Towards Data-Driven Physics-Informed Global Precipitation Forecasting from Satellite Imagery

Valentina Zantedeschi
GE Global Research

Daniele De Martini
University of Oxford

Catherine Tong
University of Oxford

Christian Schroeder de Witt
University of Oxford

Alfredo Kalaitzis
University of Oxford

Piotr Biliński
University of Warsaw

Matthew Chantry
University of Oxford

Duncan Watson-Parris
University of Oxford

Abstract

Under the effects of global warming, extreme events such as floods and droughts are increasing in frequency and intensity. This trend directly affects communities and make all the more urgent widening the access to accurate precipitation forecasting systems for disaster preparedness. Nowadays, weather forecasting relies on numerical models necessitating massive computing resources that most developing countries cannot afford. Machine learning approaches are still in their infancy but already show the promise for democratizing weather predictions, by leveraging any data source and requiring less compute. In this work, we propose a methodology for data-driven and physics-aware global precipitation forecasting from satellite imagery. To fully take advantage of the available data, we design the system as three elements: 1. The atmospheric state is estimated from recent satellite data. 2. The atmospheric state is propagated forward in time. 3. The atmospheric state is used to derive the precipitation intensity within a nearby time interval. In particular, our use of stochastic methods for forecasting the atmospheric state represents a novel application in this domain.

1 Data-driven precipitation forecasting

Climate change is already affecting precipitation events and how water is distributed over the planet. A warmer atmosphere has increased capacity at holding water vapor, resulting in increasing the frequency and intensity of heavy rainfall [6] and at the same time in draining water reserves. Direct consequences of this trend are the saturation of watersheds and sewer systems, with higher risks of flooding and landslides, and of polluting waters. This is particularly problematic for urban areas, but also for agriculture and aquaculture, endangering livelihoods and causing severe economic loss. Being able to predict these extreme events several days in advance would allow for mitigating their impact.

However, currently deployed operational weather forecasting systems are based on numerical models with computing and data demands that most developing countries cannot afford. Furthermore, they have a prediction time lag of several hours[17], which delays prompt disaster response. Given the amount of Earth Observational data and numerical model outputs available nowadays, machine learning approaches have the potential to match operational systems' performance and significantly reduce their compute needs and their inference time lags. Yet, due to the scale and complexity of the
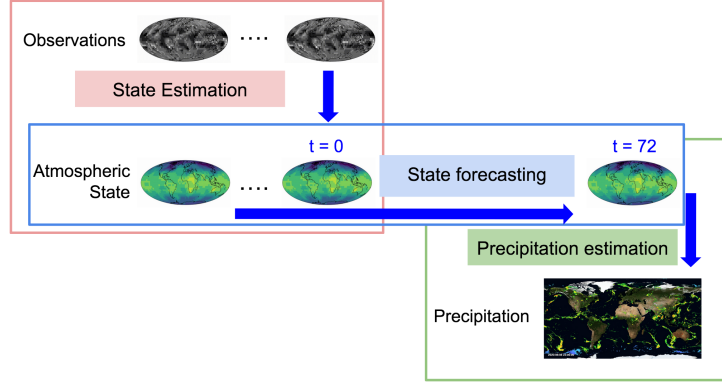
Figure 1: System overview. The three steps are highlighted by different colours. Each step can be viewed as a self-contained task, to be finally linked and fine-tuned together.

problem, state-of-the-art approaches in this field are limited to nowcasting (forecasting a few hours in advance), or to local geographical areas.

In this paper, we propose a global, physics-informed and probabilistic Deep Learning system for predicting precipitation rates up to three days ahead of time. The global scale is enabled by the use of satellite imagery and allows for surpassing the nowcasting time barrier. Physics knowledge is implicitly injected by processing a numerical model product (ERA5 [7]) at training as an intermediate representation. Due to data availability constraints elucidated in Section 2, we decouple the system into three steps and provide preliminary results for each of them.

Related Work.   Our work is motivated by three areas of related work. First, our work adds to the domain of machine-learning-based precipitation forecasting, where a common approach is to use Recurrent Neural Networks for precipitation nowcasting (i.e. forecasting few hours ahead) [20]. Agrawal et al. [1] proposed a U-Net-based nowcasting model from radar images, while Sønderby et al. [17] proposes a network with spatial and temporal encoders to predict from both radar and satellite images, although on a limited regional area. Another related area is physics-informed machine learning, where a common approach is to incorporate physics constraints as an additional loss term [4]. In this work, we instead emulate the output of a physics-based model, an approach also seen in [12]. Lastly, our work incorporates probabilistic forecasts; relevant techniques can be found in stochastic video prediction [3, 9]. In precipitation nowcasting, [17] is an example of a probabilistic model, where uncertainty is modeled by outputting precipitation histograms.

2   A physics-informed and probabilistic approach

To enable medium-range and global predictions, we incorporate physics knowledge into the precipitation forecasting system by leveraging the reanalysis dataset ERA5. This product is generated by a numerical model and provides a range of atmospheric-state variables appropriate for our precipitation estimation, e.g. specific humidity, temperature and geopotential height at different pressure levels [7]. By using this data source, we encourage the system to implicitly model physical laws. We further make use of SimSat [2] which contains simulated[1] satellite imagery from European Centre for Medium-Range Weather Forecasts (ECMWF) at 3 h frequency for three spectral channels. Finally, as precipitation ground truth, we consider IMERG, a global half-hourly precipitation estimation product provided by NASA [8]. More details about each dataset can be found in Appendix A.

Directly forecasting precipitation is an intrinsically difficult task, as these three data sources are not available at the same times. Training an end-to-end system, with all datasets simultaneously, would limit the data availability to less than 5 years, which would consequently limit the capacity of the trained model. We propose instead a three-step approach: 1) *State Estimation*: the weather satellite imagery is processed to infer the atmospheric state of the Earth, 2) *State Forecasting*: a sequence

---

[1]Using simulated satellite data in place of real ones minimizes data processing, as images are a global nadir view of Earth, avoiding issues of instrument error and large numbers of missing values.

Table 1: State estimation RMSE at $t = 0$. Best results are in bold.

	t-850 (K)	q-500 (kg kg^{-1})
Persistence	3.0941	0.8781
Climatology	13.744	1.286
Simsat	2.2289	0.4875
Simsat, Ground obs.	2.4499	0.5246
Simsat, Ground obs., ERA5	**1.8145**	**0.4821**

Table 2: Precipitation estimation F1 scores. Best results are in bold.

Class	FCNN	ERA5 tp
No-precip	**0.9607**	0.9572
Drizzle	0.2997	**0.4085**
Light	**0.2114**	0.1784
Heavy	**0.0641**	0.0632

of atmospheric states is analyzed to move the current state forward in time and 3) *Precipitation Estimation*: the predicted state is then used to estimate the precipitation probability over the globe. This approach has the advantage of leveraging all available data, as each step is tackled independently at first, allowing to train larger models that are able to capture all different dynamics. This also allows for easier comparison with existing baselines. Figure 1 shows an overview of the system.

2.1 State Estimation

The first aspect of the system extracts from weather-satellite imagery a compact representation of the state of the atmosphere, which in essence is a *data assimilation* problem. We formulate this as a regression task modeled by Recurrent Neural Networks. The network takes a historic sequence of **SimSat** data as input (in regular intervals of Δt from $t = -T$ to $t = 0$), and outputs the atmospheric state vector at $t = 0$. The atmospheric state vector consists of 17 ERA5 variables of value to the downstream precipitation prediction problem[2]. Amongst these variables, we report the prediction results on temperature at $850\,\mathrm{hPa}$ (`t-850`) and specific humidity at $500\,\mathrm{hPa}$ (`q-500`).

Other than **SimSat** data, other forms of data, which are conducive to the state estimation task, can be expected to be available as input to the network at $t = 0$. This includes readily-available near-ground observations such as `sp` and `t2m`, or a past atmospheric state vector up to 24 hours before. Therefore, we consider three alternative setups for depending on input data availability: **SimSat only**, **SimSat and ground observations** and **SimSat, ground observations and historical atmospheric states**. Additional details about each setup are provided in Appendix B.

For this task, we train a Convolutional LSTM model [20] and minimize the latitude-weighted Mean Squared Error [15] as loss function. As baselines, we consider *Persistence* (i.e. using 1-day old ERA5 variable values as prediction at $t = 0$), and *Climatology* (i.e. using the training set mean as prediction for each variable). We process all data sources at $5.625°$ resolution and use 2016-2017 time interval for training, 2018 for validation, and 2019 for test. In Table 1, we see that all three setups are able to reconstruct the state variables at $t = 0$, outperforming the considered baselines. As expected, **SimSat, ground observations and historical atmospheric states** setup achieves the best performance, most noticeably for `t-850`. We also observe an encouraging result for **SimSat only**, which predicts `q-850` similarly to the third setup.

2.2 State Forecasting

The second part of the system propagates the atmospheric state (the ERA5 variables predicted at step 2.1) forward in time. Using ERA5 as training data will implicitly allow the model to emulate the physical dynamics of the atmosphere. This ability to probe the weather state provides a route to surpass the forecast horizon of MetNet [17]. In order to handle the inherent stochasticity of the atmospheric system, we implement a *stochastic video prediction* (SVG) [13] model. Specifically, we base our solution on the model proposed in [5], which tackles the problem of predicting the most likely futures by incorporating the usage of a Variational Auto-Encoders (VAE) for sampling different, sharp futures from a probability distribution learned from the data itself. To improve

---

[2]We have chosen these through empirical correlation analysis as well as domain expert advice
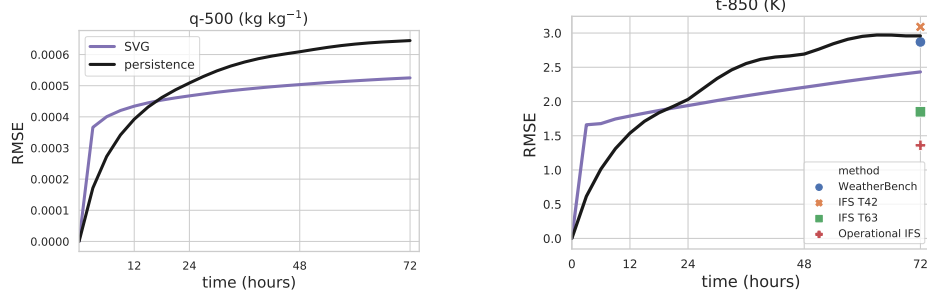
Figure 2: State forecasting RMSE as a function of lead time. `q-500` and `t-850` are given in units of kg kg^{-1} and K respectively.

prediction sharpness, we modify the architecture with the addition of adversarial losses to the learning process, inspired by [19, 9, 11]. To the best of our knowledge, this is the first attempt to forecast the atmospheric state with a probabilistic machine learning model.

For this task, we split the ERA5 dataset at $5.625°$ resolution into training/validation/test sets using the following time intervals: for training, 2010 to 2016 inclusive; for validation, 2017 and 2018; for test, 2019. Inputs consist of $12\,\mathrm{h}$-long time series, at a $3\,\mathrm{h}$ frequency. The model outputs a full state, three hours in the future; the prediction is carried out iteratively up to $72\,\mathrm{h}$. Figure 2 reports the latitude-weighted Root Mean Square Error of `t-850` and `q-500` as a function of lead time. We compare results on both variables to *Persistence* (at $t = 0$) and report for `t-850` the performance achieved by *Weatherbench* and the Integrated Forecast System (IFS) *IFS* from ECMWF on this task (available in [15]). The results show that the methodology has great potential since it performs comparably to the other methodologies in the field.

2.3 Precipitation Estimation

The final step consists in estimating precipitation intensity given the predicted state of the atmosphere. For this task, we make use of IMERG as target ground-truth and of ERA5's variables predicted in the previous steps. Both datasets are retrieved at $0.25°$ resolution. We implement a gridcell-wise Fully Connected Neural Network (FCNN) as global or regional contexts are not required when estimating the precipitation at the current time. As precipitation intensity has an extremely skewed distribution, we convert the problem to a classification task by defining four classes of increasing precipitation intensity accumulated over three hours (in $\mathrm{mm\,h^{-1}}$, No-precip $[0, 1]$, Drizzle $]1, 7.5]$, Light $]7.5, 22.8]$, Heavy > 22.8) and by sampling data equally from each class. We further optimize the Focal Loss [10]. More details are provided in Appendix D.

For this task, we split the ERA5 and IMERG datasets into training/validation/test sets using the following time intervals: for training, 2010 to 2016 inclusive; for validation, 2017 and 2018; for testing, 2019. We pre-process the target values by accumulating precipitations over $3\,\mathrm{h}$ periods; this is done with respect to the frequency of atmospheric state estimation in the second step, such that a future harmonization of the three steps can be done with ease. We report in Table 2 the obtained F1 scores on the test set compared to ERA5's *total precipitation* (tp) baseline. Performance is assessed by per-class F1 scores in order to equally study precision and recall, and to highlight results over rainy classes. Overall, the FCNN is able to achieve performance similar to ERA's one, and it even outperforms this numerical model on all classes but drizzle. However, both models show limited predictive capabilities on minority classes. We conjecture that FCNN's performance could be improved by (i) extending the input feature set, (ii) carefully tuning the class weights and proportions, and (iii) extending the training set.

3 Conclusion

We have designed a three-step approach to weather forecasting and assessed the performance of each step in isolation. The next step is to couple all three components together to produce an end-to-end forecasting system. We envisage this will require a further step of fine-tuning to harmonize the

4

models. In particular, the state forecasting component has a tendency to produce blurred images when forecasting for longer times, as commonly observed with state forecasting [15]. Future work will balance the adversarial loss importance to produce sharp images throughout the prediction window. This will be key as the precipitation estimate requires sharp images of the atmospheric state.

Acknowledgements

References

[1] Shreya Agrawal, Luke Barrington, Carla Bromberg, John Burge, Cenk Gazen, and Jason Hickey. Machine learning for precipitation nowcasting from radar images. *arXiv preprint arXiv:1912.12132*, 2019.

[2] DMA Aminou. Msg's seviri instrument. *ESA Bulletin (0376-4265)*, (111):15–17, 2002.

[3] Mohammad Babaeizadeh, Chelsea Finn, Dumitru Erhan, Roy H. Campbell, and Sergey Levine. Stochastic Variational Video Prediction. *arXiv:1710.11252 [cs]*, March 2018. URL `http://arxiv.org/abs/1710.11252`. arXiv: 1710.11252.

[4] Tom Beucler, Michael Pritchard, Stephan Rasp, Jordan Ott, Pierre Baldi, and Pierre Gentine. Enforcing analytic constraints in neural-networks emulating physical systems. *arXiv preprint arXiv:1909.00912*, 2019.

[5] Emily Denton and Rob Fergus. Stochastic video generation with a learned prior. *arXiv:1802.07687 [cs, stat]*, Mar 2018. arXiv: 1802.07687.

[6] Aman Kumar Gupta, Dipak Yadav, Priyanka Gupta, Supriya Ranjan, Vishal Gupta, and Sirpat Badhai. Effects of climate change on agriculture. *Food and Agriculture Spectrum Journal*, 1(3), 2020.

[7] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020.

[8] G.J. Huffman, E.F. Stocker, D.T. Bolvin, E.J. Nelkin, and Jackson Tan. Gpm imerg final precipitation l3 half hourly 0.1 degree x 0.1 degree v06. Technical report, 2019. ftp://arthurhou.pps.eosdis.nasa.gov/gpmdata/.

[9] Alex X Lee, Richard Zhang, Frederik Ebert, Pieter Abbeel, Chelsea Finn, and Sergey Levine. Stochastic adversarial video prediction. *arXiv preprint arXiv:1804.01523*, 2018.

[10] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[11] Pauline Luc, Aidan Clark, Sander Dieleman, Diego de Las Casas, Yotam Doron, Albin Cassirer, and Karen Simonyan. Transformation-based adversarial video prediction on large-scale data. *arXiv:2003.04035 [cs]*, Mar 2020. arXiv: 2003.04035.

[12] Ashray Manepalli, Adrian Albert, Alan Rhoades, Daniel Feldman, and Andrew D Jones. Emulating numeric hydroclimate models with physics-informed cgans. In *AGU Fall Meeting 2019*. AGU, 2019.

[13] Sergiu Oprea, Pablo Martinez-Gonzalez, Alberto Garcia-Garcia, John Alejandro Castro-Vargas, Sergio Orts-Escolano, Jose Garcia-Rodriguez, and Antonis Argyros. A review on deep learning techniques for video prediction. *arXiv:2004.05214 [cs, eess]*, Apr 2020. URL `http://arxiv.org/abs/2004.05214`. arXiv: 2004.05214.

[14] Prajit Ramachandran, Barret Zoph, and Quoc V Le. Searching for activation functions. *arXiv preprint arXiv:1710.05941*, 2017.

[15] Stephan Rasp, Peter D. Dueben, Sebastian Scher, Jonathan A. Weyn, Soukayna Mouatadid, and Nils Thuerey. WeatherBench: A benchmark dataset for data-driven weather forecasting. *arXiv:2002.00469 [physics, stat]*, June 2020. arXiv: 2002.00469.

[16] Roger Saunders, James Hocking, Emma Turner, Peter Rayer, David Rundle, Pascal Brunel, Jerome Vidot, Pascale Roquet, Marco Matricardi, Alan Geer, et al. An update on the rttov fast radiative transfer model (currently at version 12). *Geoscientific Model Development*, 11(7), 2018.

[17] Casper Kaae Sønderby, Lasse Espeholt, Jonathan Heek, Mostafa Dehghani, Avital Oliver, Tim Salimans, Shreya Agrawal, Jason Hickey, and Nal Kalchbrenner. Metnet: A neural weather model for precipitation forecasting. *arXiv preprint arXiv:2003.12140*, 2020.

[18] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[19] Yaohui Wang, Piotr Bilinski, Francois Bremond, and Antitza Dantcheva. Imaginator: Conditional spatio-temporal gan for video generation. In *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, page 1149–1158. IEEE, Mar 2020.

[20] SHI Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*, pages 802–810, 2015.

A   Dataset descriptions

SimSat SimSat emulates three spectral channels from the Meteosat-10 SEVIRI satellite [2] and is generated from ECMWF's high-resolution weather-forecasting model using the RTTOV radiative transfer model [16]. SimSat provides information about global cloud cover and moisture features and has a native spatial resolution of about $0.1°$ at three-hourly intervals. This dataset is available from 2016 onward.

ERA5 ERA5 provides hourly estimates of a variety of atmospheric, land and oceanic variables, such as specific humidity, temperature and geopotential height at different pressure levels [7]. Estimates cover the full globe at a spatial resolution of $0.25°$ and are available from 1979 to present.

IMERG IMERG is a global half-hourly precipitation estimation product provided by NASA [8]. Specifically we use the Final Run product which primarily uses satellite data from multiple polar-orbiting and geo-stationary satellites. This estimate is then corrected using data from reanalysis products (MERRA2, ERA5) and rain-gauge data. IMERG comes at native spatial resolution of $0.1°$ and is available from 2004.

These data sources are available all together only from 2016 onward.

B   State Estimation

B.1   Training setups

We define the three training setups as follows, depending on the data used as input:

1. **SimSat only.** Here the input sequence is formed of the 3 frequency bands of SimSat data ($clbt : 0 - 2$). We take $\Delta t = 3$ and $T = 12$.

2. **SimSat and ground observations.** The input sequence is formed of $clbt : 0 - 2$ and ground observations (namely sp and t2m). We take $\Delta t = 3$ and $T = 12$.

3. **SimSat, ground observations and historic atmospheric states.** We take an input sequence $\Delta t = 24$ and $T = 72$. At time steps $t = -72$ to $t = -24$, the input vector is formed of $clbt : 0 - 2$, ground observations and the ERA state variables. At $t = 0$, the input vector is formed only of $clbt : 0 - 2$ and ground observations.

Each of the input sequences specified above is then concatenated with static variables (latitude, longitude, land-sea mask, orography, soil type), repeated per time step. As output, we regress the following ERA5 variables: geopotential, temperature and humidity (sampled at $300\,\text{hPa}$, $500\,\text{hPa}$ and $850\,\text{hPa}$), cloud liquid water content and cloud ice water content (sampled at $300\,\text{hPa}$ and $500\,\text{hPa}$), as well as the surface pressure (`sp`) and 2-metre temperature (`t2m`).

B.2 Additional results

To understand the state extraction results further, Figure 3 visualizes the prediction given by the third setup **SimSat, ground observations and historic atmospheric states** of a randomly chosen time step. The model is indeed able to capture each variable's specificity.
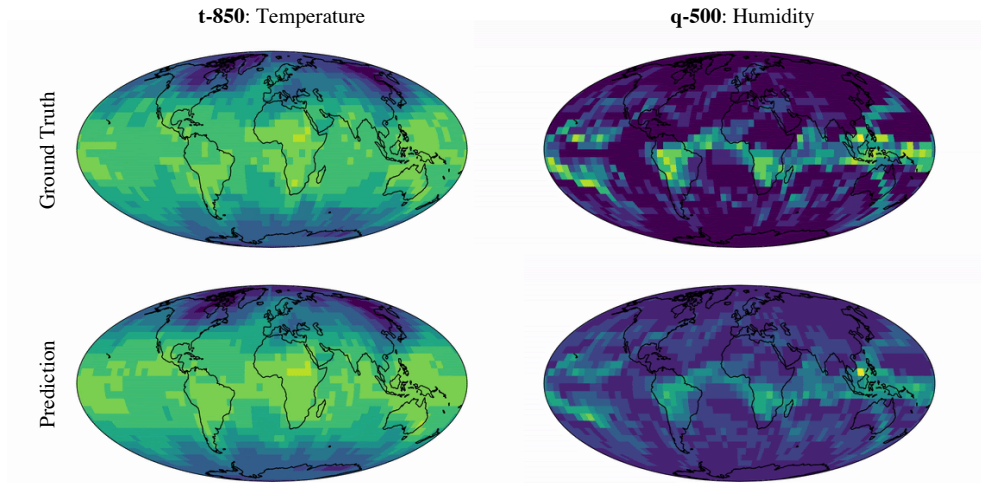

Figure 3: State Estimation Result of a randomly chosen time point.

C State Forecasting

C.1 Additional results

As a qualitative result of the methodology, we can refer to Figure 4, where we depict variable `t2m`, first globally and then focused on a different sampled future above Australia. Here we can notice the qualitative difference between the two futures, both sampled from the same current state.

D Precipitation Estimation

We provide further details about the experimental setup for performing the precipitation estimation step. We start by justifying the choice of using IMERG as target value for our experiments, to then describing the data sampling strategies for tackling the output imbalance and finally reporting additional results.

D.1 Analysis of precipitation estimates

Precipitation intensity is an extremely imbalanced variable, as its probability density decreases exponentially fast. In general, precipitation events of intensity higher than $2.5\,\text{mm}\,\text{h}^{-1}$ are rare and heavy or violent rains seldom occur. Data-driven approaches generally suffer from data imbalance, as
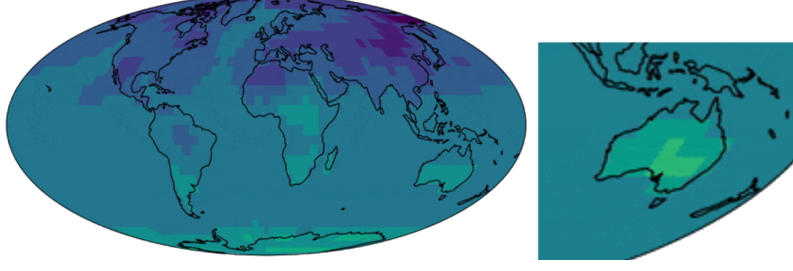
Figure 4: Qualitative Result of the stochastic video prediction algorithm on the `t2m` variable. The reader can note the different outcomes on Australia where two different futures have been sampled.

they tend to be biased towards the expected value of the target distribution. This problem is aggravated by the fact that certain precipitation estimates fail to accurately represent extreme precipitation events.

5(a) shows the distribution of precipitation for years 2010-2016 for ERA5's *total precipitation* and IMERG's *calibrated precipitation* at different resolutions. Regridding is achieved using bilinear interpolation, which does not conserve global precipitation (within an hour). However, using conservative interpolation to such a coarse resolution was found to remove all extreme precipitation events. In general, distributions are positively skewed with no precipitation and light precipitation ($0\,\mathrm{mm}$ to $2.5\,\mathrm{mm}$) events over-represented for both estimates. We further notice that precipitation distributions vary depending on the dataset, with ERA5 modelling poorly extreme events, and on the selected resolution, in particular for ERA5. The choice of target dataset is indeed crucial for training accurate models.
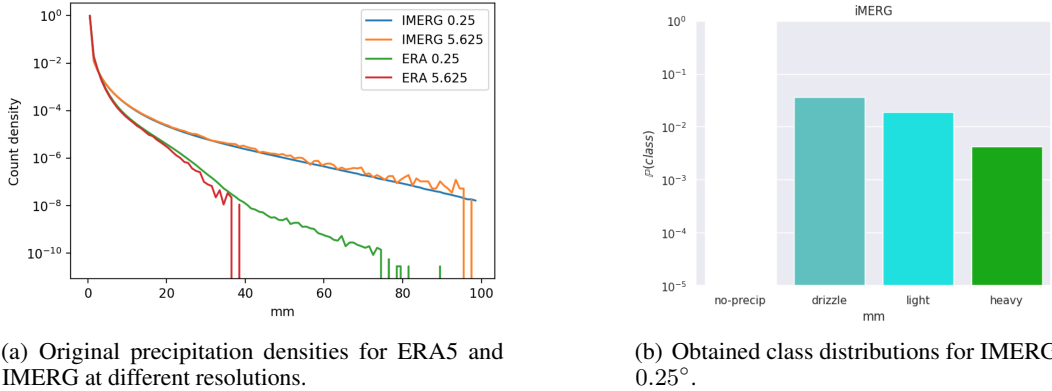
(a) Original precipitation densities for ERA5 and IMERG at different resolutions.

(b) Obtained class distributions for IMERG $0.25°$.

Figure 5: Precipitation distributions for years 2010-2016. Densities are reported in logarithmic scale.

D.1.1 Data sampling

In order to deal with the large output imbalance, we define the following four classes of precipitation intensity accumulated over three hours:

1. No-precipitation: precipitation intensity below $1\,\mathrm{mm}$;

2. Drizzle precipitation: precipitation intensity between 1 and $7.5\,\mathrm{mm}$;

3. Light precipitation: precipitation intensity between 7.6 and $22.8\,\mathrm{mm}$;

4. Heavy precipitation: precipitation intensity above $22.8\,\mathrm{mm}$.

These classes are useful for defining data sampling strategies to balance the target distribution at training and for defining metrics for assessing the quality of predictions fairly across the types of precipitation. Figure 5(b) reports the obtained class distributions for IMERG at $0.25°$ resolution. The choice of a gridcell-wise model allows us to balance the defined classes at training, through a masking process when evaluating the chosen loss function: given an input X_t and a target label mask

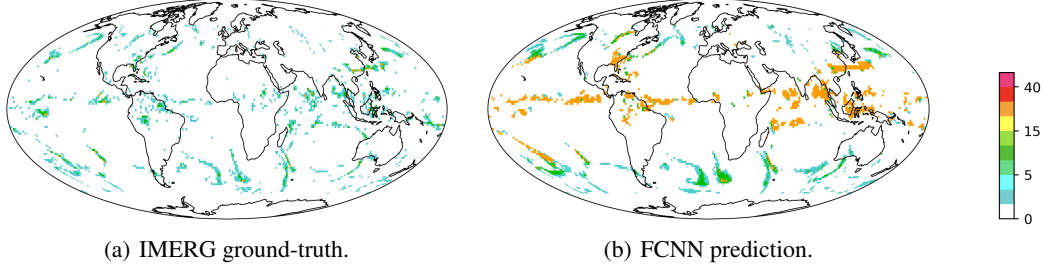(a) IMERG ground-truth.   (b) FCNN prediction.

Figure 6: Comparison of ground-truth and predicted precipitations, accumulated over 3-hour period in 2019. Globally our model accurately captures the rain no-rain dichotomy. Our best networks struggle with over prediction of heavy rain, particularly in the tropics.

y_t, we select all pixels of the minority class within the input X_t and randomly select an equal amount of pixels for each of the other classes, masking-out the remaining pixels. This procedure ensures that classes are equally represented at training and that the model can be easily fine-tuned together with the models of the previous steps. To further address the data imbalance, in the experiments we train the model by minimizing the Focal Loss [10]. This loss is a modified version of the Cross Entropy loss that decreases the weights of well-classified examples, so that the training focuses on hard examples which are typically the ones from under-represented classes.

D.1.2 Model specifications

We build a gridcell-wise Fully Connected Neural Network (FCNN) as a concatenation of a batch normalization, five fully connected layers with Swish activation [14] and dropout [18], and a final fully connected layer.

Inputs consist of ERA5's temperature (t), humidity (q), cloud liquid water content (clwc), cloud ice water content (ciwc), at $300\,\mathrm{hPa}$, $500\,\mathrm{hPa}$ and $850\,\mathrm{hPa}$ geopotential levels and surface temperature (sp), land-sea mask (lsm) and orography. Considered that at a given time t IMERG provides the precipitation accumulated over the period $[t, t+1]$, the model takes as inputs the feature values for both time steps: $X_t = [x_t; x_{t+1}]$. At training, we feed into the model only gridcells with latitude in the range $[-60°, 60°]$, as IMERG is not available at higher latitudes.

D.1.3 Additional results

Figure 6(a) provides an example of the obtained ground-truth global precipitations and Figure 6(b) reports the predictions of our FCNN for the same time step. Visually, we notice that our FCNN is accurate in distinguishing between rainy and non-rainy cells, but it has a tendency to overestimating precipitation rates. Notice that both ERA5 and FCNN models show limited predictive capabilities on minority classes, although for the opposite reasons: ERA5 underestimates precipitations, as shown in 5(a), while the trained FCNN overestimates them.