https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

Downscaling using Deep Convolutional Autoencoders, a case study for South East Asia

Oliver Levers[1], Dorien Herremans[1], Anurag Dipankar[2], and Lucienne Blessing[1]

[1]Singapore University of Technology and Design (SUTD), 8 Somapah Rd, Singapore 487372
[2]Institute for Atmospheric and Climate Science, ETH Zürich

Correspondence: Oliver Levers (oliverlevers@gmail.com)

Abstract.

Inspired by recent advancements in the field of computer vision, specifically models for generating higher-resolution images from low-resolution images, we investigate the utility of a deep convolutional autoencoder for downscaling and bias correcting climate projections for South East Asia (SEA). Downscaled projections of 2 m surface temperature are generated, using autoencoders trained with data from the Coupled Model Intercomparison Project Phase 5 (CMIP5) and data from the fifth generation ECMWF atmospheric reanalysis (ERA5) project. Using CMIP5 projections as an input, three sets of downscaled data are generated using three methods of autoencoder training, which allow us to determine how autoencoder downscaling and bias correction modify temperature values. Where possible, the downscaled outputs are compared against the Southeast Asia Regional Climate Downscaling/Coordinated Regional Climate Downscaling Experiment–Southeast Asia (SEACLID/CORDEX–SEA) project and outputs from available CMIP6 experiments, to evaluate performance. The autoencoders are found to excel at the rapid generation of highly spatially-resolved climate projections for surface temperature. Realistic spatial features due to coastal and topographic variation are generated by the autoencoder, which are not present in the CMIP5 projections. Additionally, the autoencoders are capable of generating forecast data with regional temperature profiles exceeding that of those appearing in the training set (out-of-sample extrapolation). Seasonal temperature cycles are retained after downscaling throughout the region, despite the absence of temporal information provided to the model. However, autoencoders trained to carry out bias correction display a tendency to smooth daily average temperatures and reduce daily highs and lows beyond that which can be expected to be realistic. Without bias correction, downscaled outputs have a reduced improvement in spatial resolution but the daily temperature profiles of the CMIP5 input forecasts are maintained. Autoencoders rely on the presence of structural features in the datasets to carry out downscaling, and so performance over the oceans is reduced as strong temperature gradients are absent. For this reason, ocean warming is not well represented, an artefact which is not immediately clear in the downscaled outputs. This study demonstrates the importance of rigorous analysis of 'black-box' methods, which can generate non-obvious artefacts that could potentially create misleading results. Despite these limitations, Autoencoders are clearly capable of generating much needed high-resolution climate projections, and strategies to improve upon shortcomings are numerous and well established.

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

1 Introduction

Tropical regions with large coastlines and small islands are particularly at risk from climate change (Ge et al., 2021; Lee et al., 2013), yet high-resolution climate projection data for Southeast Asia is lacking. Global Climate models (GCMs) typically have spatial resolutions in the order of 100 km, roughly equivalent to the width of many archipelagos, peninsulas and islands which make up Southeast Asia (Chotamonsak et al., 2011). This significantly reduces the utility of GCM projection data, as decision makers lack key information at sufficient spatial scales. Increasing the spatial resolution and accuracy of climate projections must therefore be considered a priority, as well as the dissemination and utilisation of this information so that adaptive measures can be taken (Schär et al., 2020).

Generating high spatial-resolution climate projections can be achieved via dynamical downscaling or via statistical methods. Dynamical downscaling using regional climate models (RCMs) generates full physical simulations of future climate, but are subject to the same limitations as GCMs including sensitivity to the representation of sub-grid atmospheric processes by parameterisation schemes (Tangang et al., 2020) (Juneng et al., 2016a), and require similar levels of computational resource to full scale GCMs. Additionally, RCMs are highly sensitive to the quality of boundary and initialisation settings (Warner et al., 1997). This is problematic for RCMs in SEA due to the regions interactions with large scale atmospheric processes such as monsoonal activity (both north-east and south-west monsoon) (McSweeney et al., 2015) and El Nin˜o/Southern Oscillation (ENSO (Thirumalai et al., 2017). The Southeast Asia Regional Climate Downscaling (SEACLID) / Coordinated Regional Downscaling Experiment (CORDEX) is currently the only dynamically downscaled dataset publicly available. It uses scenarios and models from the CMIP5 suite of experiments. The increased spatial resolution is suitable for most applications, between 25-36 km, and future experiments aim to increase the spatial resolution to 3-5 km, using data from the latest (Coupled Model Intercomparison Project) CMIP6. However, the delay in publication of CORDEX data means that decision makers now have both higher resolution CORDEX data (RCP4.5 and RCP8.5 only) or updated and mid-range resolution CMIP6 data to choose from, of which around half of the 100 models are published. CMIP6 experiments have greater spatial and temporal resolution than the previous CMIP5 experiments, a greater number of scenarios, and are published 10 years after CMIP5, representing a significant improvement in modelling and increased quantity and quality of observational data. However, the spatial resolution is still only of the order of $0.25° \times 0.25°$, 100 km nominal resolution for land, although limited HighResMIP data is already available at 50 km x 50 km resolution.

Traditional statistical downscaling methodologies derive statistical relationships between GCM ouputs and long-term observational datasets to generate high resolution datasets. Statistical methods are computationally faster and cheaper than dynamical modelling but performance is limited by the lack of region-specific long term, high quality weather station data. Long-term observational data is not readily available for SEA as other regions, which makes statistical downscaling more difficult. Additionally the inability of statistical models to replicate complex non-linear processes has hindered their usage, but offer a complimentary approach to dynamical downscaling now climate data is more widely available. In 2020, high resolution and bias corrected statistically downscaled data was made available via the CCAFS-Climate data portal (http://www.ccafs-climate.org/statistical_downscaling_delta_cmip5/) as mean monthly maximum and minimum temperature

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

and monthly rainfall Navarro-Racines et al. (2020). This was carried out using the delta method (thin plate spline spatial interpolation of anomalies) to bias correct 35 CMIP5 climate models to around 1-km spatial resolution. These products are of great value to those who need to model vegetation dynamics where temporal information is less critical than spatial variation. Statistical methods, which lack the physical basis of RCMs, are often considered less 'trust-worthy' than other techniques, and validating outputs is critical. Projects such as the VALUE framework (http://www.value-cost.eu/) has been set up to assess the performance of downscaling methodologies, but the application is limited to Europe.

Recently, the application of Machine-learning (ML) and Deep learning (DL) techniques for downscaling has been growing in interest. The translation of image enhancement methodologies to improve spatial resolution (downscaling) of regional data is now well established (Vandal et al., 2018; Sachindra et al., 2018; Chang et al., 2018; Baño-Medina et al., 2020; Ji et al., 2020; Xu et al., 2020). These methodologies offer many of the benefits of traditional statistical methods (speed and low computational complexity) but with the potential to emulate non-linear processes and complex physical and spatial relationships (Mansfield et al., 2020). Convolutional neural networks, of interest because of their ability to capture spatial relationships, appear to have advantages over more traditional neural networks. Image super-resolution has been employed to enhance the spatial resolution of a number of mostly atmospheric variables in weather and climate forecasting. ML techniques are particularly suited to variables which are stochastic in nature and have features of a finer spatial scale, such as cloud cover and precipitation. These models have also been shown to speed up forecasting requiring less time than RCM outputs (in the order of 12%)(Chang et al., 2018).

As with all statistical learning methods, neural network models require large, high-quality datasets to achieve high performance. Even with the increasing availability of high-quality remote-sensing and data assimilated datasets, DL methods are not infallible, and it is important to rigorously assess potential limitations or introduced bias when evaluating complex, 'black-box' style techniques. For example, DL models based on convolutional methods often demonstrate the ability to generalise well and routinely outperform traditional dense neural network models. However, these methods are prone to overfitting (Xu et al., 2020) and performance and generalisability is likely to be somewhat data and model architecture specific. Demonstrating rigorous performance and trust in ML methodologies is essential, especially where numerous bench-marking datasets are not available (Baño-Medina et al., 2020, 2021). Additionally, identifying a complete set of metrics to assess performance and is a challenge, and ensuring region specific characteristics are maintained may need to be taken into account. The geography of SEA may present additional challenges to convolutional methods, akin to those found in dynamical downscaling approaches or possibly unique to the technique. Once these challenges are identified, more complex model architectures can be designed which target key limitations ensuring that the iterative process of improvement is not random, and meets the individual requirements of the region.

Here, the performance of a deep convolutional image super-resolution model is assessed in the context of spatial downscaling for climate projections in Southeast Asia. Because of the limited bench-marking datasets available, the aim is to identify potential limitations of the model, any unwanted artefacts, and any region specific considerations. 2m surface temperature is explored as a variable present in most model outputs with well understood interactions and is less stochastic in nature compared to variables such as precipitation. Surface temperatures also have strong relationships to topographic features which

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

is convenient for the analysis of generated temperature gradients. Selected CMIP5 model outputs are downscaled using ERA5 Reanalysis data as a target dataset, which simultaneously increases spatial resolution and aligns model output with current observational trends (bias correction). CMIP5 data was used as an input to the model so that both CORDEX and CMIP6 projections can be used as bench-marking datasets. We aim to highlight potential pitfalls in the approach by generating three comparable datasets using autoencoders trained to isolate both the downscaling and bias correction functions of the model.

2 Proposed architecture

A deep 2D convolutional autoencoder constituting of fourteen layers was developed in this study (Figure 3). The model architecture is inspired by autoencoders applied to image super-resolution problems Johnson et al. (2016). The autoencoder is near-symmetrical and consists of an encoder and a decoder. The function of the encoder is to extract structural features from a series of images, which can be used to reconstruct an image. The convolutional encoder achieves this via a series of convolutional and MaxPooling layers which decrease in size (dimensionality reduction), resulting in the generation of a small encoding known as a 'latent representation'. This representation of the input image is then passed to the decoder, which is trained to reconstruct an image from its latent representation. In reverse of the encoder, the decoder is capable of reconstructing an image via a series of convolutional layers which increase in dimensionality, until the output matches that of the target (as well as input) image size. In image reconstruction, this process is lossy, since the encoder disregards information which is considered to be noise and unimportant for the reconstruction of an image.

In this study, the input and output images used in training can be from different models, and so the function is modified from image reconstruction. In the training phase for example, the autoencoder is trained using CMIP5 hindcast data as input and high-resolution ERA5 Reanalysis data as the target. The autoencoder therefore learns to extract spatial patterns from CMIP5 images which can be used to generate the corresponding ERA5 image. Critically, the CMIP5 data is not simply a 'low-resolution version' of ERA5 images. This means that the autoencoder may learn to 'translate' between CMIP5 and ERA5, 'correcting' for model bias and systematic errors present in the CMIP5 dataset. This function of the autoencoder may present a more difficult challenge than the generation of fine structure from low resolution images, especially since the autoencoder was originally conceived for image reconstruction. It is important to assess how well the autoencoder can achieve both downscaling and correction of CMIP5 data, which is achieved here through a series of differently trained autoencoders.

In order to determine how downscaling and correction carried out by our proposed autoencoder affects the quality of the generated results, three different autoencoder architectures are trained (Figure 1). Autoencoder A is trained to carry out correction and downscaling simultaneously, Autoencoder B is trained to carry out the correction only, and Autoencoder C carries out downscaling only. The three autoencoders are used to generate the following results as described in Figure 2:

Autoencoder (AE): Results generated using Autoencoder A. Correction and downscaling is carried out simultaneously by a single autoencoder, using CMIP5 as an input dataset and ERA5 as a target.

Stacked (ST) Stacked models often carry out two functions separately, or a single function in a step-wise manner (eg. incremental increases in spatial resolution), which can result in an improvement in performance Vandal et al. (2018). Here,

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

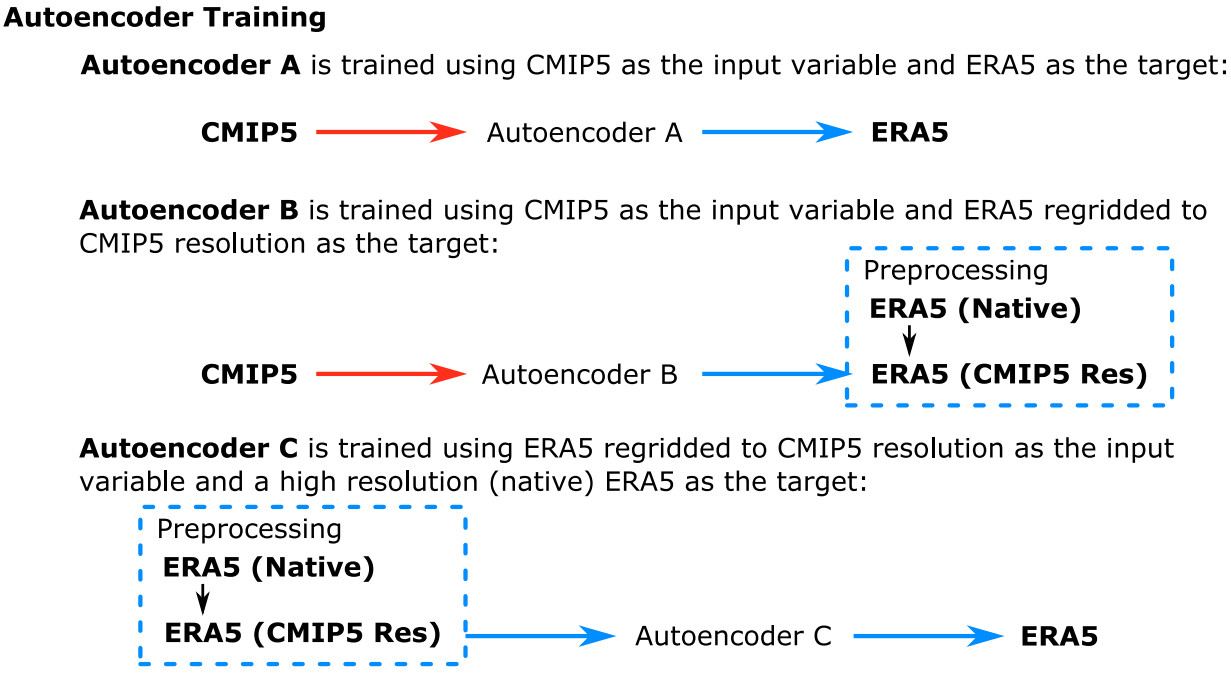

Figure 1. Three autoencoders are trained using different datasets to isolate the different functions of the autoencoder. Autoencoder A is trained to learn downscaling and correction simultaneously. Autoencoder B is trained to learn correction only. Autoencoder C is trained to learn downscaling.

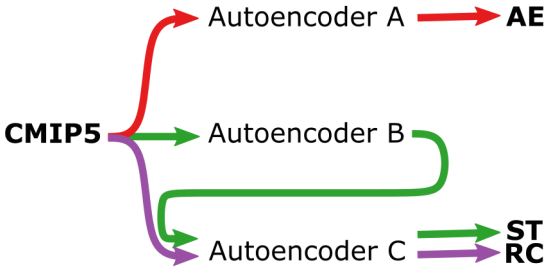

Figure 2. Downscaled outputs (AE, ST, RC) are generated using one or two of the trained autoencoders. AE: Autoencoder, generates downscaled and corrected data using Autoencoder A. ST: Stacked model, utilises Autoencoders B and C to generate downscaled and corrected data. RC: Reconstruction, mimics an image reconstruction experiment utilising Autoencoder C to generate downscaled only data.

the two functions of correction and downscaling are carried out by two Autoencoders (B and C) and both autoencoders are applied in series to generate a dataset (ST). Since this approach uses two autoencoders in series, there is an effective doubling of convolutional layers which may also result in an increase in performance. Autoencoder B is trained to convert between the
130 CMIP5 dataset and an artificially reduced resolution ERA5 dataset, carrying out a correction of the CMIP5 dataset, without

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

carrying out downscaling. To generate low-resolution ERA5 data, the ERA5 training dataset was first regridded to the CMIP5 coordinate grid, and then rescaled back to the original ERA5 coordinate grid using nearest-neighbour interpolation, resulting in a loss of spatial information. Autoencoder C, is trained using the same low-resolution ERA5 dataset as an input, with the native high resolution dataset as the target. Thus, Autoencoder C is trained to carry out image-enhancement (downscaling) only. Since the result generated by Autoencoder B would be of low resolution, results generated by Autoencoder B and C in series are used as a comparison with other datasets.

Reconstruction (RC): Autoencoder C is trained using artificially generated low-resolution ERA5 data, with the original higher resolution ERA5 data as the target dataset. Therefore, Autoencoder C does not learn to translate between CMIP5 and ERA5 datasets, isolating the downscaling functionality. This mimics an image reconstruction experiment most closely, where models are trained to reconstruct missing details in artificially degraded images. After training the model, the CMIP5 data used to generate a new downscaled dataset (RC).

The datasets generated from our three proposed models can be compared to assess a) the benefits of stacking and separating the functionality and b) how the correction and downscaling functions of the autoencoder affect the projection data.

Autoencoders A, B and C are identical in architecture, to maintain comparability. This architecture is described in Figure 3 Model optimisation was carried out with some minor improvements in MSE, but were not significant for the aims of the study and complicate dataset comparisons unnecessarily.

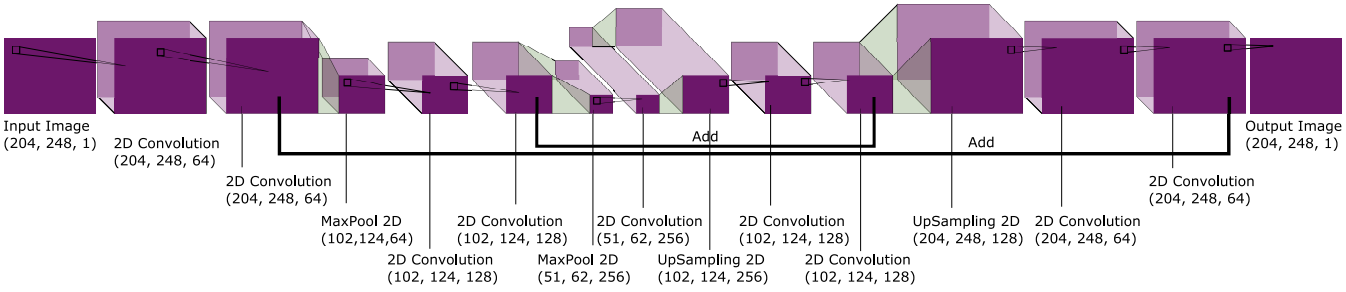

Figure 3. Model architecture of the autoencoder used for Autoencoders A-C. Brackets denote (height, width, channels) in pixels. Kernel size is (3,3) for 2D convolutional layers.

3 Experimental setup

3.1 Datasets

3.1.1 Dataset selection

Within the CMIP6 and CMIP5 deck of experiments, there are GCMs which are able to capture the seasonal and climatic patterns present in SEA better than others (Raghavan et al., 2018; Kamworapan and Surussavadee, 2017). Typically, performance is evaluated via the models' ability to capture historical trends, and large seasonal processes such as the monsoon (McSweeney

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

et al., 2015). One set of models which have consistently performed well in the SEA region are those developed jointly by CNRM-GAME (Centre National de Recherches Météorologiques—Groupe d'études de l'Atmosphère Météorologique) and Cerfacs (Centre Européen de Recherche et de Formation Avancée) (Voldoire et al., 2019; Kamworapan and Surussavadee, 2017). An analysis by Kamworapan et al (Kamworapan and Surussavadee, 2017, 2019) stated that CNRM-CM5-2 and CNRM-CM5 were the top rated models for studying climate in the region, and rated these highest for modelling temperature values within the CMIP5 deck of experiments. Fortunately, CNRM-CM5.1 (CMIP5) and CNRM-CM6-1 (CMIP6) data is available publicly via the Copernicus data store, as is CORDEX regional downscaled data generated using the CNRM-CM5 model. A cool bias has been reported for these models, however an improvement in performance from CMIP5 to CMIP6 has been noted due to improved modelling of the shortwave cloud radiative effect Ge et al. (2021). For this reason, data derived from CNRM-CM5 experiments was used as the input for the models. The output of the downscaling was compared against CNRM-CM6 models, which have resolutions of around 100 km, although high resolution (HR) data is also available. CMIP6 datasets may provide better training data in the future for downscaling using ERA5, but CMIP5 data was used to ensure a benchmarking dataset was available. CORDEX outputs which are run using modified CNRM-CM5 are also included, since this is the best regional model available at resolutions exceeding 100 km, current CORDEX data is available at 25 km resolution (CNRM-CERFACSS-CNRM-CM5). A complete set of experiments is not available for comparison however, CORDEX experiments are only available for Representative Concentration Pathways (RCP) scenarios 4.5 and 8.5. Additionally, CMIP6 experiments do not use RCPs, and report projections using the Shared Socioeconomic Pathways (SSPs). SSPs and RCPs are not equivalent, as SSPs contain updated and more variable socio-economic pathways, but result in the same total radiative forcing by 2100. This means that some RCPs and SSPs can be compared, especially in the latter years of the projections. By utilising GCMs and RCMs which use the same CNRM model as a base, it is hoped that regional biases and systematic errors are consistent and easier to identify.

CORDEX simulations are constrained to a smaller region than that selected for downscaling in these experiments. This is because RCMs for the SEA region are destabilised by the presence of the Tibetan Plateau at the north-western boundary. For this reason, a domain of 90E-145E and 15S-27N is available for CORDEX experiments. A larger domain is used for autoencoder downscaling, to assess the effect and performance over the Tibetian Plateau and to include the islands of the Philippines. For comparison with CORDEX, the results in this study are occasionally 'cropped' to the CORDEX domain so that they remain comparable.

ERA5 Reanalysis data is available hourly up to 2020 on a 30 km grid at hourly intervals. It utilises hindcast models, observational and remote sensing data and uses data assimilation to generate high resolution and highly accurate estimates of past climate. Due to the limited observational data available in this region, ERA5 Reanalysis data was used as the high-resolution target dataset for training.

A summary of the datasets used in this study is included in Table 1.

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

Dataset	Abbreviation	Spatial Resolution	Source
CNRM-CM5	CMIP5	$1.4008^\circ \times 1.40625^\circ$ (150 km grid)	(Voldoire et al., 2012)
ERA5 Reanalysis data	ERA5	$0.25^\circ \times 0.25^\circ$ (30 km grid)	(Hersbach et al.)
CNRM-CM6-1-HR	CMIP6	$0.25^\circ \times 0.25^\circ$	(Voldoire, 2019)
SEACLID/CORDEX, CNRM-CERFACSS-CNRM-CM5	CORDEX	25x 25 km grid	(Juneng et al., 2016b; Ngo-Duc et al., 2017) (Cruz et al., 2017; Tangang et al., 2018)

Table 1. A table summarising the datasets used in this study. CMIP5 and CMIP6 is used for model training, CMIP6 and CORDEX datasets are used for validation and bench-marking.

3.1.2 Data preprocessing

Model training was carried out using CMIP5 and ERA5 Reanalysis data from 1990-01-01 to 2004-12-31, constituting 5,478 days of which an out-of-time validation set of 20% was removed. The year of 2005 (01-01-2005 - 12-31-2005) was excluded from the training set and used as a test set, to evaluate the performance. We selected this data range as hindcast data was available in this date range for all datatsets CMIP5, CMIP6, CORDEX and ERA5 experiments.

Climate data was converted from NetCDF to numpy arrays using python. This resembles a stack of grey-scale images, each with one channel. Due to the differences in projection and resolution, ERA5 and CMIP5 are provided with differing spatial grids. To ensure that the numpy arrays generated from the climate data contained information in each pixel representing the same geographical area, CMIP5 data was first regridded to match the regular 30 km grid of ERA5 Reanalysis. CMIP5 data was regridded and upsampled using nearest-neighbour interpolation. Nearest-neighbour interpolation was used over other interpolation methods which could smooth values, resulting in the loss of local minimas and maximas. Regridding was carried out using the xESMF package in Python. The ERA5 data was downloaded as hourly data, and daily means were generated using NCO netCDF operators using bash in terminal.

Scaling, normalisation or standardisation of input data is commonly carried out to improve the training of ML models. Although commonplace, this is generally beneficial when using a range of datasets with large variations in absolute values. All datasets were normalised using the same min-max scaling factor, which covers the full temperature range of both the CMIP5 projections and ERA5 target datasets. Temperature data was normalised to be between 0-1 within a range of -50 to 50 °C. This encompasses the lower and upper range of all datasets used, however, it also constraints the majority of the datapoints to a narrower range following normalisation.

3.2 Dataset analysis

An evaluation of the training datasets and benchmarking datasets was carried out to determine bias and to gauge acceptable limits for error. Historical data from each of the datasets was compared in the region between the dates of 01-01-2000 and 01-

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

01-2006 (Figure 4). Where CORDEX experiments are included for comparison, results are cropped to the CORDEX region. Surface temperatures averaged over the region appear to be very similar for CMIP5 and ERA5 datasets. CMIP5 has a narrowly larger range of average daily temperature values, tending towards more cooler temperatures in the cooler months. ERA5 Reanalysis gives warmer yearly-averaged surface temperatures than CMIP5, and this difference is increased over the full region used in these experiments, suggesting that stronger than average differences exist between CMIP5 and ERA5 in the northerly regions missing from the CORDEX experiments. The similarity in the training datasets may increase the performance of the model, as the correction element will be less significant. CMIP6 is significantly cooler, with the mean, inter-quartile range and range shifted to by around 1.7 °C below that of both ERA5 and CMIP5. A cold bias has been observed previously for both CMIP5 and 6, but it appears more exaggerated for CMIP6 (Voldoire et al., 2019) despite reported improvements. Additionally, there appears to be a seasonal dip in temperatures observed in the OCT-NOV-DEC months for CMIP6, not present in other datasets, a seasonal trend more characteristic of mainland Southeast Asia than the region as a whole. A slight cold bias was observed for the CORDEX dataset, the inter-quartile range for temperatures appears to be very similar. The CORDEX data suggests that temperature ranges do not increase as the result of increased resolution, as do the similarities in the boxplots of all datasets.

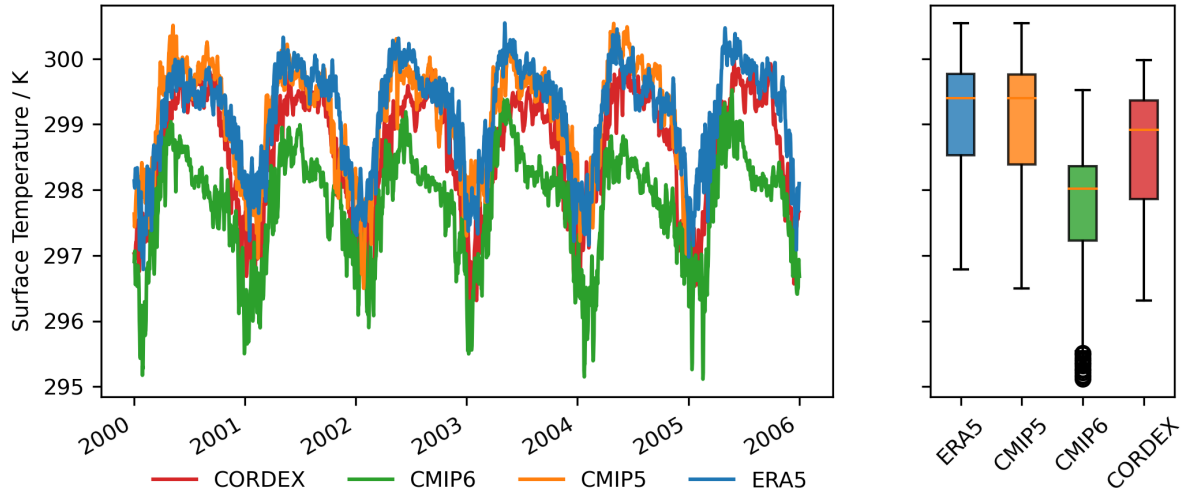

Figure 4. Daily surface temperature averaged over the SEA region defined by CORDEX between 01-01-2000 and 31-01-2005 (left). Average values are displayed only for clarity. The daily mean, interquartile range and range is shown as boxplots (right).

There is a clear distinction between surface temperatures on land and sea. With surface temperatures on land tending to be cooler and with larger seasonal variation than temperatures over sea (Appendix: Figure A1). Land constitutes only around 22% of the area of the region, and so ocean temperatures are likely to dominate averaged trends observed over the region. Both CMIP6 and CORDEX datasets show cooler seasonal cycles than CMIP5 and ERA5 datasets on land. CORDEX however, has much warmer seas which appear more similar to the CMIP5 and ERA5 datasets, with land values more closely resembling that of CMIP6.

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

3.3 Model training and evaluation metrics

Data from the period 31-12-2004 to 01-01-2006 was used as a test set to analyse the performance of all three of our proposed models. CMIP5 data was used as an input to generate results AE, ST and RC. These are then compared with ERA5 data. Training was carried out for 500 epochs but early-stopping was triggered in all cases around 200 epochs. The validation loss for all models plateaued around $2 - 4 \times 10^{-4} K^2$ epochs.

Mean square error (MSE) was used as the primary metric for both the evaluation of model performance. It was also used for model training over other loss functions which have been optimised to give more visually pleasing results in image resolution applications (Johnson et al., 2016) for simplicity. MSE provides the clearest indicators of differences between generated model data and is the easiest to interpret. Other metrics commonly employed for image analysis such as structural similarity index (SSIM) were also used but results were either complimentary to MSE or less intuitive to interpret. We also do a visual observation of the final generated temperature maps to find patterns and general observations.

4 Results

Data generated by the autoencoders showed increases in the number of spatial features and sharpness of spatial features present in the climate data. The improvement of temperature gradients representing topographic features and the reduction of pixels which spread across coastline boundaries can be observed when comparing the the predicted model outputs (AE, ST and RC) to the lower CMIP5 model input, and the ERA5 data (Figure 5). AE and ST appear to have higher resolution around topographic features as compared to CMIP5 and RC, and were closer to ERA5. RC retained features present in CMIP5 but not ERA5, such as the cooler band which appears in the Philippine Sea (Figure 5). RC shows significant enhancement in terms of spatial resolution, indicating that the translation function of the autoencoders in AE and ST are not essential for downscaling. Other forms of interpolation do not perform nearly as well in resolving coastal regions especially in regards to the 'hard edges' of land-sea borders.

4.1 Autoencoder performance

The performance of the autoencoders was assessed by comparing the output datasets AE, ST and RC against the ERA5 dataset for the year of 2005 (Table 2). MSE values were calculated by averaging daily differences over the region, and then averaging over the year to give a single value. Standard deviation of the MSE is reported (σ). Additionally, the differences between the benchmarking datasets and ERA5 were calculated and included for comparison. MSE was used over other measures such as mean absolute error to attempt to increase model performance with regards to outliers and extremes. The MSE is far higher for CMIP5 and CMIP6 than the autoencoder generated outputs, with ST and AE performing the best. Only CORDEX improves on RC, although the geographical region for CORDEX is smaller and this may reflect the improved predictions owing to the absence of the Tibetan plateau. However, the autoencoder generated results report reduced differences between daily average maximum and minimum temperature values. This contraction in temperature range is an important metric and a contraction

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

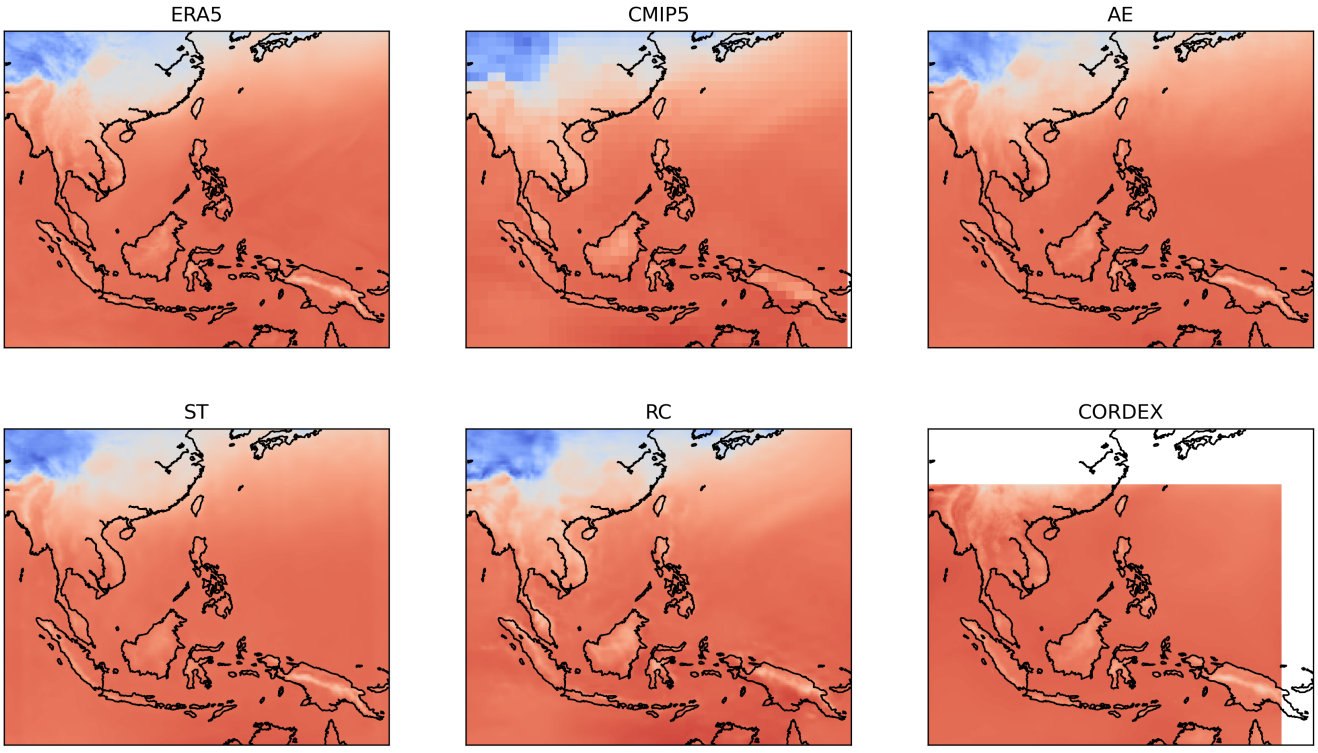

Figure 5. A single day taken from the test set, showing outputs from each model AE, ST and RC, with ERA5, CMIP5, CORDEX as comparisons. The 'pixels' of the CMIP5 input are clearly visible, overlapping coastal boundaries. Visually, the output of the autoencoders (AE, ST, RC) appear extremely similar, with AE more closely resembling that of ERA5. The output from RC appears to be more similar to that of the CMIP5 input but with significant increases in spatial resolution. The CORDEX region is more constrained than the outputs of this model, and so the north and east sections are blank. Coastal boundaries (black line) are overlayed, and are not present in the input or output images.

in these values may not be desirable. This effect is not reflected in low values of MSE which appear to suggest the model is performing well. RC narrowly overestimates the temperature range. This suggests that the autoencoder translation function present in AE and ST results in a reduction in the range of predicted temperature values. The difference in the average daily temperature maximum is more pronounced than the minimum for all models, this may be because cooler temperatures are found in regions of higher topography where climate is more stable. RC is closest for average daily maximum to ERA5, underestimating the daily average maximum by 1K.

Error over the region is not strongly temporally dependent, although larger errors are reported in the Spring months (March, April, May: Figure 6) The error is localised to mainland China, but approaches zero within the islands of the Maritime Continent. April heatwaves over mainland SEA were observed in 2005, likely exacerbated by El Nin˜o/Southern Oscillation (ENSO) phenomenon Thirumalai et al. (2017), but this does not result in significantly increased error for mainland SEA during the

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

Model	MSE / K^2	σ / K^2	Av. Daily Max / K	Av. Daily Min / K	Difference / K
ERA5	0.00	0.00	301.51956	291.05258	10.46698
ST	2.81	4.50	300.2947	292.96204	7.33266
AE	2.85	4.74	300.42657	292.68814	7.73843
CORDEX*	3.41	6.09	301.7253	294.37817	7.34713
RC	6.10	10.85	301.71252	290.02182	11.6907
CMIP5	6.86	13.93	301.74002	289.66425	12.07577
CMIP6	9.81	18.37	300.5624	288.33856	12.22384

Table 2. Model performance listed in order of increasing mean square error (MSE) calculated between model outputs (in bold) and ERA5 for the period 01-01-2005 to 31-01-2005. Benchmarking and training datasets are also included for comparison. The standard error of the mean is also given (σ). The average daily temperature maximum and minimum are given, as are the differences between these values, as an indication of the models ability to replicate the range of temperature vales. * The geographical region for CORDEX is smaller than geographical area covered by the other models and so performance may appear to be better.

spring season as might be expected. Out of the years used in the training dataset, April heatwaves in SEA were observed for 1992, 1995, 1998, 2001, 2003 and 2004.

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

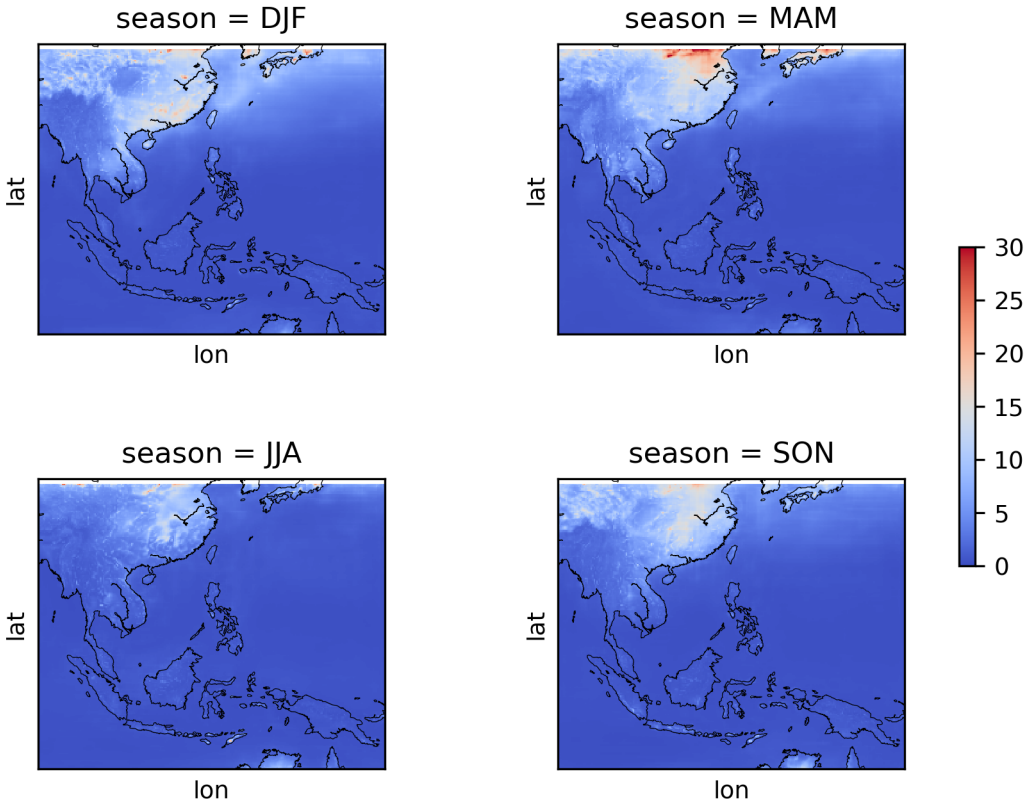

Figure 6. MSE averaged over four seasons, Winter (December January Feb), Spring (March, April, May), Summer (June, July, August), Autumn (September, October, November).

4.1.1 Spatial resolution

Regional downscaling is achieved in order to obtain more realistic temperature profiles over complex topography and around coastlines. Since the autoencoder output is not determined by any physical laws, it is important to determine if increases in resolution are realistic. Autoencoder outputs increase spatial resolution significantly with respect to the CMIP5 input, and temperature gradients appear to strongly resemble the ERA5 dataset for AE and ST (Figure 7). Temperature profiles correlate with elevation and land-sea borders (Figure 7, middle panel) and higher resolution datasets (ERA5) approximate boundaries and topography at higher resolution. Land-ocean boundaries are better represented as are the presence of small islands for AE and ST which overlap the profile of ERA5 almost exactly, especially over mountainous regions. RC is an improvement on CMIP5, and goes someway to improve land-sea boundaries. Clear differences between RC and AE/ST are observed where mountainous regions are present. Mountainous regions are accompanied by large changes in surface temperature over a short spatial scale, which may suit ML methods due to the presence of 'static' topographic features which form repeated structures in the data. AE and ST perform extremely well over mountain ranges. CMIP5 and RC underestimates the surface temperatures in

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

regions around Kalimantan in Borneo and overestimates the temperatures of West Papua. This clearly indicates the importance of translation for spatial downscaling, and the significant improvements which are achieved over areas with large changes in topography (mountains).

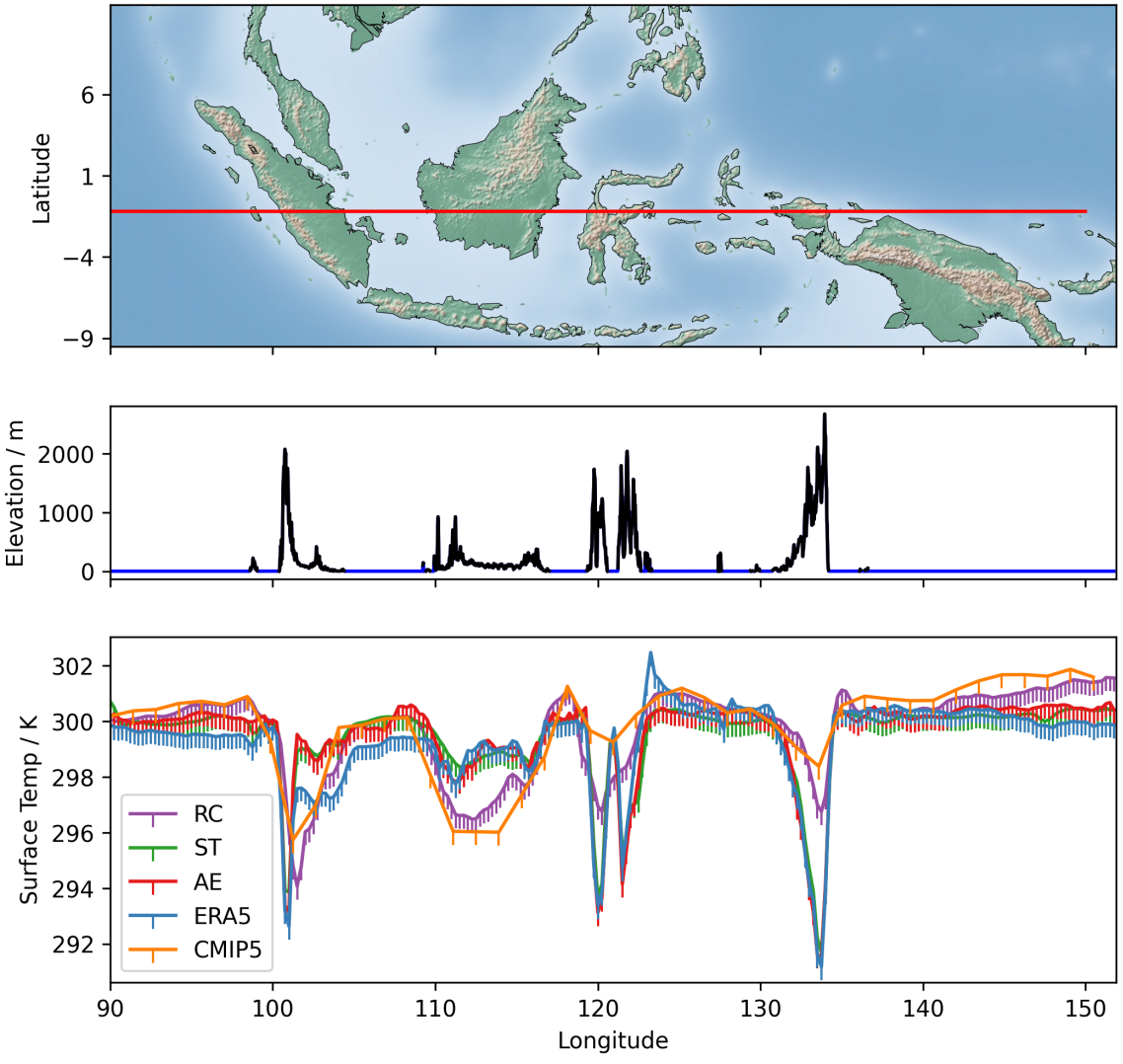

Figure 7. The spatial features of the surface temperatures correlate strongly with topographic features. Here a line is drawn across a number of islands in the region at latitude -1.2 (Top). The elevation, where ocean (blue) is set to 0 elevation can be shown to correlate strongly with the temperature profiles (middle and lower figure). The lower figure shows the surface temperature across the longitude with tick marks indicating data points. AE correlates strongly with the ERA5 target but the RC follows the low resolution CMIP5, which only loosely indicates topographic variations in temperature.

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

4.1.2 Temporal Resolution

The autoencoder is optimised for image resolution applications and temporal information is not provided to the autoencoder. The autoencoder can process images in any order and so temporal information my be lost. In order to determine how seasonal temperature cycles are affected by downscaling, 12 locations around SEA representing regions of interest are compared. These locations are representative of a few of the micro climates present within the region (Cruz et al., 2017). The locations are listed in Figure 8 and Table 3.

Despite the absence of spatial information provided to the autoencoder, seasonal cycles within the downscaled results remain realistic in all micro-climates of SEA, especially where there is good agreement for CMIP5 and ERA5 (Figure 9). Where CMIP5 and ERA5 diverge, AE and ST resemble ERA5 more closely, whereas RC retains the temperature profile of CMIP5 (Figure 9, d,h,j,k). For this reason, it is not likely that the autoencoder is introducing a strong regional bias or introducing unrealistic temporal temperature cycles.

Despite the retention of seasonal patterns and trends, it is clear that temperature ranges are modified following downscaling (Table 2). Critically, the range and difference between daily average max and min are more constrained for ST, AE. The contraction in temperature values observed in regionally averaged data (Table 2) is not due to a localised error, but present over the entire region. In Figure 9, a contraction of daily temperature values result in the smoothing of seasonal profiles for AE and ST, which is not observed for ERA5. Artificially smoothing temperature extremes, is not desirable as the number and duration of hotter days in a location are important metrics to assess climate risk. The contraction in temperature range for AE and ST for the downscaled models is significant, and beyond that which could be considered realistic when compared to those of other datasets. This is a significant limitation of AE and ST, and a trade off between temporal and spatial resolution is present when selecting between autoencoders which carry out translation. Temperature ranges for RC, which does not attempt to correct the temperature profiles to match ERA5 are more realistic and match the daily fluctuations of CMIP5 (Figure 9)

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

Letter	Latitude	Longitude	Location
a	24.797037	102.833037	Kunming Yunnan
b	30.573978	114.302702	Wuhan Hubei
c	26.061373	137.804832	Philipine Sea North
d	13.681586	100.494526	Bangkok
e	10.823328	106.648856	Ho Chi Minh City
f	14.584877	121.074389	Manila
g	1.316031	103.82534	Singapore
h	-1.180855	112.569417	Central Kalimantan
i	3.176664	142.27385	Philipine Sea South
j	-3.267858	103.866476	South Sumartra
k	-2.172308	120.336446	Sulawesi
l	-6.704243	140.995719	Papua New Guinea

Table 3. The latitude and longitude, and location of points selected within the SEA region.

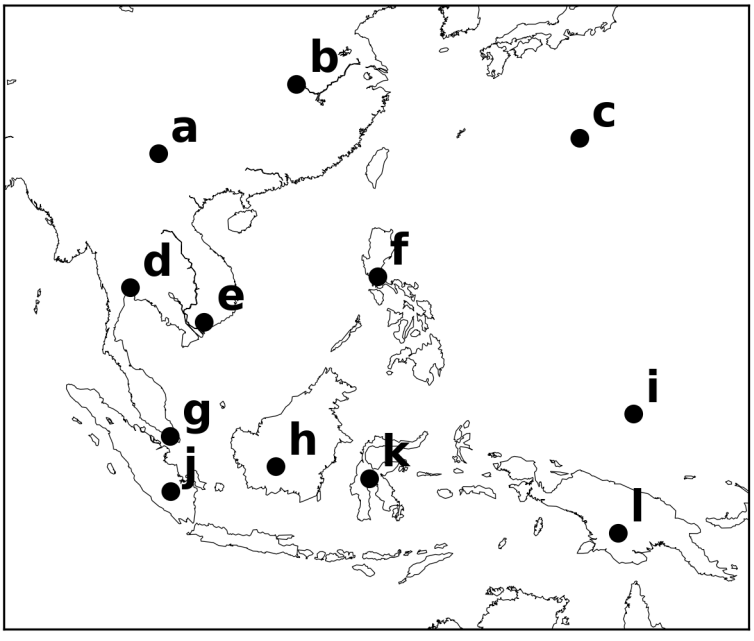

Figure 8. 12 regions within the region, representing the different micro-climates in the region, and major cities likely to be significantly affected by climate change.

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

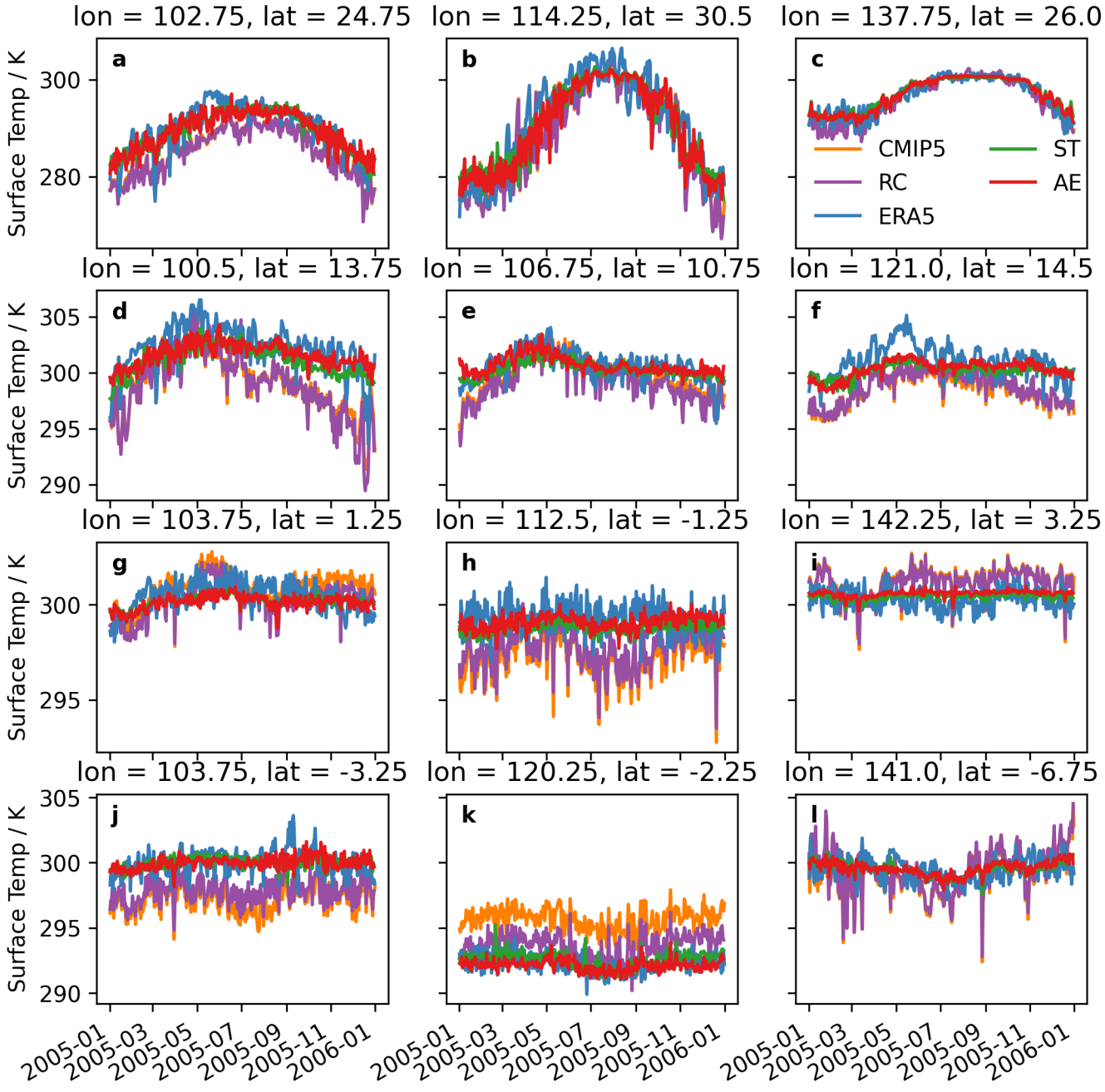

Figure 9. Selected locations within the SEA region, showing the CMIP5, ERA5, AE, ST and RC outputs between 01-01-2005 to 31-12-2005. Note: Each row has different y-axis range

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

4.2 Downscaling CMIP5 climate projections

CMIP5 climate projections for RCP 2.6 and RCP 4.5 are downscaled using the trained autoencoders and compared to the CMIP5 and CMIP6 datsets. The autoencoders are clearly able to predict temperature values outside those present in the training set. Yearly averages are modified significantly for both RCPs by AE and ST, however RC retains the temperature profile of CMIP5 data (Figure 10). This demonstrates the significant modifying effect of the 'translation' function present in AE and ST datasets. The surface temperatures are consistently warmer for both AE and ST in the early years for both projections, following the observed trends for the ERA5 training set. Notably, AE and ST are not simply offset from the CMIP5 and RC projections by a constant value. For RCP4.5 this results in a reduced gradient of warming, with regional averaged temperatures eventually approximating the same regional average surface temperature as the other datasets by 2100. Why temperature gradients are modified is not clear, however this suggests the autoencoders are predicting temperatures using non-linear relationships.

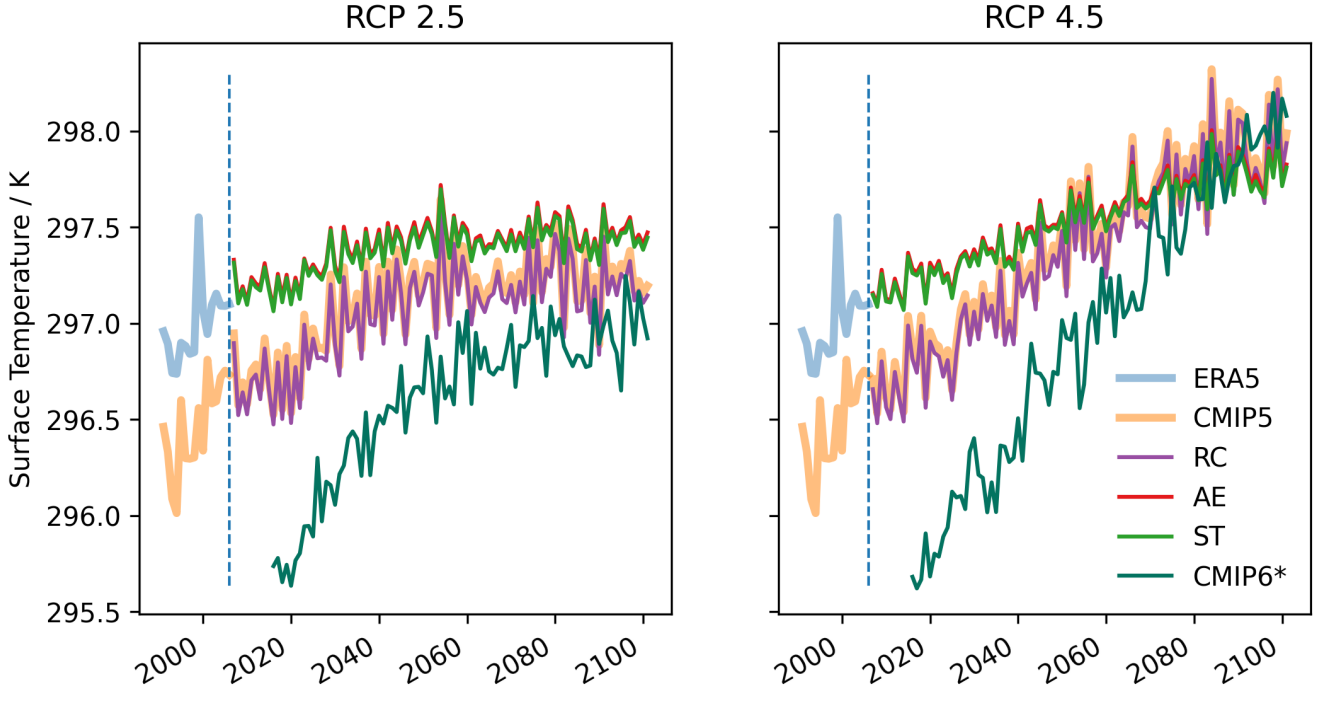

Figure 10. Yearly-averaged daily surface temperatures over the SEA region for all datasets for RCP 2.6 and RCP 4.5. Vertical dotted line represents transition from training data to climate projections of daily surface temperatures. Yearly average values are presented only to maintain clarity as the range of values within these averages are significant and overlap considerably. *CMIP6 data shown are for SSP126 and SSP245 which are not equivalent to the RCPs shown.

Mapping the temperature differences between the downscaled outputs (AE, ST, RC) and the equivalent CMIP5 projections was used to indicate where significant changes to the input CMIP5 datasets had been introduced by the autoencoder downscaling (Figure 11). These results indicated that significant differences were present in mainland Asia, and warmer temperatures

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

were introduced for land than over the oceans. The effect of the higher resolution around islands can be seen by the cooling of temperatures in areas of higher topography, for example, the central part of New Guinea. 'Residues' due to the difference between gradients generated by the finer resolution downscaled outputs and the outputs from the coarser pixels of CMIP5 experiments can also be observed. This creates 'hotspots' at the edges of these larger pixels, which means that areas with finer topographic detail appear to have greater differences from the CMIP5 datasets. This maybe why the ocean is appears to have little change from the CMIP5 input, as there are no topographic features or strong changes over small spatial scales.

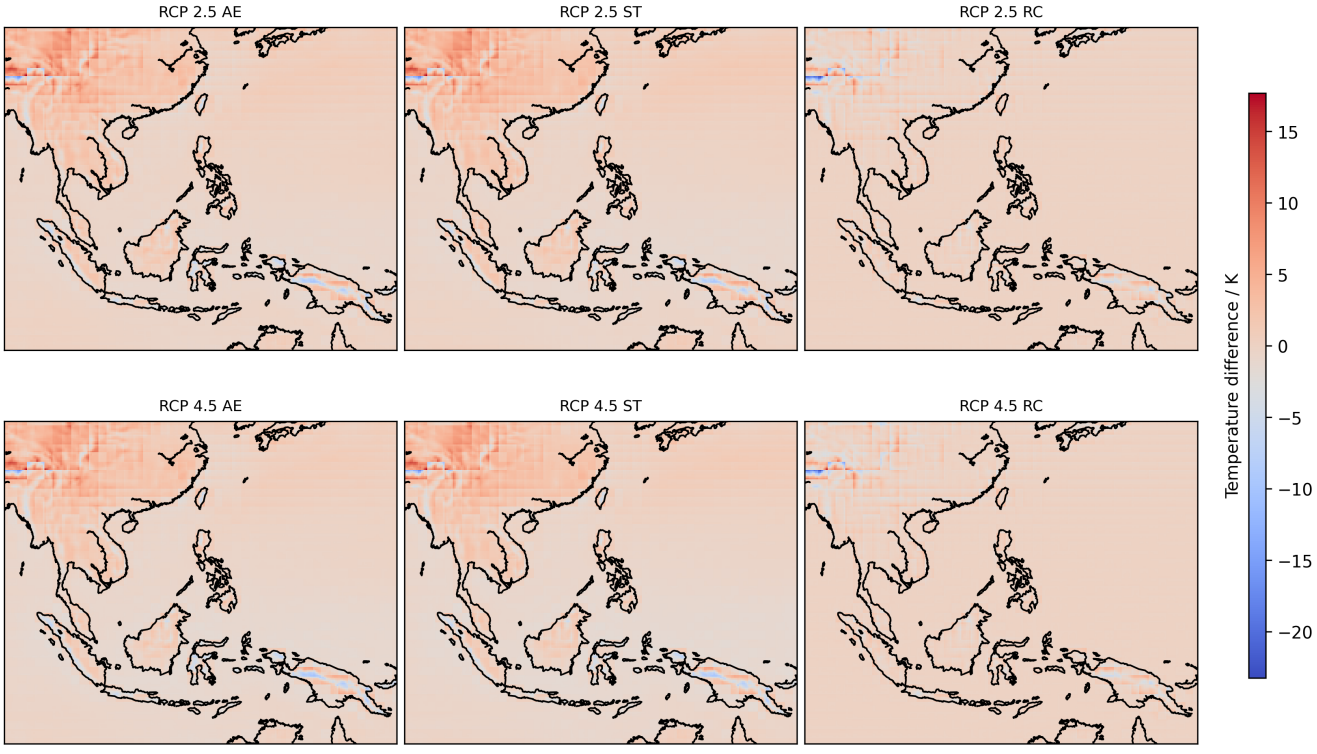

Figure 11. The difference between average daily temperatures of the downscaled outputs (AE, ST and RC) and the input dataset, CMIP5. Results shown are for differences which occur between the years of 2060 and 2100.

4.2.1 RCP 2.6: Are these changes realistic?

RCP 2.6, which represents the RCP with the smallest forcing from greenhouse gasses in the CMIP5 experiments, is the set with temperature values closest to that of the training set. Currently, regional projections for CORDEX are not available for comparison, only CMIP6(SSP125) is available from CMIP6 experiments. The SSP126 scenario of CMIP6 is not an ideal surrogate for CORDEX under RCP 2.6 since it is represents a 'best-case' scenario reaching 2.6 W/m2 by 2100 but at a much slower trajectory (Figure 12). SSP125 reaches equivalent forcing to RCP2.5 past 2060, and so projections in this date range are used for comparison.

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

Temperature profiles: AE, ST and RC all display seasonal cycles very similar to that of both CMIP5 and CMIP6 and no regional bias could be determined. Despite introducing more dynamic gradients in temperature on a spatial scale, daily temperature ranges are again reduced. This means that the smoothing effect of AE and ST is carried forward into the projections.

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

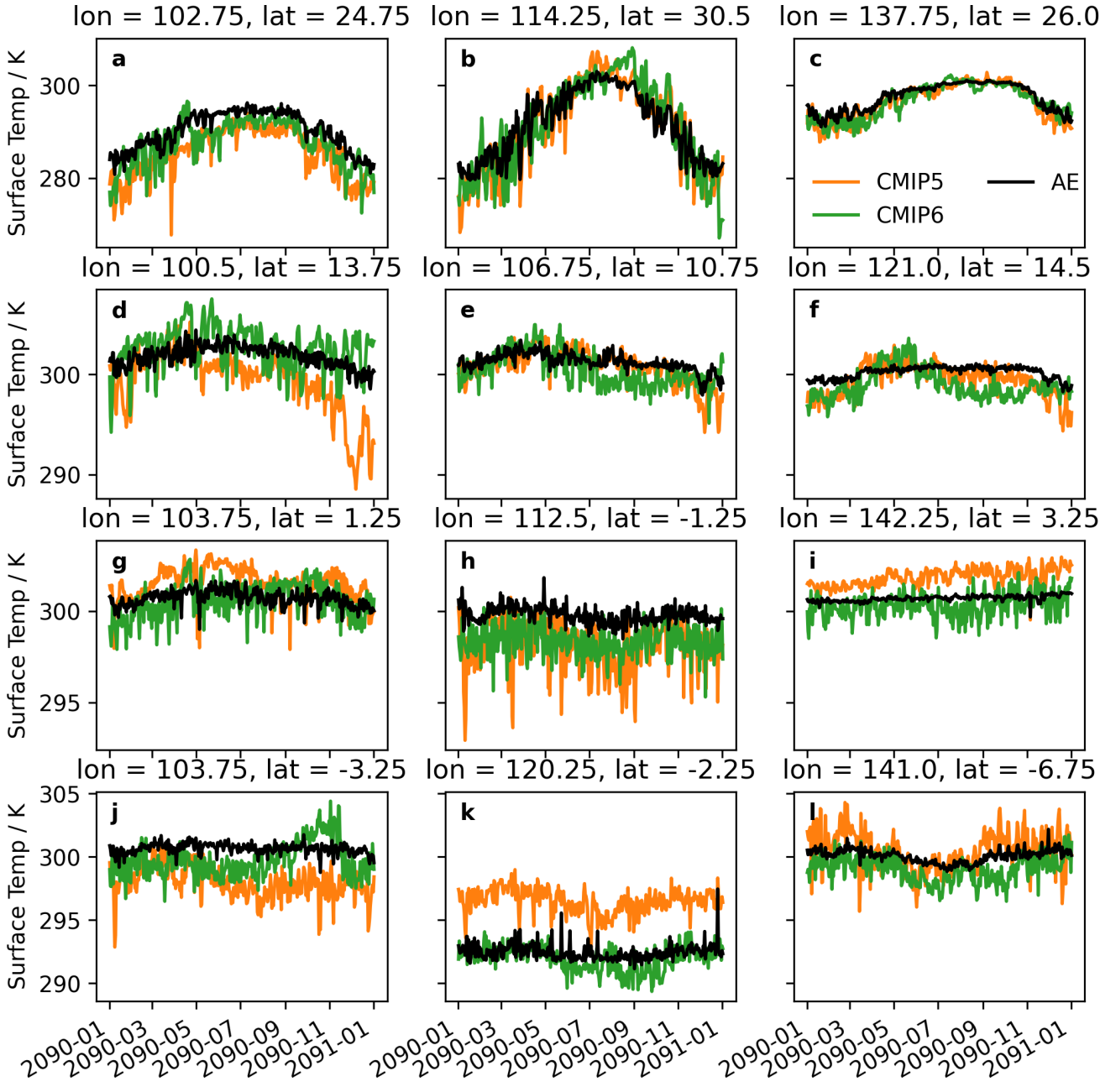

Figure 12. Grid locations compared for the input dataset CMIP5, the autoencoder downscaled output AE and the benchmarking dataset CMIP6 for the year 2090 to 2091. AE sits well between projections made for CMIP5 and CMIP6, however the smoothing effect is also present, reducing daily temperature fluctuations.

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

4.2.2 RCP4.5: Are warming trends realistic?

Downscaled predictions have a reduced rate of warming as compared to the CMIP5 equivalents. The reduction in warming over time means that the differences between the scenarios RCP 2.6 and RCP 4.5 in terms of regional average temperature increase are smaller for the downscaled results. It is not clear why the autoencoder results have a shallower gradient of warming, and if these results are realistic.

Warming due to climate change in the SEA region is not evenly distributed and will be more pronounced in some regions. Comparing regional warming can be used to indicate if the autoencoders are able to project warming correctly. Since the absolute differences in temperature for the autoencoder generated data, RCP and SSP scenarios are dissimilar, qualitative changes can only be compared. As expected, the temperature increase experienced under RCP4.5 or SSP245 as compared to RCP2.6 or SSP126 is not uniform over the region but localised to specific regions (Figure 13). For CMIP6 projections, the difference between scenarios results in a disproportionate increase in average temperature over land, especially within the greater land mass of central China. Additionally, warming at the ocean surface is strongest towards the centre of the Phillipine Sea, and strong changes occur around the Tibetian Plateau which extends into lowland China. A similar warming profile is observed for CMIP5 although less warming is observed over the sea, and reduced warming over landmasses is observed, but this is due to differences in compared scenarios. The downscaled result (Figure 13, AE) does not mirror the changes observed for the other datasets. AE results indicate a warming on the Eastern coast of mainland China which is not present in CMIP6, and significantly reduced in the CMIP5 projections. Additionally, ocean temperatures are cooler away from land masses and unlike the CMIP5 and CMIP6 projections, are cooler towards the centre of the Philippine Sea. Artefacts, such as the 'warm border' present in the lower South-Western region of AE indicate that the autoencoder may be failing to forecast realistic temperatures where features (such as islands or landmasses) are absent. This is why temperatures over the ocean may not change considerably in the downscaled outputs as the autoencoder cannot map spatial patterns where no strong features are present. This is also why islands appear to have a warm outline, which is not present in other projections. This disparity may indicate that this approach may bias warming to regions with topographic features, especially when datasets are limited. This may also explain the plateau observed in the RCP projections, which may be caused by the lack of ocean warming in downscaled results, which constitutes the greatest area of the region. This observed 'patchiness' is not immediately clear from other comparisons and could easily be overlooked since the absolute temperature differences are small and do not contrast sharply in the downscaled outputs.

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

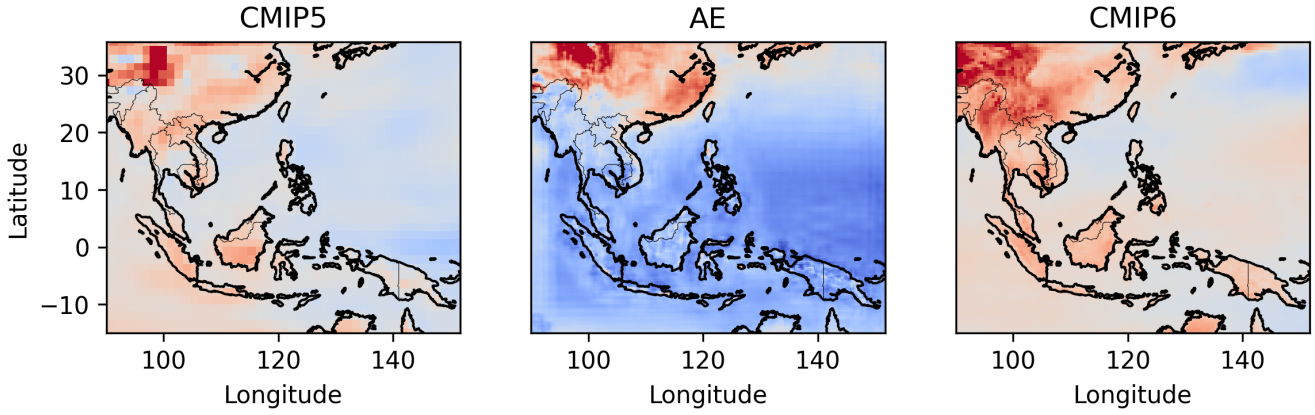

Figure 13. The difference between average surface temperatures between 2080-2101 for RCP 2.6 and RCP 4.5 (CMIP5 and AE) and SSP126 and SSP245 (CMIP6). Maps are enhanced to maximise contrast between temperature changes within the region in order to assess differences in spatial distributions of temperature change only and colours do not represent quantitative differences in temperature which can be inter-compared.

4.2.3 RCP 4.5: Comparison with CORDEX projections

The CORDEX outputs use the same climate model and initialisation parameters as the CMIP5 dataset. The regional average temperatures for CORDEX follow almost the same warming profile as the CMIP5 dataset, features in the CMIP5 are mirrored by features in the CORDEX dataset. CORDEX is cooler over the region, the difference between CORDEX and CMIP5 appears to be constant over the entire temporal range (Figure 14, a). As with RCP 2.5 the RC downscaled data replicates the yearly average temperatures of CMIP5 (data not shown for clarity), whereas AE and ST resemble CMIP5 more in the early years but have the shallowest rate of warming, becoming very similar to the CORDEX outputs in the final 20-30 years. Daily variability in temperatures for the downscaled outputs AE and ST is reduced, as with other studies (Figure 14, b). Daily variability in surface temperatures is expected to increase following downscaling from CMIP5, due to the addition of cooler temperatures at higher elevation. This is observed when comparing the standard deviation in daily average surface temperatures of CMIP5 and CORDEX, as variability becomes narrowly larger. For AE, there is a clear contraction of the standard deviation, and therefore the range of temperature values in the downscaled output. Daily variability in surface temperature vales for CMIP5, CMIP6 and CORDEX is roughly equivalent, so the contraction of AE and ST is likely to be due to a limitation in the autoencoder performance and an unrealistic result.

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

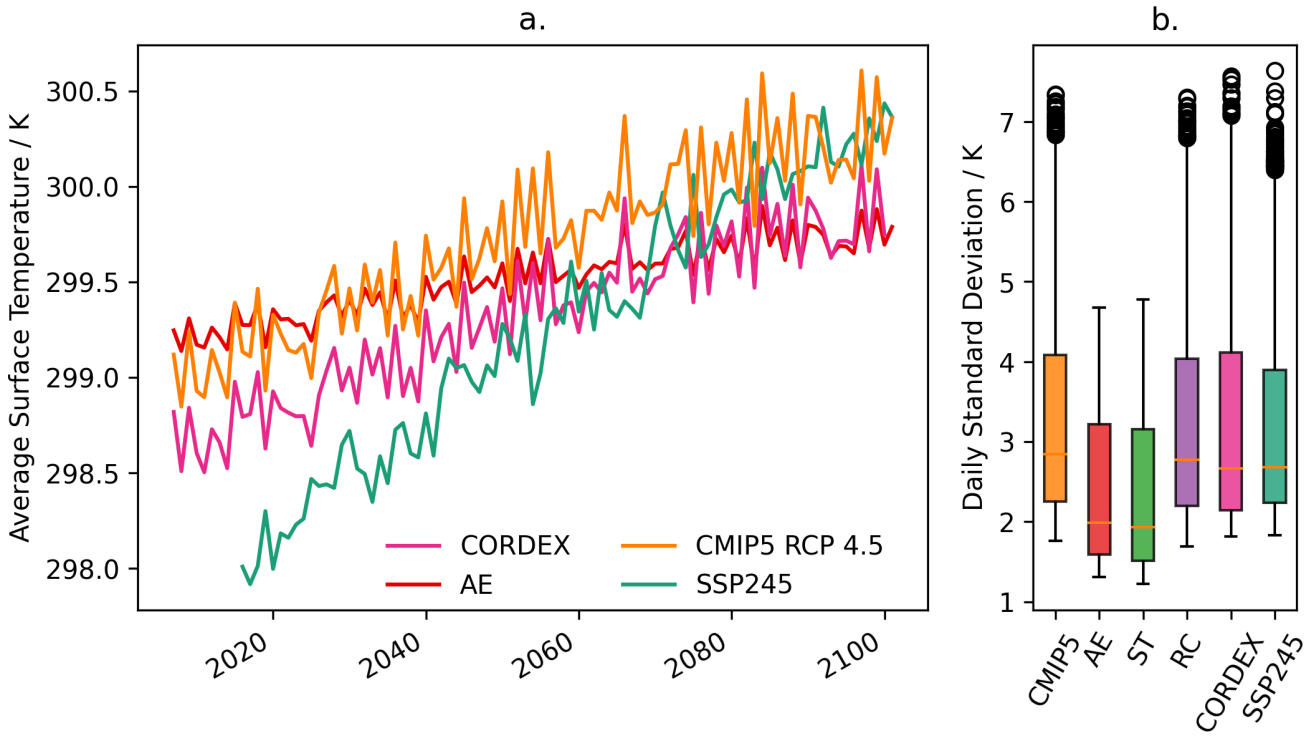

Figure 14. a. Yearly-averaged daily surface temperature for RCP 4.5 and SSP245, averaged over the region defined by the CORDEX experiments. b. Daily temperature variability, presented as standard deviations from daily means.

The spatial features in the AE and ST downscaled outputs appear as well resolved as CORDEX and CMIP6-HR experiments, this is clear around islands, coastlines and mountinous regions. AE and ST are a strong approximation of CORDEX projections (Figure 15), especially in the latter years and spatial features are very similar. There is a step difference between land and sea surface temperatures in the CORDEX datasets and land features appear well resolved against the sea because of this. These differences are not so resolved in the case of AE and ST, but this may be an artefact of dynamic modelling, and may not be realistic.

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

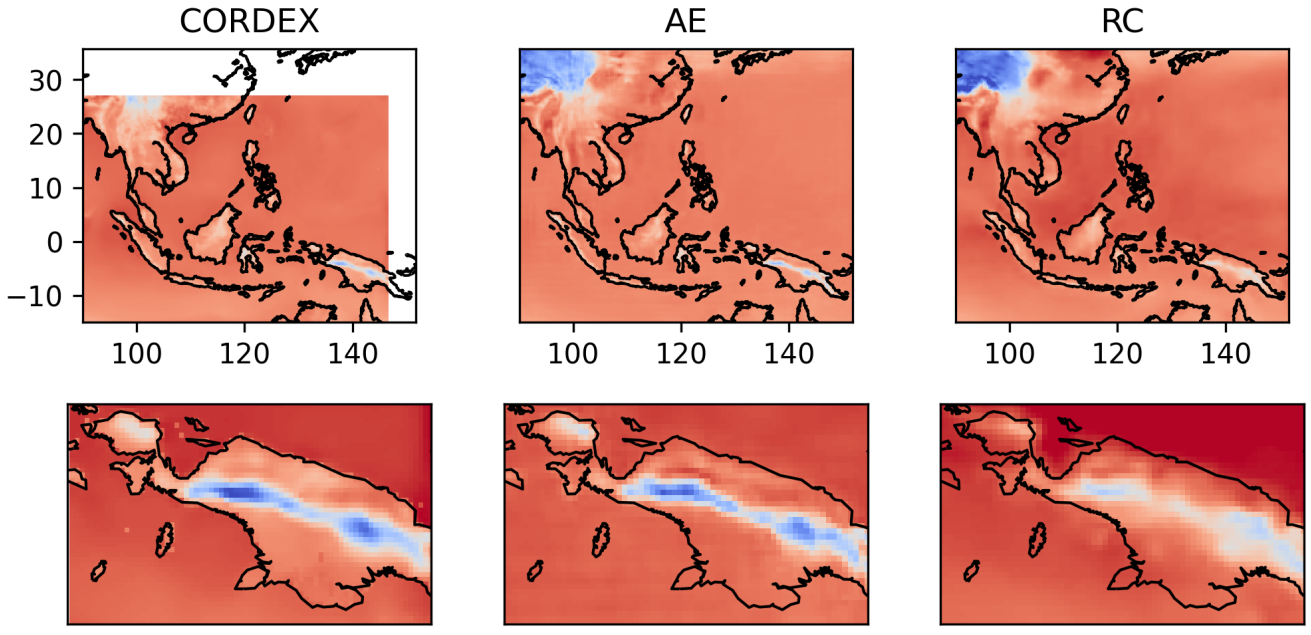

Figure 15. Surface temperatures at 01-07-2099 for CORDEX, AE and CMIP5 for the whole region (top row) and zoomed in (bottom row). Top and bottom row use different temperature scales. This timepoint was selected as regionally averaged surface temperatures were most different at this point for this year.

5 Conclusions

These results show that a deep convolutional autoencoder architecture, utilising limited datasets, is capable of generating realistic high-resolution climate projections at low computational cost. The images are scaled up by a factor > 5, and 'corrected' based on the target observational data. The three approaches employed (AE, ST, RC) demonstrate that despite the similarities in regional average temperatures, the autoencoder 'translation' function of CMIP5 to ERA5-like data can significantly modify the model output and subsequent projections. The translation has the benefit of increasing spatial resolution significantly, and ensuring that projections follow current observational trends. Critically however, the translation results in a 'dampening' of temperature highs and lows which has the effect of artificially stabilising the climate beyond that which could be considered realistic. This is likely due to an inherent limitation of the autoencoder, which is reliant on the identification of reoccurring spatial patterns which can be used to generate 'patches' of higher resolution information. This results in a model with high generalisabilty, but this comes at the cost of temporal resolution. AE and ST behaved similarly and 'stacking' did not lead to any significant improvements despite the doubling of ML layers and separation of functions. Omitting the 'translation' element (RC), temporal cycles are maintained and a moderate increase in spatial resolution is achieved. The inclusion of additional variables such as topography and other atmospheric variables may reduce the magnitude of this dampening effect for AE and ST, but may not eradicate this issue completely. The generation of low probability events, especially ones of an apparent

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

stochastic nature will require a new approach to model accurately, however models which accurately predict time series data are also established. The apparent patchiness in comparing different scenarios also hints at another limitation of this technique. Models perform extremely well in areas with strong topographic features, however, in featureless regions such as large expanses of ocean, model performance is likely to be less reliable. This is because convolutional methods rely on the presence of structure and pattern in a 'image'. The models inability to increase the oceanic surface temperatures away from coastlines can be can be seen as 'borders' which appear around islands when comparing different projections. This effect is not immediately obvious in the downscaled data, and could be overlooked, especially as there is large variability in oceanic temperatures for benchmarking datasets. The patchiness of auto encoder outputs suggest that there may be issues with teleconnection and the maintaining of seasonal regional temperature variations, however seasonal cycles in the region seem to be well maintained. Once issues with temperature anomalies can be rectified, studies which demonstrate that mesoscale atmospheric processes are maintained should be employed. The frequency of April heatwaves, corresponding to El Nino events could be used in the SEA region.

No strategy was identified to eradicate the issue of temperature smoothing by optimising or adding components to the model architecture, and only minor improvements to MSE could be achieved with model optimisation. This suggests that the performance is limited by the features which could be extracted from the training datasets. Halving the date range, and therefore the amount of training data, did not considerably impact the model performance either, suggesting that the model training set was adequate. For this reason, other variables will need to be added to increase performance. Decreasing the 'squeeze' of the autoencoder by reducing the number of max-pooling layers did not improve results. Since the initial CMIP5 dataset is upscaled, the compression of the image using max-pooling layers results in almost no loss of information. This suggests that the encoder function is the more critical component of the autoencoder. However, removing the decoder element results in a large reduction in performance, which is due to the Add layers, which ensure that larger spatial features are carried forward. These studies suggest that the performance cannot be optimised further without adding additional co-variants, or adding new model architectures. Additionally, new cost-functions which penalise reductions in the spread of temperature values might improve training.

Regions which are traditionally difficult to model using RCMs appear to still provide some difficulty for the autoencoder, although uncertainty in this region is high within GCM and RCM models (Gu et al., 2018). Because of the reduction in temperature extremes, the cooler region around the Tibetian plateau appears to be poorly represented. This is a region where additional data (topographic information, atmospheric variables) may considerably improve the model performance and a region where autoencoders may excel due to the presence of strong highly-resolved static features (topography).

Since ML techniques are inherently an emulation of climate, it is important that 'black-box' models are not relied on and simple metrics are not used to demonstrate performance. Since there is no guarantee that physical constraints present in numerical models are maintained following downscaling using ML, some checks are required to ensure that downscaled projections are accurate. It is likely that hybrid approaches for projecting future climate are required, to maximise the benefits of ML approaches whilst forcing model outputs to respect a number of physical constraints (Barbier et al., 2021; Willard et al., 2020; Beucler et al., 2019). Despite the issues highlighted in this paper, clear and established methodologies are available which could potentially mitigate these issues. A number of ML techniques excel in time-series forecasting and climate projection,

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

and the addition of one or two more co-variables would not impact computational time considerably and likely lead to large improvements in performance. Treating issues such as patchiness and the poor performance in 'featureless' ocean regions should be addressed too, as convolutional methods may not be able to solve this problem without interfacing with another methodology.

With these observations in mind, the next stages for developing a model for SEA will take the following steps:

- **Addition of covariants**: Simply adding more covariants such as as topography or additional atmospheric variables to generate surface temperature may increase issues with patchiness and temporal resolution but may not remove them completely. Identifying a minimum set of covariants will be necessary to maintain low computational load. Building on an image reconstruction approach (RC) may result in the best temporally resolved and realistic output.

- **Methodologies to increase precision over the ocean** Featureless areas perform less well, and the autoencoder does not 'correct' these areas. Identifying methodologies to ensure that corrections over the oceans are carried out, and that patchiness is removed will improve regional average results.

- **Identify methodologies to improve temporal resolution**: It is critical to maintain daily temperature ranges, including the frequency of highs and lows. Many ML methodologies have been developed to model time series data, and to identify anomalous events in time series data, but these will need to be adapted to become generative. **Identify methods for demonstrating that teleconnection is maintained** It is important that downscaled climate projections still maintain relationships with other climatic processes in GCMs. This builds confidence that the downscaled projections are realistic.

- **Increase the number of scenarios** In this study the aim is to generate models which emulate climate, for applied modelling. For this reason, the results may be of less interest for those interested in climatic processes and more of interest for those who which to use climate data. In order to increase the utility of this model, strategies to increase the flexibility or number of scenarios available at higher resolution should be developed.

Acknowledgements

We acknowledge the World Climate research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups CNRM-GAME (Centre National de Recherches Météorologiques—Groupe d'études de l'Atmosphère Météorologique) and Cerfacs (Centre Européen de Recherche et de Formation Avancée) for producing and making available their model output for use in this study. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

Hersbach, H. et al., (2018) was downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store.

The results contain modified Copernicus Climate Change Service information 2020. Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains.

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

The downscaled data is a product of the SEACLID/CORDEX Southeast Asia Project which is funded by the Asia-Pacific Networks for Global Change Research (APN) (ARCP2015-04CMY-Tangang)

This research was made possible by research by the IDC-MIT International Design Centre at Singapore University of Technology and Design.

Code availability. Trained models and code can be accessed via:

Levers, Oliver. (2022). Trained models and code accompanying 'Downscaling using Deep Convolutional Autoencoders, a case study for South East Asia' [Data set]. Zenodo. https://doi.org/10.5281/zenodo.6986257

Appendix A: Appendix: Land vs Sea

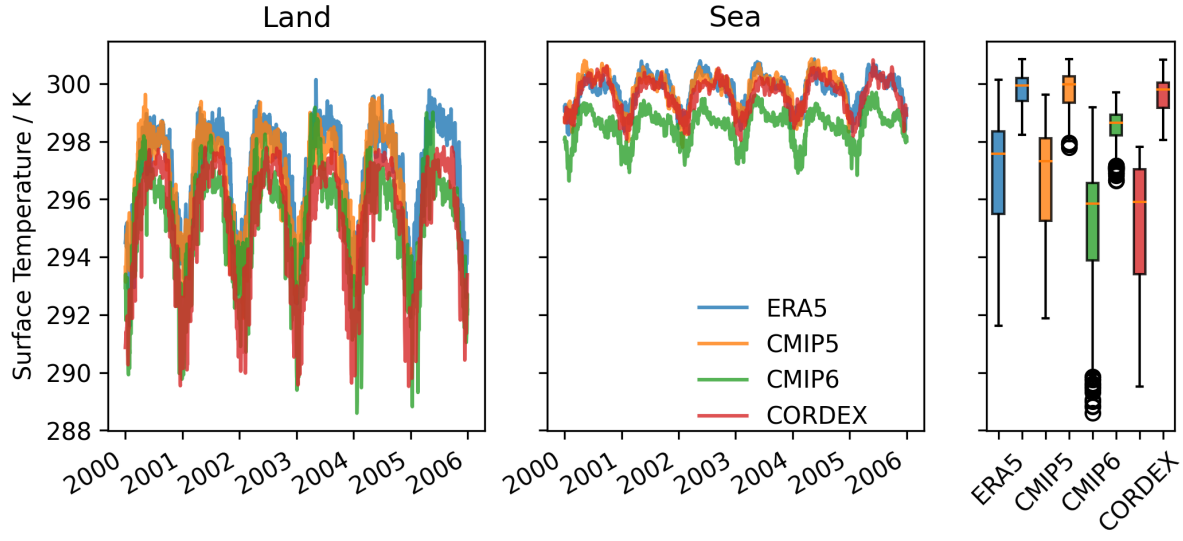

Figure A1. Daily surface temperature averaged over the region. Seasonal temperature cycles are more pronounced on land (left) than in the sea (middle figure). CMIP6 values are cooler in both land and sea values in comparison to ERA5 and CMIP5 (right figure). CORDEX temperature values are similar in the ocean to ERA5 and CMIP5 in the sea, but are far cooler on land.

Appendix B: Appendix: Singapore at RCP 2.6

GCMs are able to generate high temperature predictions caused by physical phenomenon, however ML methods using limited datasets are capable only of generalising, and therefore project a more stable climate. This can be seen in Figure B1) where the stable climate of Singapore is used as an example. ERA5 and CMIP5 have very similar temperature profiles, but dramatic changes in temperature of >1 °C which are not present in both datasets simultaneously are removed by AE. This property

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

of the autoencoder artificially stabilises the climate over the entire region whilst maintaining a statistically likely temperature profile.

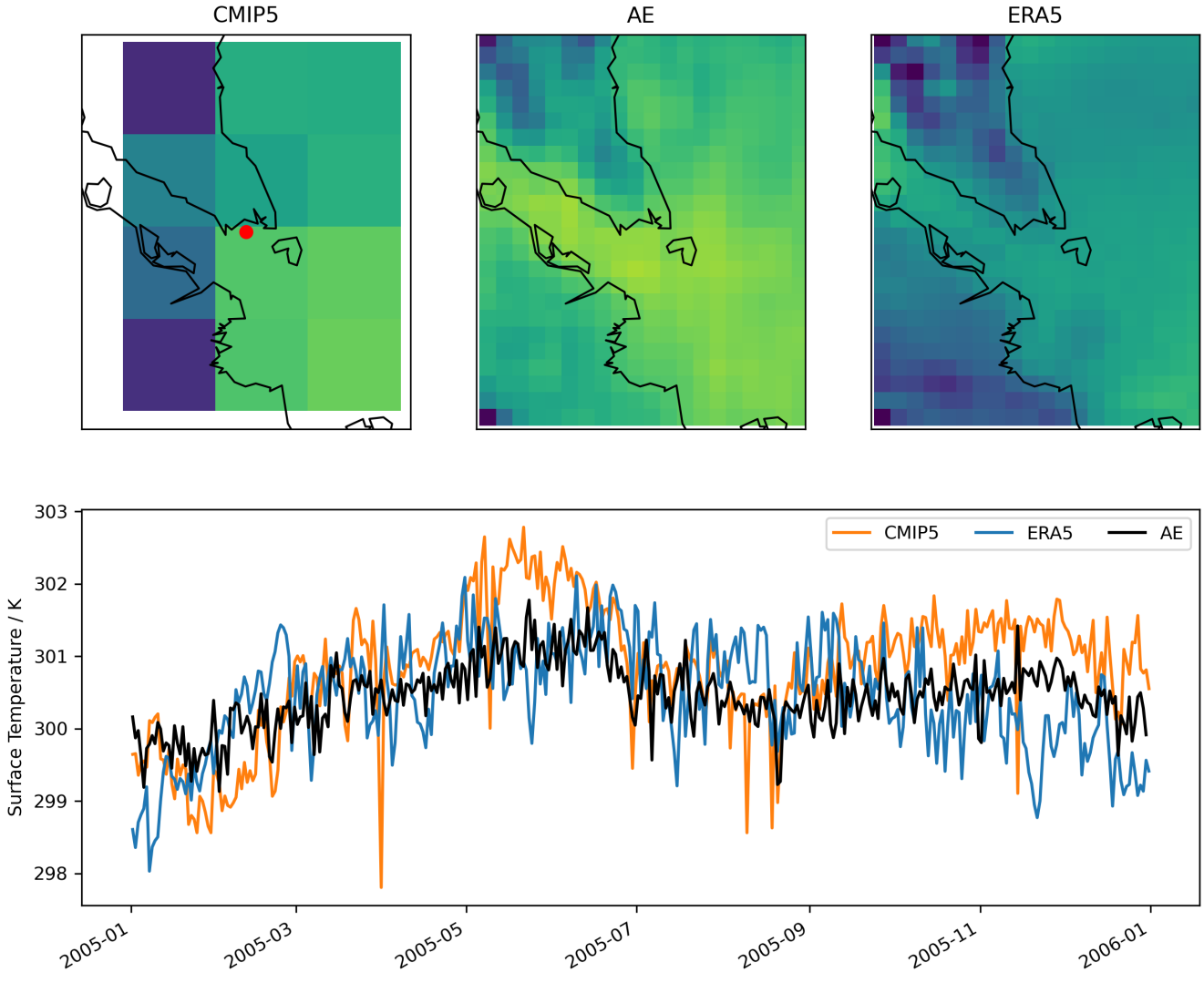

Figure B1. Top row: CMIP5, AE and ERA5 gridded outputs around Singapore, location used to generate lower plot shown as red dot in top left figure. Bottom row: Daily surface temperatures for Singapore between 01-01-2005 to 31-12-2005.

Appendix C: Appendix: China resolution

Mainland china appears to be a region where large changes occur following downscaling and a region which is problematic for traditional downscaling methods. This is due to the complex topography and dramatic increases in altitude present towards the north west of the region. Figure C1 shows how downscaling can generate extreme temperature differences around steep

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

topography which is absent from the CMIP5 experiments. Kunming, Yunnan is surrounded by rugged topographic features and bodies of water which effect local surface temperature gradients, and capturing this information is of high utility for decision makers. However, introducing these temperature gradients can result in significant changes in the projected temperature for individual points on the map. Here, the local temperature for Yunan is increased by almost 4 °C by AE and ST by the addition of complex gradients.

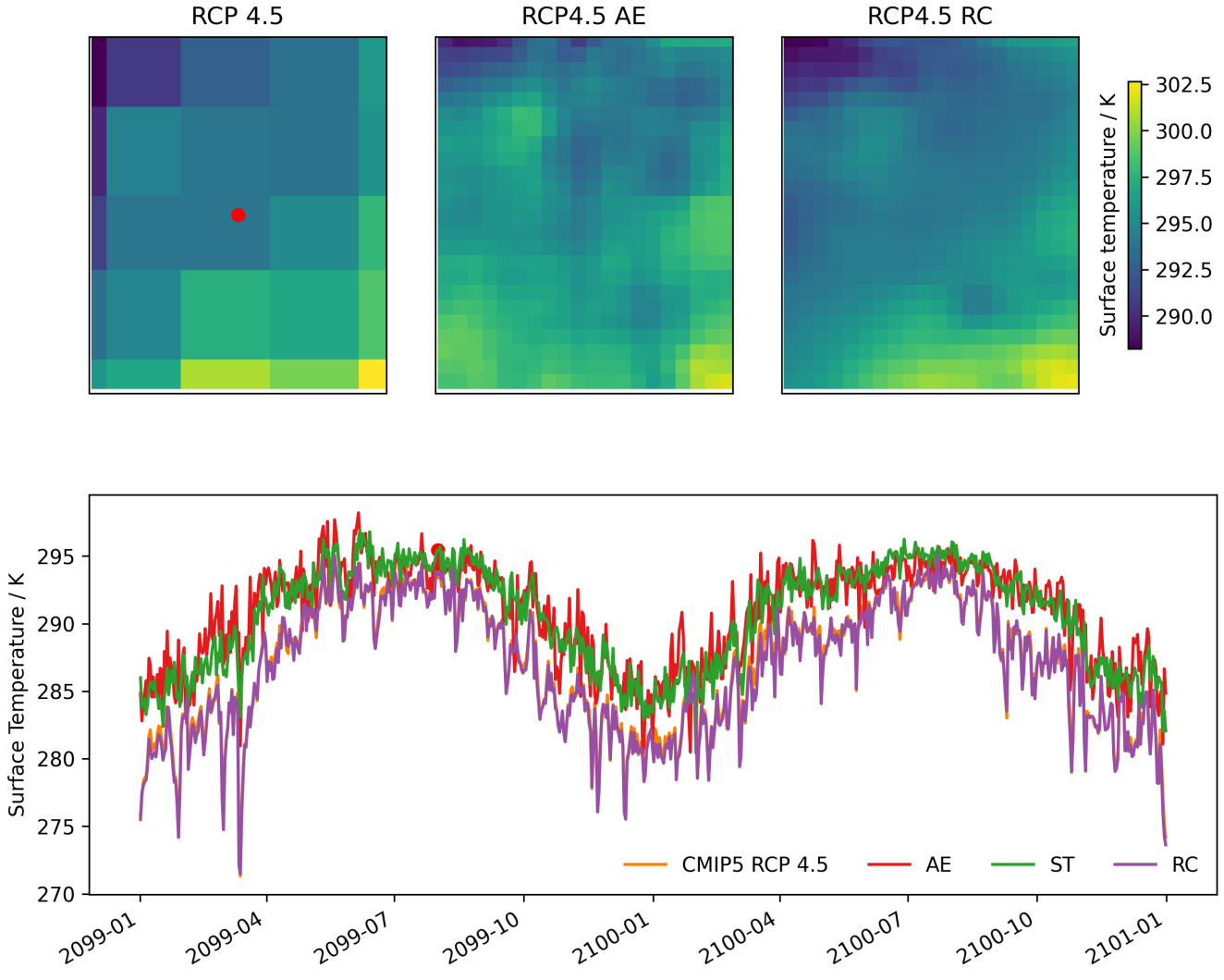

Figure C1. Point a (see Figure 9) located within mainland China where larger differences are recorded between CMIP5 and downscaled outputs. Top row show outputs from CMIP5, AE and RC experiments for the day of 01-09-2099. Lower panel shows the seasonal temperature cylces for the period 01-01-2099 to 01-01-2101. MIGHT BE COOL TO ADD TOPOGRAPHIC LINES AS OVERLAYS TO TOP SECTION

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

Appendix D: Appendix: Bangkok

Increase in spatial resolution: Figure D1 demonstrates the utility of auto-encoder down scaling to identify climatic features not present in CMIP5 forcasts. Bangkok exists in a region of variable topography and is often hotter than the surrounding mountainous areas. The 'hot spot' present around Bangkok, a populous low-lying alluvial region where the Chao Phraya River meets the sea is likely to be strongly affected by climate change and changes in surface temperature. The absence of this spatial information in the CMIP5 projections could result in an underestimate of surface temperatures due to the presence of regions of high topography either side of the city, Additionally, Bangkok exists in the nexus between four large pixels in the CMIP5 dataset so assigning a temperature value to this point is difficult with only CMIP5 data available. The CMIP6 dataset is from the CNRM-CM6-1-HR experiments, which is one of the few higher resolution CMIP6 experiments, so has finer spatial resolution not present in the majority of CMIP6 experiments. Temperature gradients related to the topographic features are resolved in both the CMIP6 and downscaled outputs (AE, ST > RC) and follow similar spatial patterns, indicating that the generated temperature gradient is realistic. Figure D1 also demonstrates the high degree of variability in projected temperatures for areas such as this. Large temperature differences approaching 10 K are present at the end of 2099 where AE and SSP126 are far warmer than the CMIP5 predictions. Daily temperatures are more stable with the AE output, with the daily fluctuations reduced in range as compared to CMIP5 and CMIP6126. This could be problematic since the frequency and magnitude of temperature highs are important measures. Reducing the number of hot days artificially could result in misleading climate projections.

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

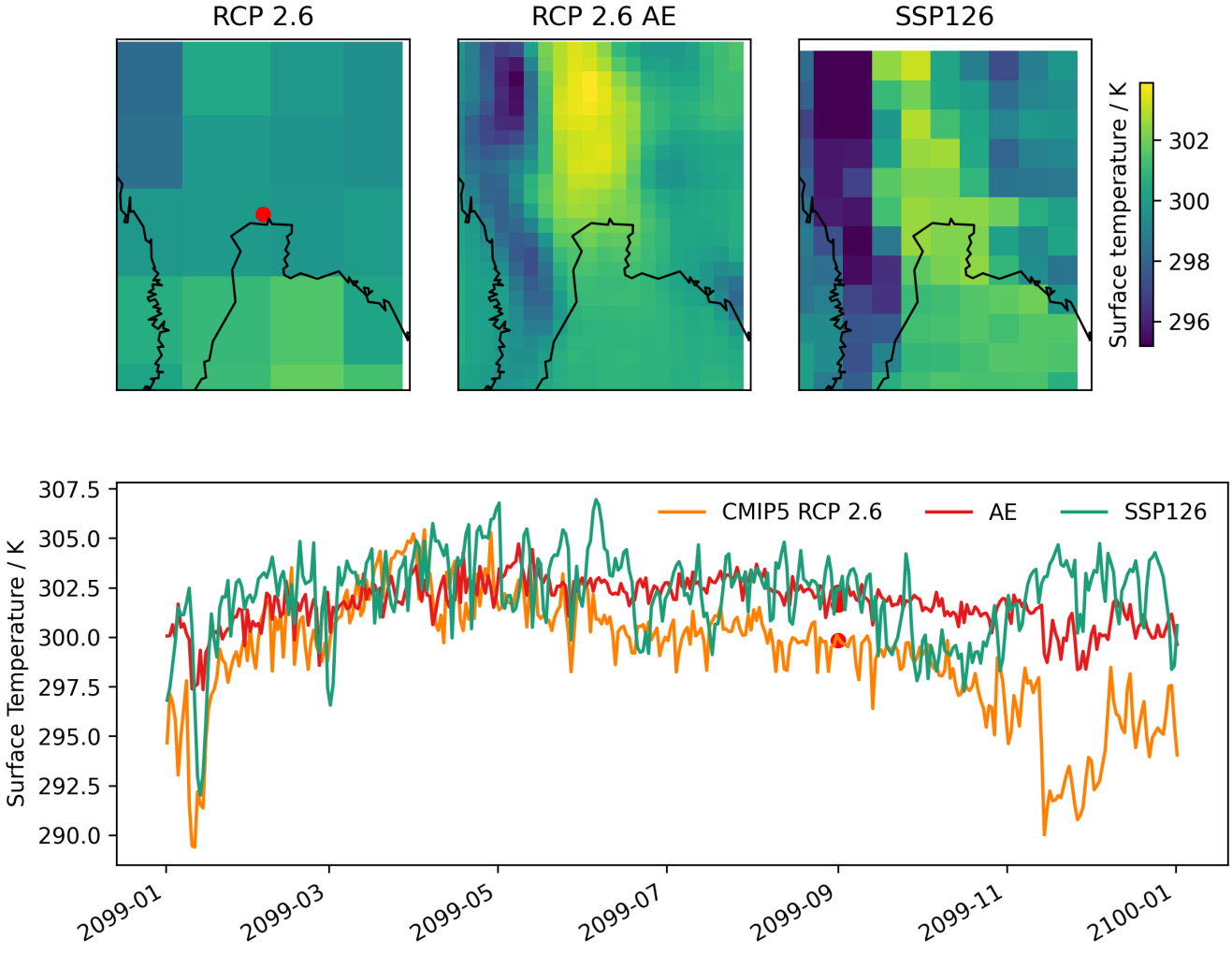

Figure D1. Top row: Maps of temperature distributions around Bangkok, Thailand for 01-01-2080 (date marked on bottom row with red dot). This date is illustrative of a day in which the datasets are disimilar. Bottom row: Temperature profiles for the year 2080 under RCP 2.6 (CMIP5 and AE) and SSP126 (CMIP6).

D1

Author contributions. OL, LB and DH conceived the study, AD was an advisor and helped the development of the study. The experimental design was formulated by OL, LB and DH. OL carried out the analysis, wrote the code and manuscript. DH helped with the coding and computational requirements. All authors contributed to analysing the results and writing, editing of the manuscript.

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

Competing interests. There are no competing interests

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

References

Baño-Medina, J., Manzanas, R., and Gutierrez, J. M.: Configuration and intercomparison of deep learning neural models for statistical downscaling, Geoscientific Model Development, 13, 2109–2124, https://doi.org/10.5194/GMD-13-2109-2020, 2020.

Baño-Medina, J., Manzanas, R., and Gutiérrez, J. M.: On the suitability of deep convolutional neural networks for continental-wide downscaling of climate change projections, Climate Dynamics, 1, 3, https://doi.org/10.1007/s00382-021-05847-0, 2021.

Barbier, R., Le Masson, P., and Weil, B.: Transforming Data into Added-Value Information: The Design of Scientific Measurement Models Through the Lens of Design Theory, in: ProceedingsoftheInternational 1 Conference on Engineering Design (ICED21), 16-20 August, pp. 3239–3248, Gothenburg, Sweden, https://doi.org/10.1017/pds.2021.585, 2021.

Beucler, T., Pritchard, M., Rasp, S., Ott, J., Baldi, P., and Gentine, P.: Enforcing Analytic Constraints in Neural-Networks Emulating Physical Systems, Physical Review Letters, 126, https://doi.org/10.1103/physrevlett.126.098302, 2019.

Chang, Y.-C., Acierto, R., Itaya, T., Akiyuki, K., and Tung, C.-P.: A Deep Learning Approach to Downscaling Precipitation and Temperature over Myanmar, Tech. rep., The University of Tokyo, https://ui.adsabs.harvard.edu/abs/2018EGUGA..20.4120C/abstract, 2018.

Chotamonsak, C., Salathé, E. P., Kreasuwan, J., Chantara, S., and Siriwitayakorn, K.: Projected climate change over Southeast Asia simulated using a WRF regional climate model, ATMOSPHERIC SCIENCE LETTERS Atmos. Sci. Let, 12, 213–219, https://doi.org/10.1002/asl.313, 2011.

Cruz, F. T., Narisma, G. T., Dado, J. B., Singhruck, P., Tangang, F., Linarka, U. A., Wati, T., Juneng, L., Phan-Van, T., Ngo-Duc, T., Santisirisomboon, J., Gunawan, D., and Aldrian, E.: Sensitivity of temperature to physical parameterization schemes of RegCM4 over the CORDEX-Southeast Asia region, International Journal of Climatology, 37, 5139–5153, https://doi.org/10.1002/JOC.5151, 2017.

Ge, F., Zhu, S., Luo, H., Zhi, X., and Wang, H.: Future changes in precipitation extremes over Southeast Asia: Insights from CMIP6 multi-model ensemble, Environmental Research Letters, 16, https://doi.org/10.1088/1748-9326/ABD7AD, 2021.

Gu, H., Yu, Z., Yang, C., Ju, Q., Yang, T., and Zhang, D.: High-resolution ensemble projections and uncertainty assessment of regional climate change over China in CORDEX East Asia, Hydrology and Earth System Sciences, 22, 3087–3103, https://doi.org/10.5194/HESS-22-3087-2018, 2018.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N.: ERA5 hourly data on single levels from 1979 to present., https://doi.org/10.24381/cds.adbb2d47, (Accessed on 17-NOV-2020).

Ji, Y., Zhi, X., Tian, Y., Peng, T., Huo, Z., Ji, L., Ji, Y., Zhi, X., Tian, Y., Peng, T., Huo, Z., and Ji, L.: Downscaling of Precipitation Forecasts Based on Single Image Super-Resolution, EGUGA, p. 8533, https://ui.adsabs.harvard.edu/abs/2020EGUGA..22.8533J/abstract, 2020.

Johnson, J., Alahi, A., and Fei-Fei, L.: Perceptual Losses for Real-Time Style Transfer and Super-Resolution, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9906 LNCS, 694–711, https://arxiv.org/abs/1603.08155v1, 2016.

Juneng, L., Tangang, F., Chung, J. X., Ngai, S. T., Tay, T. W., Narisma, G., Cruz, F., Phan-Van, T., Ngo-Duc, T., Santisirisomboon, J., Singhruck, P., Gunawan, D., and Aldrian, E.: Sensitivity of Southeast Asia rainfall simulations to cumulus and air-sea flux parameterizations in RegCM4, Climate Research, 69, 59–77, https://doi.org/10.3354/cr01386, 2016a.

Juneng, L., Tangang, F., Chung, J. X., Ngai, S. T., Tay, T. W., Narisma, G., Cruz, F., Phan-Van, T., Ngo-Duc, T., Santisirisomboon, J., Singhruck, P., Gunawan, D., and Aldrian, E.: Sensitivity of Southeast Asia rainfall simulations to cumulus and air-sea flux parameterizations in RegCM4, Climate Research, 69, 59–77, https://doi.org/10.3354/CR01386, 2016b.

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

Kamworapan, S. and Surussavadee, C.: Performance of CMIP5 global climate models for climate simulation in Southeast Asia, in: IEEE Region 10 Annual International Conference, Proceedings/TENCON, vol. 2017-Decem, pp. 718–722, Institute of Electrical and Electronics Engineers Inc., https://doi.org/10.1109/TENCON.2017.8227954, 2017.

Kamworapan, S. and Surussavadee, C.: Evaluation of CMIP5 Global Climate Models for Simulating Climatological Temperature and Precipitation for Southeast Asia, Advances in Meteorology, 2019, 1067 365, https://doi.org/10.1155/2019/1067365, 2019.

Lee, J.-W., Hong, S.-Y., Chang, E.-C., Suh, M.-S., and Kang, H.-S.: Assessment of future climate change over East Asia due to the RCP scenarios downscaled by GRIMs-RMP, Climate Dynamics 2013 42:3, 42, 733–747, https://doi.org/10.1007/S00382-013-1841-6, 2013.

Mansfield, L. A., Nowack, P. J., Kasoar, M., Everitt, R. G., Collins, W. J., and Voulgarakis, A.: Predicting global patterns of long-term climate change from short-term simulations using machine learning, npj Climate and Atmospheric Science, 3, https://doi.org/10.1038/s41612-020-00148-5, 2020.

McSweeney, C. F., Jones, R. G., Lee, R. W., and Rowell, D. P.: Selecting CMIP5 GCMs for downscaling over multiple regions, Climate Dynamics, 44, 3237–3260, https://doi.org/10.1007/s00382-014-2418-8, 2015.

Navarro-Racines, C., Tarapues, J., Thornton, P., Jarvis, A., and Ramirez-Villegas, J.: High-resolution and bias-corrected CMIP5 projections for climate change impact assessments, Scientific Data 2020 7:1, 7, 1–14, https://doi.org/10.1038/s41597-019-0343-8, 2020.

Ngo-Duc, T., Tangang, F. T., Santisirisomboon, J., Cruz, F., Trinh-Tuan, L., Nguyen-Xuan, T., Phan-Van, T., Juneng, L., Narisma, G., Singhruck, P., Gunawan, D., and Aldrian, E.: Performance evaluation of RegCM4 in simulating extreme rainfall and temperature indices over the CORDEX-Southeast Asia region, International Journal of Climatology, 37, 1634–1647, https://doi.org/10.1002/JOC.4803, 2017.

Raghavan, S. V., Liu, J., Nguyen, N. S., Vu, M. T., and Liong, S. Y.: Assessment of CMIP5 historical simulations of rainfall over Southeast Asia, Theoretical and Applied Climatology, 132, 989–1002, https://doi.org/10.1007/s00704-017-2111-z, 2018.

Sachindra, D. A., Ahmed, K., Rashid, M. M., Shahid, S., and Perera, B. J.: Statistical downscaling of precipitation using machine learning techniques, Atmospheric Research, 212, 240–258, https://doi.org/10.1016/j.atmosres.2018.05.022, 2018.

Schär, C., Fuhrer, O., Arteaga, A., Ban, N., Charpilloz, C., Girolamo, S. D., Hentgen, L., Hoefler, T., Lapillonne, X., Leutwyler, D., Osterried, K., Panosetti, D., Rüdisühli, S., Schlemmer, L., Schulthess, T. C., Sprenger, M., Ubbiali, S., and Wernli, H.: Kilometer-Scale Climate Models: Prospects and Challenges, Bulletin of the American Meteorological Society, 101, E567–E587, https://doi.org/10.1175/BAMS-D-18-0167.1, 2020.

Tangang, F., Supari, S., Chung, J. X., Cruz, F., Salimun, E., Ngai, S. T., Juneng, L., Santisirisomboon, J., Santisirisomboon, J., Ngo-Duc, T., Phan-Van, T., Narisma, G., Singhruck, P., Gunawan, D., Aldrian, E., Sopaheluwakan, A., Nikulin, G., Yang, H., Remedio, A. R. C., Sein, D., and Hein-Griggs, D.: Future changes in annual precipitation extremes over Southeast Asia under global warming of 2°C, APN Science Bulletin, 8, 3–8, https://doi.org/10.30852/SB.2018.436, 2018.

Tangang, F., Chung, J. X., Juneng, L., Supari, Salimun, E., Ngai, S. T., Jamaluddin, A. F., Mohd, M. S. F., Cruz, F., Narisma, G., Santisirisomboon, J., Ngo-Duc, T., Van Tan, P., Singhruck, P., Gunawan, D., Aldrian, E., Sopaheluwakan, A., Grigory, N., Remedio, A. R. C., Sein, D. V., Hein-Griggs, D., McGregor, J. L., Yang, H., Sasaki, H., and Kumar, P.: Projected future changes in rainfall in Southeast Asia based on CORDEX–SEA multi-model simulations, Climate Dynamics 2020 55:5, 55, 1247–1267, https://doi.org/10.1007/S00382-020-05322-2, 2020.

Thirumalai, K., DInezio, P. N., Okumura, Y., and Deser, C.: Extreme temperatures in Southeast Asia caused by El Ninõ and worsened by global warming, Nature Communications, 8, https://doi.org/10.1038/ncomms15531, 2017.

https://doi.org/10.5194/egusphere-2022-234
Preprint. Discussion started: 19 August 2022

Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., and Ganguly, A. R.: Generating high resolution climate change projections through single image super-resolution: An abridged version, IJCAI International Joint Conference on Artificial Intelligence, 2018-July, 5389–5393, https://doi.org/10.24963/IJCAI.2018/759, 2018.

Voldoire, A.: CNRM-CERFACS CNRM-CM6-1-HR model output prepared for CMIP6 HighResMIP, https://doi.org/10.22033/ESGF/CMIP6.1387, 2019.

Voldoire, A., Sanchez-Gomez, E., Salas y Mélia, D., Decharme, B., Cassou, C., Sénési, S., Valcke, S., Beau, I., Alias, A., Chevallier, M., Déqué, M., Deshayes, J., Douville, H., Fernandez, E., Madec, G., Maisonnave, E., Moine, M.-P., Planton, S., Saint-Martin, D., Szopa, S., Tyteca, S., Alkama, R., Belamari, S., Braun, A., Coquart, L., and Chauvin, F.: The CNRM-CM5.1 global climate model: description and basic evaluation, Climate Dynamics 2012 40:9, 40, 2091–2121, https://doi.org/10.1007/S00382-011-1259-Y, 2012.

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., Colin, J., Guérémy, J. F., Michou, M., Moine, M. P., Nabat, P., Roehrig, R., Salas y Mélia, D., Séférian, R., Valcke, S., Beau, I., Belamari, S., Berthet, S., Cassou, C., Cattiaux, J., Deshayes, J., Douville, H., Ethé, C., Franchistéguy, L., Geoffroy, O., Lévy, C., Madec, G., Meurdesoif, Y., Msadek, R., Ribes, A., Sanchez-Gomez, E., Terray, L., and Waldman, R.: Evaluation of CMIP6 DECK Experiments With CNRM-CM6-1, Journal of Advances in Modeling Earth Systems, 11, 2177–2213, https://doi.org/10.1029/2019MS001683, 2019.

Warner, T. T., Peterson, R. A., and Treadon, R. E.: A Tutorial on Lateral Boundary Conditions as a Basic and Potentially Serious Limitation to Regional Numerical Weather Prediction, Bulletin of the American Meteorological Society, 78, 2599–2617, https://doi.org/10.1175/1520-0477(1997)078<2599:ATOLBC>2.0.CO;2, 1997.

Willard, J., Jia, X., Xu, S., Steinbach, M., and Kumar, V.: Integrating Physics-Based Modeling With Machine Learning : A Survey, 1, 1–34, 2020.

Xu, R., Chen, N., Chen, Y., and Chen, Z.: Downscaling and Projection of Multi-CMIP5 Precipitation Using Machine Learning Methods in the Upper Han River Basin, Advances in Meteorology, 2020, https://doi.org/10.1155/2020/8680436, 2020.