

1 Response to RC3

2

3 ➤ General comments

4 This study proposes a framework for assessing the potential risks of climate change
5 induced by river discharge and storm surges using multiple numerical models and
6 the Pseudo Global Warming (PGW) method. The topic is timely, as the results can
7 be applied to evaluate compound floodings near coastal areas and provide insights
8 into disaster mitigation. However, the manuscript has several drawbacks that the
9 authors should address. Please refer to the following comments for further revision.
10 Also, I recommend using a professional editorial service to improve the writing
11 quality, as there are several grammatical errors and typos.

12 **Response:**

13 We thank the reviewer for the positive assessment of the study's relevance and potential
14 applications to compound flood risk and disaster mitigation. We appreciate the
15 constructive comments provided.

16 In response to the reviewer's suggestions, we have carefully revised the manuscript to
17 address the issues raised in the detailed comments below. We have also thoroughly
18 edited the manuscript to correct grammatical errors and typographical issues and to
19 improve overall clarity and readability.

20 We believe that these revisions have substantially improved the quality of the
21 manuscript.

22

23 Specific comments

24

25 Line 106: The computational settings across the three models appear to differ
26 significantly. For instance, the NHRCM model utilizes only the cumulus scheme,
27 whereas the CReSS model utilizes a slab ocean model without applying spectral
28 nudging. These differences may over-emphasize the variability of the CPMs, potentially
29 undermining the robustness of future projections. The authors should explain and
30 justify the rationale behind these settings in the main text.

31 **Response:**

32 We would like to clarify that the cumulus or convective parameterization scheme is
33 applied only in the coarser-resolution NHRCM simulations (20 km and 5 km;
34 NHRCM20 and NHRCM05), as summarized in Table S1. The 2 km resolution NHRCM
35 simulation (NHRCM02), which is used for the main analysis and intercomparison with

36 the 2 km WRF and CReSS simulations, does not employ a cumulus parameterization
37 scheme (all are CPMs). We have clarified this point in the revised manuscript in lines
38 119-120 and 128-130 to avoid misunderstanding.

39 More generally, each CPM configuration was designed to follow its standard high-
40 resolution setup, with physical parameterizations chosen according to model resolution
41 to maximize the reproducibility of the meteorological fields. Our intention was not to
42 harmonize parameterization schemes across models, but rather to evaluate structural
43 uncertainty arising from independently developed CPM systems operating under their
44 optimized configurations.

45 We have added text in the revised manuscript (lines 347–351) to clarify that differences
46 in model configurations contribute to broadening the range of model behaviour
47 captured by the multi-model framework, thereby enhancing its ability to represent
48 uncertainty and increasing its overall utility.

49

50 Line 108: It is necessary to describe the criteria for selecting the 5 cases from the 27
51 ensembles. I also recommend indicating the selected TCs in Fig. 3a to improve clarity.

52 **Response:**

53 We acknowledge that the selection criteria for the five cases from the original 27
54 ensemble simulations were not clearly described in the original manuscript.

55 In the revised manuscript in line 121–127, we have clarified that the five cases were
56 selected based on their ability to realistically reproduce the observed landfall location of
57 Typhoon Hagibis. Because the primary objective of this study is to assess compound
58 flooding risks in the Tokyo Bay and surrounding coastal regions, accurate reproduction
59 of the landfall position is essential to ensure a physically meaningful hazard assessment.
60 The five selected cases are not explicitly shown in Fig. 3a to avoid making the figure
61 overly complex. They are instead presented in Fig. S1.

62

63 Line 114: The exact timing used to calculate the differences in SST and air temperature
64 should be described. For instance, what specific period does “future” refer to? In
65 addition, d4PDF does not include results for 2023, which seems inconsistent with the
66 statement in Line 270. Please correct this or add an explanation if necessary.

67 **Response:**

68 We thank the reviewer for pointing out the need for clarification regarding the reference
69 period. While the publicly available d4PDF historical experiment covers 1951–2011, we
70 used an extended dataset provided directly by the project collaborators, which includes

71 simulations up to 2017. The present-day climatology was therefore defined as the 2008–
72 2017 mean. Although Typhoon Hagibis occurred in 2019, the selected reference period
73 is sufficiently close in time to represent the contemporary climate state. Moreover,
74 because the PGW method evaluates thermodynamic changes relative to a baseline
75 climatology, small differences in the reference year do not materially affect the
76 assessment of warming-induced impacts.

77 We have clarified this point in the revised manuscript in lines 151–155.

78

79 Section 2.4.2: The bulk formula used to calculate the sea drag coefficient and the
80 treatment of the wind speed threshold should be described. Additionally, a description
81 of the modeled bottom stress is essential for storm surge simulations and should be
82 included.

83 **Response:**

84 The description of the wind drag formulation, wind speed threshold treatment, and
85 bottom stress parameterization has been added in in line 168–192.

86

87 Line 164: The author should explain whether the lack of data pertains to all selected
88 analysis sites, or if there are other contributing factors. Moreover, given that river
89 discharge is generally measured at multiple stations along the target rivers, it is
90 necessary to clarify whether these locations also experienced data gaps or if the data
91 quality was too poor to be utilized in this study.

92 **Response:**

93 We appreciate this comment. A large number of observation stations (753 sites) are
94 available from the Ministry of Land, Infrastructure, Transport and Tourism (MLIT).
95 Among these, we selected stations that correspond to locations where the river models
96 output peak discharge, in order to ensure consistency between the observations and
97 model simulations for validation purposes. Data were therefore provided only for these
98 selected sites. These stations are commonly used for validation purposes and have been
99 well examined in previous studies; therefore, they are considered suitable for evaluating
100 the performance of the present model.

101

102 Lines 174-175: The spread and Fig. 3a indicate that TC tracks from some ensemble
103 members were not closely aligned with observations near Tokyo Bay. The phrase “model
104 results closely follow” should be revised to accurately reflect the simulated results.

105 **Response:**

106 We agree that the previous wording was overly strong and have revised the text
107 accordingly. The manuscript in line 223-225 now states that the simulated tracks are
108 broadly consistent with the best track while acknowledging the ensemble spread near
109 Tokyo Bay. We further clarify that this spread reflects the initial-condition uncertainty
110 inherited from the Global EPS, and that spectral nudging in NHRCM constrains the
111 simulated tracks to remain close to those of the Global EPS, whose track reproduction
112 skill is considered high.

113

114 Line 184: Even if the translation speeds of the TCs differ among ensemble members,
115 the authors should verify the accuracy of the time series for precipitations, sea level
116 pressure, wind speed and wind direction, as these directly affect the temporal evolution
117 of river discharges and storm surges. At least, this comparison should be conducted for
118 the five selected TCs.

119 **Response:**

120 We appreciate this comment. As described in the manuscript, the comparison of
121 typhoon central pressure is conducted at the same latitude rather than at identical time
122 points, because the northward translation speed differs among ensemble members and
123 from the observations. This spatially aligned comparison avoids artificial timing
124 discrepancies due to translation speed differences.

125 We have clarified this methodological choice in the revised manuscript to better explain
126 how the temporal evolution of key meteorological variables is evaluated in line 232-235.

127

128 Figure 4: The simulated central pressures suggest that NHRCM (CReSS) tend to
129 simulate the highest (lowest) wind speeds according to the gradient wind balance;
130 however, the simulated maximum wind speeds show the opposite trend. This is counter-
131 intuitive. The authors should elaborate on the possible reason for this gap.

132 **Response:**

133 We agree that the relationship between central pressure and maximum wind speed
134 shown in Fig. 4 may appear counter-intuitive when considering the gradient-wind
135 balance. In the revised manuscript in line 253–258, we have added an explanation
136 clarifying that Fig. 4b shows the *maximum surface wind speed*, which is strongly
137 influenced by boundary-layer and surface-layer parameterizations in addition to the
138 storm structure (e.g., the radius of maximum wind). Therefore, the ranking of
139 maximum surface wind speed among the CPMs does not necessarily correspond to the

140 ranking in central pressure. This clarification has been added to better explain the
141 differences among the models.

142

143 Lines 196-200: There appear to be no results in Fig. 4 that correspond to the statistics
144 (e.g., 10.9 hPa in line 196) mentioned in the text. Please revise the figure or the text to
145 address this inconsistency.

146 **Response:**

147 The statistics reported in the text (e.g., the 10.9 hPa decrease in central pressure)
148 represent the ensemble-mean differences averaged across the three CPMs. In Fig. 4,
149 individual ensemble members are shown by crosses, while filled circles indicate the
150 ensemble mean calculated from the five initial-condition members for each CPM. To
151 avoid confusion, we have revised both the figure caption and the corresponding text to
152 clarify how these statistics are derived.

153

154 Figure 7-1: I recommend showing the storm surge waveform with the highest accuracy
155 (i.e., the lowest RMSE) among the ensemble members rather than the mean values.
156 Averaging storm surges may obscure local peaks due to phase differences. Indeed, the
157 mean lines do not capture the seiche observed around 10-12-19, which is typical for
158 Tokyo Bay.

159 **Response:**

160 We agree that averaging storm surge time series may obscure important features due to
161 phase differences among ensemble members. Following the reviewer's suggestion, we
162 revised Fig. 7 and S5 to show the storm surge waveform from the ensemble member
163 that produces the maximum storm surge level in each experiment (present and future),
164 which are shown as solid blue and red lines, respectively.

165 In the present-climate simulations, this case also shows the highest correlation with the
166 observed tide level, indicating that it provides the best representation of the observed
167 storm surge evolution, including the seiche signal around 12 October 19:00 in Tokyo
168 Bay. The figure and corresponding text have been revised accordingly in line 295-302
169

170 Figure 7-2: The mean lines indicate that the storm surge models underestimated the
171 observed tidal residuals despite the tendency to overestimate TC intensities in the
172 CPMs. This is also a counter-intuitive result. Please investigate and elaborate on the
173 reason for this.

174 **Response:**

175 The ensemble-mean storm surge was lower than the observed tidal residual because the
176 average included ensemble members that produced relatively small surge levels. As
177 noted in our previous response, averaging storm surge time series can obscure local
178 peaks due to phase differences among ensemble members.

179 To avoid this issue, we revised the figure to show the storm surge waveform from the
180 ensemble member that produces the **maximum surge level** in each experiment, instead
181 of the ensemble mean. The corresponding discussion of the ensemble mean has also
182 been removed from the manuscript.

183

184 Figure 8: The y-axis should show the ratio rather than the maximum surge levels, as the
185 caption indicates that the same definition used in Fig. 6 applies to both the bars and
186 error bars.

187 **Response:**

188 We appreciate the reviewer's comment. In the previous version, the caption could cause
189 confusion regarding the definition used in Fig. 8. The figure shows the maximum storm
190 surge levels, not the ratios used in Fig. 6. Therefore, we revised the caption to clarify the
191 definition and avoid inconsistency with Fig. 6.

192

193 Technical comments

194

195 Line 38: Avoid using the word "like," as it makes the research targets (e.g., typhoons
196 and heavy precipitation) sound ambiguous.

197 **Response:**

198 We agree with the reviewer's suggestion and replaced the word "like" with "such as" to
199 avoid ambiguity when describing the research targets.

200

201 Line 40: Please provide a more specific description of "inner variability." Does this refer
202 to uncertainty in the meso-scale system arising from natural variability rather than
203 model differences?

204 **Response:**

205 During the language revision of the manuscript, the structure and placement of several
206 sentences were modified. However, the paragraph was revised to clearly convey the
207 original intention. Specifically, we replaced the term "*inner variability*" with "*natural*
208 *variability*".

209

210 Lines 43-44: The phrase “SOLA” in the references is unnecessary and should be
211 deleted.

212 **Response:**

213 We agree with the reviewer’s suggestion and removed the unnecessary phrase “SOLA”
214 from the reference list.

215

216 Line 53: Rephrasing “integration” to “forecasting” would improve the clarity of the
217 paper.

218 **Response:**

219 Following the reviewer’s suggestion, the sentence has been revised. The expression has
220 been changed to “These results highlight the importance of forecast initialization.”

221

222 Line 70: The phrase “natural variability” is inappropriate for weekly ensemble
223 simulations, as it typically refers to interannual to decadal scales.

224 **Response:**

225 We thank the reviewer for pointing this out. The description “weekly ensemble” was
226 inappropriate and was due to our mistake. The simulations used in this study are based
227 on the global Ensemble Prediction System (EPS). We have corrected this expression in
228 the revised manuscript.

229 In addition, the phrase “natural variability” was removed to avoid misunderstanding,
230 and the variability is now described as initial-condition uncertainty inherited from the
231 Global EPS (e.g., line 226-227).

232

233 Line 71: The meaning of “Both trials” is unclear. Please clarify.

234 **Response:**

235 We appreciate the reviewer’s comment. The sentence has been revised to clarify the
236 meaning of “both trials,” and the description has been rewritten for clarity (line 78–82).

237

238 Line 78: The phrase “wave height” usually refers to the amplitude of wind waves or
239 swells, not storm surges. Please correct the terminology accordingly.

240 **Response:**

241 We appreciate the reviewer’s comment. The phrase “wave height” has been replaced
242 with “storm surge height” in the revised manuscript.

243

244 Line 102: I recommend revising the description of the DDS procedure for the GCM.
245 The current text does not clearly explain the sequential downscaling steps through
246 domains with different grid sizes.

247 **Response:**

248 We appreciate the reviewer's comment. The DDS procedure has been described in
249 more detail in Sections 2.1 and 2.2. Please refer to this section for the sequential
250 downscaling steps and the case selection step between the 5 km and 2 km simulations.
251

252 Line 104: The meaning of "it" is unclear, making the sentence difficult to understand.
253 Please provide a more detailed description of the DDS procedure.

254 **Response:**

255 The description in Sections 2.1 and 2.2 has been revised, and the ambiguous use of "it"
256 has been removed.

257

258 Line 129: It is preferable to state the spatial resolution of CaMa-Flood in the main text.

259 **Response:**

260 The spatial resolution of CaMa-Flood ($1/60^\circ$) has now been explicitly described in the
261 main text (line 172).

262

263 Line 237: Please add the name of the tide station to the manuscript.

264 **Response:**

265 The name of the tide gauge station (Tokyo) has been added to the manuscript (Line
266 294).

267

268 Figure 3: Please add "(a)" immediately after the figure number.

269 **Response:**

270 Following the reviewer's suggestion, "(a)" has been added immediately after the figure
271 number.

272

273 Table S2: Delete the unnecessary question mark in the table. Additionally, please cite
274 the reference related to the J-FlwDir data.

275 **Response:**

276 The unnecessary question mark in the table has been removed. In addition, the
277 reference related to the J-FlwDir dataset has been added to the manuscript.

278