We are grateful to the reviewers for their insights and believe their comments have substantially improved our manuscript. We address all of their comments below, point-by-point, in blue. The major revisions include:

- Revised introduction to improve the overall narrative and to clarify the objectives of the study.
- Updated description of the CORDEX-SEA ensemble, including names and clarification of the simulations available at the time of analysis (as of September 2025) and the inclusion of newly available simulations produced as part of the ongoing modelling effort (available as of July 2026 and not included in this study).
- Explanation of the characteristics of the seasonal (precipitation and temperature) cycle over the Maritime Continent
- Discussion of a potential hypothesis on the "pair model biases" in daily maximum temperature and precipitation
- Updated and additional explanation on dendrograms from cluster analyses

We trust that our changes to the manuscript will satisfy the reviewers and the Editor.

Response to Reviewer 1's comment

1. The introduction is comprehensive and well supported by relevant literature. However, the overall narrative could be clearer. The progression from CORDEX to CMIP5 and CMIP6 is somewhat non-linear. The distinction between background, previous studies, and the specific contribution of this study is not always clear. As a result, the main focus of the paper, standardized evaluation and model selection for future city-scale downscaling, is not sufficiently emphasized. A clearer structure would improve readability and highlight the key contribution.

We thank the reviewer for this insightful comment. We agree that the chronological progression and ultimate focus of the study could be framed more clearly.

In accordance with the reviewer's recommendations, we have reorganised and rewritten the Introduction section to ensure a logical, linear structure that clearly separates global background information, previous regional studies, and our novel contributions. The specific modifications are detailed below:

- Linear progression: We have restructured the narrative timeline to prevent overlapping transitions. The text now introduces the initial global CORDEX framework and then sequentially moves to the historical CMIP5-driven CORDEX-SEA experiments, contextualizing their performance limitations in the literature. After that, we introduce the transition into the next-generation CMIP6/AR6 models and the updated experimental designs.

- Explicit demand for next-generation projections with a cross-model evaluation: To bridge previous studies with our current work, we added text that explicitly highlights systematic wet biases and inter-model variability in legacy CMIP5 runs, as well as the demand for a new generation of CMIP6-downscaled CORDEX simulations.

Please refer to Section 1 (L57-104) in the revised manuscript to view these tracked changes.

"Since its launch in 2009, CORDEX has facilitated the generation of regional climate simulations across 14 continental-scale domains, moving in parallel with the development of the Coupled Model Intercomparison Project (CMIP) phases (e.g., CORDEX-CMIP5). To provide a more consistent modelling framework across regions, the CORDEX Coordinated Output for Regional Evaluation (CORDEX-CORE) further delivered a standardised set of projections for all CORDEX domains using three global climate models (GCMs) from CMIP5 (Giorgi et al., 2021).

Within this global framework, Southeast Asia (SEA), a home to more than 700 million people, is one of the key domains (Fig. 1). The Southeast Asia Regional Climate Downscaling (SEACLID)/CORDEX-SEA project, launched in November 2013, represents the most comprehensive regional climate modelling initiative for this region, aiming to provide high-resolution climate information to support mitigation and adaptation planning. In its first phase, fourteen dynamical downscaling experiments with a grid spacing of 25 km were conducted over the SEA domain (89.5°E–146.5°E, 14.8°S–27°N, Fig. 1) through a global collaboration among seven modelling centres (Tangang et al., 2020). Subsequently, the WCRP's CORDEX-CORE initiative contributed three additional CMIP5-driven simulations for the SEA domain. Evaluation of CMIP5-driven CORDEX-SEA simulations indicated that while their multi-model ensemble (MME) generally improves representation of climatological precipitation compared to their driving GCMs, significant systematic wet biases remain, particularly within specific regional climate models (Tangang et al., 2020). However, CMIP5-forced regional climate models (RCMs) exhibit larger inter-model variability than their driving counterparts. Furthermore, Nguyen et al. (2022) demonstrated that these RCMs are not necessarily closer to the observed precipitation distribution than their driving GCMs. This highlights the ongoing challenge of regional climate modelling with a demand for a new generation of regional climate projections that provide more reliable and actionable data for climate risk assessments.

Building on these legacy developments, CMIP Phase 6 introduced a new generation of GCMs integrated with the Shared Socioeconomic Pathways (SSPs) (O'Neill et al., 2016), with their key results presented in the Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6) (IPCC, 2021). These advances enabled more comprehensive scenario-based climate projections. In response, the CORDEX community released an updated experimental design tailored for the dynamical downscaling of CMIP6 models (CORDEX, 2021), and modelling groups worldwide have since contributed simulations to provide well-grounded climate information for regional climate resilience and adaptation planning.

Following the CMIP6 CORDEX guidance, several modelling centres from SEA, Hong Kong, Australia, and the United Kingdom have recently initiated new downscaling experiments using CMIP6 GCMs for an expanded CORDEX-SEA domain (17.3°S–31.9°N, 86.5°E–149.3°E; Fig. 1), which now also fully spans Myanmar. This coordinated effort establishes a valuable foundation for generating updated and improved regional climate projections for impact and adaptation studies (Table 1). It is worthy to note that a fundamental requirement for the reliable use of dynamically downscaled climate projections is that a model can adequately simulate the observed climate, including the extreme events that threaten vulnerable natural and socioeconomic systems."

- Sharpened study focus: We have isolated our specific objectives into an independent, clear paragraph at the end of the section. This ensures that our primary focus (e.g., utilizing a standardized benchmarking methodology to isolate a "fit-for-purpose" model selection tailored for kilometer-scale urban climate downscaling under the CARE for SEA Megacities project) is properly emphasized as the core contribution of this paper.

Please refer to Section 1 (L143-162) in the revised manuscript to view these tracked changes.

2. The description of the study objectives is somewhat unclear. In Section 2.3.1, model evaluation is referred to as the "second objective", and in Section 2.3.2, model selection is described as the "third objective". However, the first objective is not explicitly defined. It seems that the construction or documentation of the CMIP6 CORDEX-SEA ensemble (Section 2.1) may correspond to the first objective, but this is not clearly stated. It would help to clearly define all study objectives earlier in the introduction. In addition, the distinction between Sections 2.2 (data) and 2.3.1 (evaluation methodology) could be made clearer to avoid possible overlap.

Thank you so much for these helpful suggestions. In response, we have revised the closing paragraphs of Section 1 (Introduction) to explicitly define all three core objectives in the manuscript (Lines 143–162).

The revised closing paragraph now explicitly states the three objectives as follows:

"To bridge those gaps and to support the robust application of dynamically downscaled datasets in climate impact studies, this study presents the first comprehensive cross-ensemble evaluation of CMIP6 CORDEX-SEA simulations. Specifically, we assess their capacity to reproduce historical climate characteristics, thereby benchmarking the ensemble for subsequent kilometre-scale dynamical downscaling. The objectives of this research are threefold:

(1) to document the experimental design for producing the updated CMIP6 dynamical downscaled climate projection over SEA;
(2) to present the first assessment of CMIP6 CORDEX-SEA simulations, which serves as a standard reference for CMIP6-downscaled CORDEX simulations over Southeast Asia, specially by providing a quantitative evaluation of how well each GCM–RCM combination reproduces different characteristics (e.g., climatological mean, spatial patterns, seasonal cycle and the whole distribution) of maximum, mean, and minimum temperature as well as precipitation for the entire SEA region and its two sub-regions: the Mainland and the Maritime Continent (MC);
(3) to benchmark precipitation and near-surface temperature from CMIP6 CORDEX-SEA ensembles using the BMF developed by Isphording et al. (2024) for subsetting models that can be considered for further kilometre-scale downscaling over the SEA megacities under the CARE for SEA Megacities project and to also consider other factors, including model dependence and spread in future response, given the fact that a balanced selection of models across the different model "families" spanning a reasonable range of future projections is desirable (Brunner et al., 2020; Di Virgilio et al., 2022; Nguyen et al., 2024; Grose et al., 2023; Sobolowski et al., 2025)."

As correctly noted by the reviewer, the first objective corresponds to documenting the updated experimental setup and domain expansion of the new CMIP6 CORDEX-SEA ensemble. We have revised the text in the subsequent sections (Lines 168 and 241) to sharpen the distinction between the baseline data descriptions (Section 2.2) and our evaluation and benchmarking methodologies (Sections 2.3.1 and 2.3.2). These modifications eliminate any potential thematic overlap and ensure a much cleaner structural flow.

3. Around line 321, the text refers to Fig. 1 for tas biases, while the results appear to correspond to Fig. 2.

Revised (L368).

"These are the major contributions to the cold biases in tas found in Fig. 2."

4. The discussion of the seasonal-cycle benchmark over the Maritime Continent is reasonable. However, a brief explanation of why the unimodal benchmark is not applicable in this region would improve clarity. Adding a supporting reference may also help.

Thank you for this constructive suggestion. We agree that adding an explanation of why the unimodal benchmark does not apply to the Maritime Continent (MC) improves clarity. MC spans the equator; the semi-annual overhead passage of the sun naturally drives a bimodal surface temperature regime. This is characterised by two periods of rapid warming in March–April and October–December, respectively, and a period of rapid cooling in June–August (Li et al., 2018). In addition, changes in oceanic dynamics (including ocean heat content changes), interactions with localised monsoonal wind shifts, and regional cloud-convection feedback further complicate this bimodal cycle (Chang et al., 2005; Aldrian and Susanto, 2003).

We have revised the text in Section 3.3.2 (L623-631) accordingly to incorporate this explanation and have included the foundational references (Aldrian & Susanto, 2003; Chang et al., 2005; Li et al., 2018)

"Note that the MC spans the equator, where the semi-annual overhead passage of the sun naturally drives a bimodal surface temperature regime (Li et al., 2018). This regime is characterised by two periods of rapid warming in March–April and October–December, respectively, and a period of rapid cooling in June–August (Li et al., 2018). In addition, changes in oceanic dynamics (including ocean heat content change) and the interaction with localised monsoonal wind shifts and regional cloud-convection feedbacks complicate this bimodal cycle (Chang et al., 2005; Aldrian and Susanto, 2003). Given that the original benchmark assumes a strict unimodal annual cycle, this benchmark is not strictly applicable there. Thus, all models are considered to pass for the BMF of the temperature seasonal cycle over the MC, although 11 of 21 simulations fail to capture the observed solar-irradiance peak in Oct–December (ranks 10–11–12) over the MC, instead simulating cooler conditions around ranks 5–6."

*Aldrian, E., Dümenil-Gates, L., Jacob, D. et al. Long-term simulation of Indonesian rainfall with the MPI regional model. Climate Dynamics **22**, 795–814 (2004). https://doi.org/10.1007/s00382-004-0418-9*

*Li, Y., Yang, S., Deng, Y. et al. A process-level attribution of the annual cycle of surface temperature over the Maritime Continent. Clim Dyn **51**, 2759–2772 (2018). https://doi.org/10.1007/s00382-017-4043-9*

Chang, Chih-Pei, Zhuo Wang, John McBride, and Ching-Hwang Liu. "Annual cycle of Southeast Asia—Maritime Continent rainfall and the asymmetric monsoon transition." Journal of climate 18, no. 2 (2005): 287-301.

Overall, this is a solid and valuable study with clear relevance to regional climate modelling. The methodology is sound, and the results are useful, although the presentation and narrative structure could be improved to better highlight the key contributions.

We sincerely thank the reviewer for these encouraging and positive remarks and valuable structural recommendations. We have thoroughly revised the presentation and narrative flow throughout the manuscript in light of the reviewer's comments, which we believe has significantly sharpened the clarity of the manuscript and effectively highlighted the key contributions of our study. We hope these comprehensive updates meet your approval.