and "wolf search" on a reduced dimension state space with 50 dimension and test their performance against a reference method called "the ADJ method". However, it does not become clear, whether this reference method is used to solve the same problem, which should give identical results provided that all methods find the global minimum. Also, solving a 50-dimensional problem with 200 (resp. 420; see swarm size from table 1) model integrations at each solver step in 20 to 30 steps (Fig. 2) does not look like a dramatic improvement over conventional methods, and no direct comparison to those is offered.

Response: It is really true as Rreview2 suggested that we should give identical results provided that all methods find the global minimum. And we run the PSO, WSA and ACPW programs 10 times and then compare their results. It is commonly known that all intelligent algorithms are stochastic; that is, even when the input is the same in different trials, the output may be different. Hence, it hard to obtain the same result. But we can use the maximum, minimum and mean objective values as well as the RMSE to evaluate the algorithm. Therefore, we have illustrated this in the Section 4.1.

"To evaluate the advantages of the ACPW algorithm, we run the PSO, WSA and ACPW programs 10 times and then compare the maximum, minimum and mean objective values as well as the RMSE.

4.1 The advantages of the ACPW algorithm

Because the statistical analysis results are similar for the two TCs with the two resolutions, we only describe the analysis of Fitow at a resolution of 60 km. Table 3 presents the maximum objective value, the minimum objective value, the mean objective value and the RMSE of the 10 results.

Table 3: The analysis results of the PSO, WSA and ACPW methods.

Algorithm	Maximum Value	Minimum Value	Mean Value	RMSE
PSO	1034.192573	724.086002	900.7488578	**0.121400896**
WSA	1628.841294	323.7493169	930.9103862	0.431193448

ACPW	**2240.275956**	1243.377921	**1542.505251**	0.216750584

In Table 3, the maximum objective value is gained from the ACPW algorithm, and its mean value is also more than the other two algorithms. However, the RMSE of PSO is the smallest, which shows the best stability.

For additional analysis, we draw a box-plot of the 10 results for the PSO, WSA and ACPW algorithms, as shown in Fig. 3.

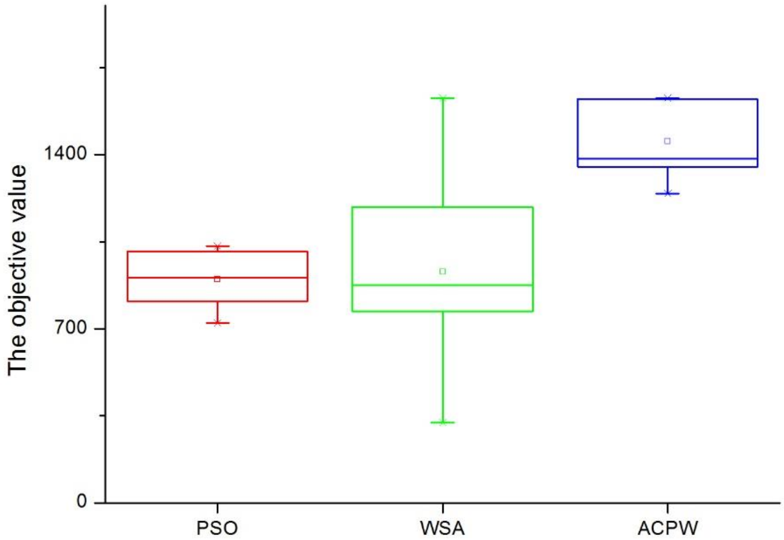

Figure 3: Box-plot of the PSO, WSA and ACPW methods for TC Fitow at 60 km resolution. The red box denotes PSO, the green box is for the WSA, and the blue box shows the results of the ACPW algorithm.

PSO has the narrowest range of values, although the objective values are smaller than the other two algorithms. The WSA has the widest range of values, although the objective values are also smaller than the ACPW algorithm. The ACPW algorithm has the second-best stability, although it has the best objective values. The experiments display the stability of PSO and the exploitation of the WSA. We combine the advantages of them and develop the ACPW algorithm to solve CNOPs. The analysis results demonstrate that the hybrid strategy and cooperation co-evolution is useful and effective."

2. Response to comment: "The ADJ method" is used as a benchmark, but it is ambiguously defined and no attempts on parallelisation are made, not even in, the case of multiple starting points, which supposedly can be parallelised trivially. Also the article leaves the impression that "the ADJ method" is run on the full state space, rather than the 50 dimensional PC space. In summary, the comparisons in terms of computational performance are not convincing.

Response: As Rreview2 suggested that we inserted the reference of the ADJ-method in L5-6, p.3. "Specific details of the ADJ-method can be found in Zhou (2009)."

As Rreview2 mentioned that the multiple starting points can be paralleled, but the time consumption will not be less than using one starting point under the same computer hardware environments. When we analyze the efficiency of the ACPW algorithm in Section 4.5, the ADJ-method using one initial guess field (starting point) is compared with the ACPW algorithm. And the speedup of the ACPW reaches 4.53 and 3.84 for the different resolutions.

"To promote the efficiency of the ACPW algorithm, we parallelize it with MPI technology. The time consumption of each case is nearly the same. Hence, we can use one group of experimental results to elucidate the efficiency of the ACPW algorithm. Because the ADJ-method cannot be parallelized because each input depends on the output of the previous step, its time consumption is not changed. Moreover, because this method generally uses 4~8 initial guess fields to obtain the optimal value, we use one and four initial first guess fields to determine the CNOPs. The time consumptions of the ADJ-method and ACPW algorithm are shown in Table 8.

Table8: The time consumption of the ADJ-method and ACPW algorithm (unit: minutes).

Methods	60 km	120 km
ADJ-method (1)[1]	79.9	12.4
ADJ-method (4)[1]	321.1	49.7
ACPW	**20.8**	**2.74**

1. ADJ-method (1) means using 1 initial guess field and ADJ-method (4) means using 4 initial guess fields.

At 120 km resolution, the time consumptions of the ADJ-method using 1 and 4 initial guess fields are 12.4 minutes and 49.7 minutes, respectively. At 60 km resolution, the time consumptions are 79.9 minutes and 321.1 minutes, respectively. Unlike the ADJ-method, the ACPW algorithm can be parallelized. When using 22 cores, the ACPW method requires much less time, i.e., 2.74 minutes at 120 km resolution and 20.8 minutes at 60 km resolution. Obviously, the ACPW has higher efficiency. Compared to the ADJ-method (1), the speedup reaches 4.53 and 3.84 for the different resolutions. Compared to the ADJ method (4), the speedup reaches 18.14 and 15.44. Although the different initial guess fields are calculated in parallel, the time consumption must be more than for the ADJ-method (1); the ACPW algorithm is also faster than the ADJ-method.
"

In addition, when the ACPW algorithm calculated the objective value, we use the nonlinear model on the full state space, only update the individual with the 50 dimensions.

3. Response to comment: The experiments with the reduced amplitude CNOPs are hard to follow. I had difficulties to understand section 4.3., which is the motivation for the verification and forecast experiments.

Response: For the Section 4.3, we want to investigate the validity of the sensitive regions identified using CNOPs, and we have two assumptions:

"When adding adaptive observations in sensitive regions, the surrounding environment is idealized, and the improvements from adding observations reduces the original errors by a factor of 0.5.

The obtained CNOPs can be seen as the optimal initial perturbations. Once we reduce them in the

sensitive regions, the benefits are highest.
”

Therefore, we design two groups of idealized experiments. CNOPs are optimal initial perturbations having the maximum nonlinear evolutions at the forecast time. Under these assumptions, by reducing the CNOPs to W×CNOPs and inserting them into the initial states we can investigate how the reductions in the CNOPs influence TC forecast skill.

4. Response to comment: In the presentation of the resulting CNOPs, the surface pressure patterns are neither shown nor discussed. No information on the vertical structure of the CNOPs is given. Moisture, an important energy source for tropical cyclones, is not included in the state vector and no justification for this omission is given. The authors do not address the the role of the fixed PC space dimension (and basis?) when comparing patterns at different resolutions. No information on how the excitation of numerical modes is avoided, both in the computation of the CNOPs and when making perturbed forecasts.

Response:

As Review2 mentioned that we did not discuss the surface pressure patterns and the vertical structure of the CNOPs, because the purpose of this paper is to identify the adaptive observation sensitive areas, we follow the study of Dr, Zhou that the total dry energy have higher benefits than other strategies (Zhou and Zhang, 2014). Therefore, the information of the surface pressure patterns and the vertical structure of the CNOPs are contained in the total dry energy. In addition, Dr. Zhou has proved that the sensitive regions gained by the dry energy have higher benefits than those obtained from the moist energy (Zhou, 2009). In this paper, we only considered the total dry energy.

For the question that “The authors do not address the role of the fixed PC space dimension (and basis?) when comparing patterns at different resolutions”, the numbers of PCs in this paper are determined by the many experiments, and the analysis of the different numbers are plotted in our previous studies.

Finally, in this paper, we also use the nonlinear model, but avoid using the adjoint model to calculate the gradient.

5. Response to comment: Many formulations in the abstract and the article are confusing on a language level, to name only a few: "...suggest that the use of an ocean coupled model needs to be conscious,..." (page 2, line 13), "the mutual affection of binary typhoons" (page 2 line 14), "[wolf search] ... takes long consuming time." (page 4, line 6). Language editing is encouraged.

Response: As Review2 suggested that we have improved the quality of our manuscript by American Journal Experts editing service and tracked the changes using revisions in the manuscript ‘Revised Manuscript with Track Changes’.

6. Response to comment: What is the update for ui if neither of the two conditions is satisfied?

Response: We are very sorry about errors in this paper and have corrected them in L2-9 Page 5. “

$$\begin{cases} u_i^{k+1} = u_i^k + \theta \cdot r \cdot rand(\ \ ) & Prey \\ u_i^{k+1} = u_i^k + \theta \cdot s \cdot escape(\ \ ) & Escape \end{cases} \tag{6}$$

where the superscript k or $\mathrm{k}+1$ is also the iterative step, θ is the velocity, r is the local optimizing radius, which is smaller than the global constraint radius δ, *rand( )* is the random function, whose mean value is distributed in [-1,1], *escape( )* is the function for calculating a random position, which is 3 times larger than r, and s is the step size of the updating individual.

As described in Eq. (6), the wolf has two behaviours, i.e., prey and escape. The prey behaviour uses the first sub-formula, and the second one is for the escape function, which happens in every iteration when the condition $p > p_a$ is satisfied, where p is a random number in [0,1], and p_a is the probability of individual escaping from the current position. ”

7. Response to comment: page 8, formula 10: Is this using the same energy norm as formula 10? If not, how are the different variables combined?

Response: formula (10) is used to calculate the similarity between the CNOPs, every CNOP has the same components, so we did not use the norm. Actually, the formula is for solving the Cosine similarity.