

Review of ‘A Promising Downscaling Strategy for Topographic Heavy Rainfalls over the Asian-Australian Monsoon Region by Leveraging Multi-Scale Dynamical Control’

This manuscript proposes a process-based downscaling framework for topographic heavy rainfall that combines (1) VAE (variational autoencoder - a generative machine learning model) trained on daily IVT (vertically integrated vapor transport) fields to define large-scale regimes, (2) an ‘upstream IVT’ analysis to distinguish two rainfall drivers, and (3) large-eddy simulations (LES) over realistic Taiwan topography.

The goal of connecting a multi-scale moisture environment to high-resolution impacts is timely and of scientific interest. However, I think that the study’s core conclusions are not supported by the results or methodology. I have some concerns with the assumptions going into the methodology (see below). The framework ultimately reduces to two well-known drivers (local-circulation vs. upstream-inflow), while the ML component neither discovers new physics nor demonstrably improves predictive skill. Regime identification is heuristic and arbitrary, and the validation/uncertainty quantification in supplementary materials raise some questions. The manuscript consists of very dense, wordy paragraphs which do not clearly explain the concepts, methodology or results.

Unfortunately, because of these factors, I do not recommend that this paper is published.

I have summarised some of my major concerns below:

1) Incorrect calculation of IVT

I am a little concerned by the authors’ definition of IVT. IVT is a vector field but it is written as if it is a scalar field. The written equation (Eq 1, pg 4) is incorrect. The scalar IVT should be the magnitude of IVT. I.e.

$$(IVT_x, IVT_y) = \left(-\frac{1}{g} \int_{1000}^{700} qu \, dP, -\frac{1}{g} \int_{1000}^{700} qv \, dP \right)$$
$$IVT = \sqrt{IVT_x^2 + IVT_y^2}$$

Where IVT_x and IVT_y are the x and y components of the (vector) IVT field, u and v are the x and y components of the velocity field, q is the specific heat, g is the acceleration due to gravity and P is the pressure.

2) Added value of machine learning approach is not clear

The title of the paper claims that ‘multiscale moisture dynamical control’ is leveraged providing a ‘promising downscaling strategy’ for topographic heavy rainfall. However, despite the VAE approach, the downscaling ultimately boils down to two pre-known drivers (local-circulation vs. upstream-inflow; Sec. 3.2–3.3; Fig. 4–6). The upstream in-flow regime is well documented, as the authors state. The spatial structure of IVT could be found simply by compositing IVT during days dominated by upstream-inflow. The VAE methodology seems unnecessary here and ultimately identifies only one relevant large-scale regime (upstream inflow). The difference between the EOF method and the VAE method on the test data in S5, does not look significant.

3) Lack of validation of LES

LES evaluation uses ~30 cases per type (Fig. 6, page 15 and ll 303, page 13). It is claimed that there is a qualitative pattern agreement but from examining Fig 6, it looks like the results from LES don’t agree well with the observations. Both 6b and 6e (observed) look different compared to 6c and 6f (LES model). No formal skill metrics (e.g. spatial correlation) are provided..

4) Rainfall over complex topography is not investigated

The manuscript cites complex topography as the motivation for the framework, yet later sections refer only to ‘localised’ vs. ‘widespread’ rainfall without explicitly linking these patterns to terrain-driven processes. LES simulations could have been used to quantify orographic lifting, slope orientation effects, and land–sea breeze interactions. Indeed, rain gauges on complex terrain are less reliable since they are difficult to access and are generally sparser. Analysis of this sort would be of scientific interest. Without explicit topographic diagnostics (e.g., upslope vs. lee-side differences), the study does not fulfil its stated goal of providing interpretable projections for complex terrain.

5) Incorrect or unclear calculation of rainfall metrics (R and A)

The study uses metrics R (ratio of ‘area summed’ rainfall over Taiwan compared to that of southwest Taiwan) and A the area averaged rainfall over southwest Taiwan (Eq. 3 and Eq. 4, respectively, page 10). It is unclear what ‘area summed’ means in the calculation of R and A. Furthermore, it is unclear what A is measuring. A (defined by Eq. 4) appears to estimate the regional rainfall intensity by summing raw rain gauge totals and dividing by gauge count for averages. This is incorrect. Furthermore, if R does faithfully provide a measure of how localised the rainfall is, Fig 4c seems to show that there is no relationship between R and IVT intensity or direction.

6) Over-reliance on IVT-only for regime definition

The VAE is trained solely on IVT maps (Sec. 2.2), after daily min–max normalization and coarsening (see point 8). Two environments with similar IVT can produce very different rainfall outcomes. This use of IVT alone is not adequately justified.

7) Heuristic classification of regimes

Regime grouping (36 latent sub-cells divided into categories A, B and C) is based on visual inspection rather than an objective clustering approach (Sec. 3.1; Fig. 3). This undermines reproducibility and weakens claims of ‘interpretable ML’. No quantitative mapping is provided between the latent coordinates and physically distinct processes beyond arbitrarily grouping different patterns in the IVT field.

8) Assumption of Gaussian statistics and issues with min-max normalisation

The VAE framework assumes IVT follows a Gaussian distribution (Sec. 2.2). IVT is non-Gaussian, tends to be heavy tailed with extreme values occurring more often than predicted by a Gaussian distribution. This assumption would underrepresent the extremes, which are the central process being investigated in this study.

Furthermore, I am concerned about the daily min–max normalisation of IVT in Sec 2.2, Eq.2. This places low IVT days (that would have very little impact on the terrain driven rainfall) on equal footing with high IVT days. The authors own analysis in S6 shows that the normalisation is flawed. The normalisation should account for the variability in IVT intensity. The normalisation used also makes tropical-cyclone cores appear weaker, as the authors acknowledge in S6, this undermines the choice of defining category C as ‘tropical cyclone related’ (Sec. 3.1).