Dear referee 2,
We thank you for the attention that you paid to this review and your helpful comments and suggestions.

1. Experimental design: A typical grid ratio in WRF is 3:1, or any odd integer ratio for computational efficiency and accuracy (Skamarock et al. 2008). The nesting ratios of the third, fourth, fifth domains in this study are not exactly, but close to 3. Some words are needed for specific consideration.

Our answer:

We confirm that in the present study we used a 3:1 typical grid ratio in WRF experimental design. The rounded values of resolution 830 m, 280 m, 90 m and 30 m correspond respectively to the accurate values of grid scales 833.333 m, 277.778 m, 92.592 m, 30.864 m.

To clarify this point in "Method" section 3,

- Line 121, the sentence *"These six nested domains have a respective resolution of 7.5 km, 2.5 km, 830 m, 280 m, 90 m and 30 m (Fig. 2a and b)."* will be replaced by:

These six nested domains have a respective resolution of 7.5 km, 2.5 km, 833.333 m (approx. 830 m), 277.778 m (approx. 280 m), 92.592 m (approx. 90 m), and 30.864 m (approx. 30 m) (Fig. 2a and b). For simplicity, in the following, we will use the approximate values (i.e. 830 m, 280 m, 90 m, 30 m) to describe the grid scales of the four innermost domains.

- In Table 1, we will add the following line at the top of the table:

Innermost domain scale (m): 833.333 277.778 92.592 30.864

And the second line corresponding to the approximate values will be:

Approx. innermost domain scale (m): 830 280 90 30

2. More detailed description is needed about the experiments. For example, when are the inner domains added in the three experiments? When are the topography and land-use removed in the NOIS and NOTP experiments? How many vertical layers are used in these experiments below 1-km altitude?

-When are the inner domains added in the three experiments?

Our answer:

The numerical experiments are described in Table 1 with the innermost domain scale corresponding at each experiment.

To improve the understanding, the sentence *"To examine resolution effects avoiding two-way child domain perturbations, all presented model outputs correspond to the innermost domain of the numerical experiments."* (Line 155) will be moved to the end of the section (i.e. Line 161).

We will also add the following sentence:

For example, while the REAL280 experiment includes four nested domains with the innermost domain resolution of 280 m, the REAL090 experiment includes five nested domains with the innermost domain resolution of 90 m.

-When are the topography and land-use removed in the NOIS and NOTP experiments?

Our answer:

Firstly, we fixed typing errors in the experiment names at the end of Table 1. The corrected two last lines of Table 1 are:

Innermost domain scale (m)	833.333	277.778	92.592	30.864
Approx. innermost domain scale (m)	830	280	90	30
Number of pts (x*y)	718*520	202*202	133*103 St. Barth	295*208 St. Barth
			238*172 St. Martin	478*412 St. Martin
Timestep (s)	2.5	0.833	0.278	0.093
Turbulence scheme	YSU	TKE or NBA	NBA	NBA
Real terrain Exp.	REAL830	REAL280	REAL090	REAL030
No island Exp.	-	NOIS280	NOIS090	**NOIS030**
No topography Exp.	-	-	**NOTP090**	NOTP030

While only the topography is removed in the NOTP experiment, both topography and land-use are removed in NOIS experiment.

For a better understanding, we can rewrite the sentence Line 155 (Method section):

"Three experiment types are run: REAL, NOIS and NOTP, corresponding respectively to real island terrain (i.e. with topography and land-use), removed island terrain (i.e. without topography and land-use), and removed topography."

This way:

"Three experiment types are run: REAL, NOIS and NOTP, corresponding respectively to real island terrain (i.e. with real topography and real land-use), removed island terrain (i.e. with topography set to constant zero value and land-use set to constant water category), and removed topography (i.e. with topography set to constant zero value and real land-use)."

-How many vertical layers are used in these experiments below 1-km altitude?

Our answer:

Line 125, in the Method Section,

After "The model has 99 terrain following vertical levels in a logarithmic resolution that is finer in lower levels, and the top is at 30 hPa (Jury et al., 2019)."

we will add this sentence: "Near surface, below 1-km altitude, 16 vertical levels are used with the first level at 13 m above ground level."

3. How many tornado-scale vortices (TSVs) are found in the NOIS90-NBA and NOIS30-NBA experiments? Since the previous studies (Wu et al. 2018, 2019) mentioned that TSVs are prevalent inside the TC eyewall. Has the structure of the simulated TSV been carefully examined? It seems that the scale is much smaller than those in Wu et al. (2019).

Our answer:

Because of the very small area of the 30-m scale studied fixed domain (i.e. 9 km per 6.5 km), we did not particularly analyse the number of tornado-scale vortices found in the NOIS90-NBA and NOIS30-

NBA experiments. Moreover, the aim of the present numerical analysis was not to carefully examine the structure of the simulated TSVs. We agree that the scale of the TSV seems much smaller than those in Wu et al. (2019). We think that these differences may be caused by: the stronger TC studied here (cat 5 Hurricane Irma with maximum sustainable winds of 80 m s^{-1}), the finer vertical and horizontal grid, the use here of the Nonlinear Backscatter and Anisotropy (NBA) SFS stress model instead of the 1.5-order Turbulence Kinetic Energy (TKE) linear eddy-viscosity SFS stress model…

However this kind of structure with multiple very small-scale vortices (diameter lower than 500 m) have been already observed in a violent tornado and defined as a structure of multiple subtornadic-scale vortices by Wurman (2002). In our case, the 30-m scale is necessary to simulate the multiple subtornadic-scale vortices structure leading to the extreme peak gust at 08:57 UTC (Fig. 5 and 6).

Wurman, J.: The Multiple-Vortex Structure of a Tornado, Wea. Forecasting, 17, 473–505, https://doi.org/10.1175/1520-0434(2002)017<0473:TMVSOA>2.0.CO;2, 2002.

To highlight that the 30-m grid spacing is necessary to simulate an intense structure of multiple subtornadic-scale vortices and to show the differences between the tornado-scale vortices found in the NOIS090-NBA and NOIS030-NBA experiments, we can add the following vorticity figure in the "Effects of resolution on gusts and small-scale vortices" section.

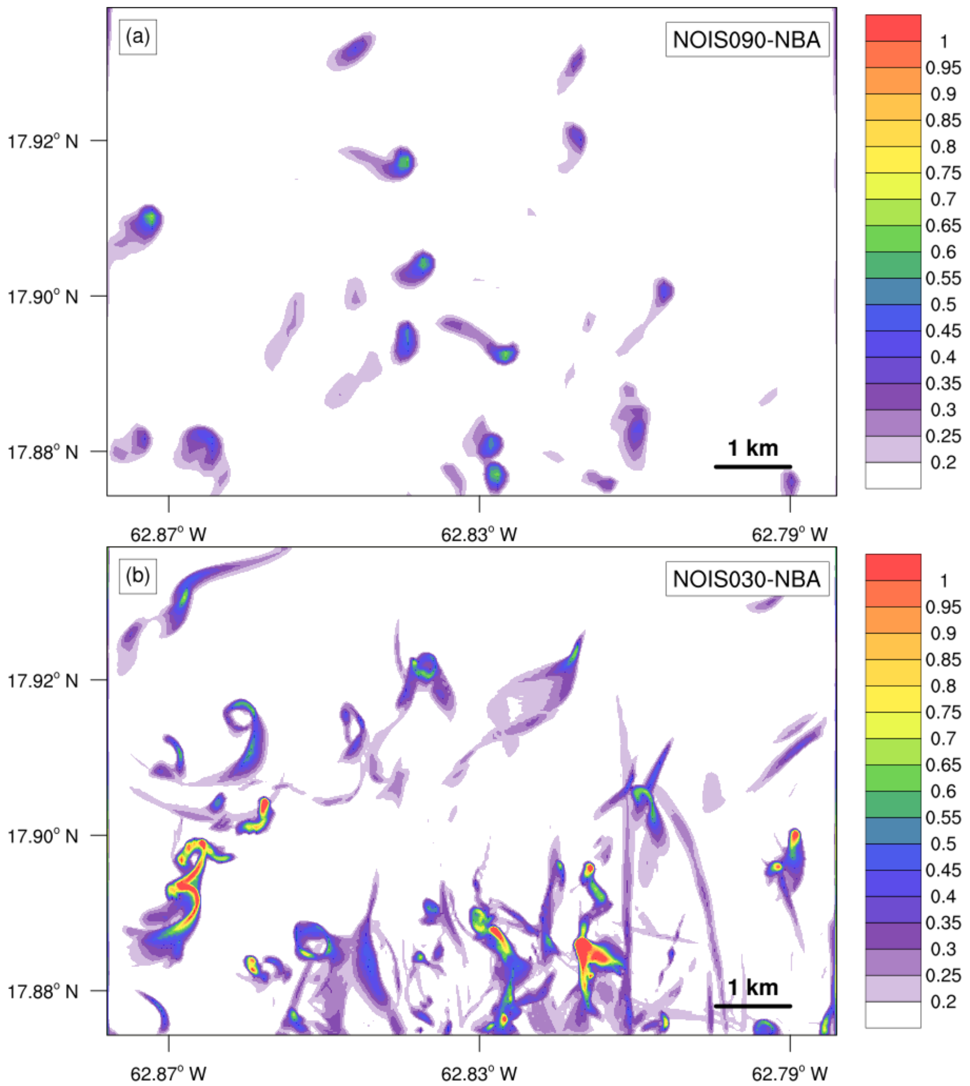

Figure: Maximum vertical vorticity (s^{-1}) over the 6 hours of simulation (history output interval of 1 min) and the vertical column below 600-m level: 90 m resolution (a) and 30 m resolution (b).

As shown in this Fig. while the 90-m scale model well reproduces tornado-scale vortices with a maximum vertical vorticity of 0.68 s^{-1} over the 360 minutes of simulation and the vertical column below 600-m level, the 30-m scale is necessary to simulate structures of multiple subtornadic-scale vortices link with maximum vertical vorticity above 1 s^{-1}.

Based on Wurman (2002), we think that the following sentences need to be modified as below.

In Abstract, Line 19:

"The results pointed out that the 30-m scale seems necessary to simulate **an intense structure of multiple 400-m scale vortices (i.e. subtornadic-scale vortices)** leading to extreme peak gusts like 132 m s^{-1} over sea"

In the "Conclusion" section, Line 292:

« Moreover while the 280-m resolution and the 90-m resolution allowed to reproduce medium kilometer-scale vortices and the associated surface instantaneous gust of 110 m s^{-1} with location errors, the 30-m resolution seems necessary to simulate **an intense structure of multiple 400-m scale vortices (i.e. subtornadic-scale vortices)** leading to extreme peak gusts like 132 m s^{-1} in open-water conditions."

4. Why are the upstream surface winds over the sea in the REAL030 experiment generally higher than those in the NOIS030 experiment in Fig. 8a?

Our answer:

In Fig. 8a, the upstream surface winds over the sea from the REAL030 outputs are not compared with those from the NOIS030 experiment, but with the hilltop surface winds from the REAL030 outputs.

5. Physical explanation is strongly suggested to understand the enhancing effect of topography.

Our answer: We agree that this point needs to be clarified.

Firstly, in the "Study area" section, we will add the following sentences Line 105:

"As described by Done et al. (2020), the high Froude number induces the flow to pass directly over the hill crest. Under mass continuity, this flow is accelerated at the hilltop due to the local constriction of the air column. These orographic wind speed-up effects have been found during Hurricane Fabian (2003) over the low hill crest of Bermuda (i.e., 86 m) (Miller et al., 2013)."

Then in the "Effects of Saint Barthélemy island terrain on gusts" section, the following figure will be added after Fig. 8.

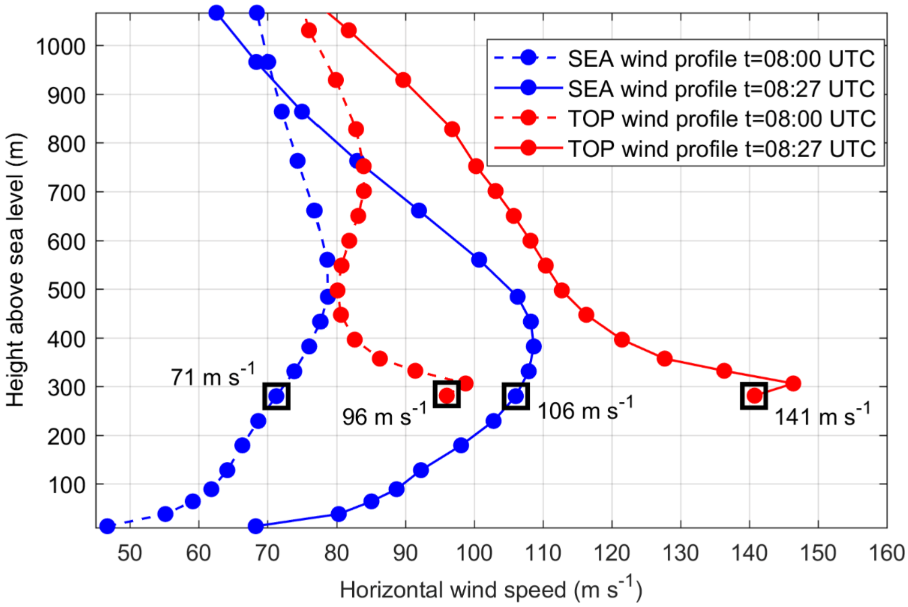

Figure 9: Vertical profile of the REAL030 instantaneous horizontal wind speed (m s^{-1}) at 08:00 UTC and at the peak gust time 08:27 UTC: comparison between the upstream winds over sea (SEA) and the orographic winds over the mountain top (TOP).

The sentence at the Line 229 “The TOP/SEA gust enhancement factor reaches 1.84 at this time” will be replaced by the following analysis:

For a better understanding of the TOP/SEA wind enhancement factor, the vertical profile of the wind speed was examined at SEA and TOP locations before and during the peak gust time (i.e. 08:00 and 08:27 UTC). The study of the wind speed at 280 m above sea level (i.e. the height of the surface winds at the hilltop of Saint Barthélemy) highlights the fact that the same level winds flowing upstream over the sea are accelerated at the hilltop (Fig. 9). This local wind speed-up factor induced by the air column constriction at the hilltop has closed values at the two times: 1.35 and 1.33, respectively at 08:00 and 08:27 UTC.

6. Why are some characteristics of discontinuities in the maximum instantaneous gusts in Fig. 9a (shading)?

Our answer:

Saint Martin island is located further North than the path of the most intense eyewall quadrant. These sparse characteristics of discontinuities in the maximum surface instantaneous gusts (over the six hours of simulation) may be induced by the weaker hurricane boundary layer vortices and their destabilization by this mountainous island.