Training of RainForests to calibrate rainfall forecasts from the ECMWF ensemble over Singapore

Tan Jun Yuan, Sylvester

March 25, 2024

Abstract

This report serves as a summary of the work I have done during my internship here with the Department of Weather Research at the Centre of Climate Research Singapore (CCRS). Recently, it has become more and more clear to me that machine learning is a force to be reckoned with especially in data-heavy industries/fields like numerical weather prediction (NWP). Machine learning presents to such industries a new dimension of possibilities and opportunities

The overlap between machine learning and weather prediction is a rapidly developing space and it seems so far that machine learning is successful in improving NWP capabilities. More notably, when deployed as a supplement to weather forecasts, the improvement in skill of the forecasts are immense compared to the additional computational costs incurred from deployment. In the case of my internship, *I will be training and deploying a machine learning model, 'RainForests' to calibrate and improve the weather forecasts over Singapore generated by the ECMWF Ensemble (ENS).*

1 ECMWF ENS

The ECMWF ensemble forecast suite (ENS) is a global model that models that evolution of the weather and also captures land-atmosphere coupling and ocean-atmosphere coupling over the entire planet. It is a probabilistic[1] suite that contains 51 ensemble members which includes 1 control member and 50 perturbed members.

The control member takes in the best estimate of the initial state of the weather while the other 50 perturbed members take in a perturbed instance of the initial state. The idea is that the range of model outputs give us an insight on how uncertain or how sensitive the output is to changes in the initial condition which will help in making a more informed weather forecast. Overall, the output is most helpful in providing an estimate of the probability of an event happening. For example, if most of the ensemble members agree that more than 5.0 mm of precipitation will fall tomorrow, then the probability of that event really occurring is quite high, subject to further calibrations and adjustments.

2 Postprocessing

ECMWF ENS being a global model means that it is incredibly informative and that anyone from anywhere can tap on the outputs provided. However, it being a global model also means that it will not be able to accurately capture the weather biases that exist in specific locations.

To make the ECMWF ENS outputs still valuable, methods have been developed to adjust and adjust these outputs to remain valuable, even in local contexts. These methods are known as ‘postprocessing’.

2.1 Bias correction

Referring to Gneiting’s [1] paper on calibration of weather forecasts, bias corrections refer to adjustments made to a single deterministic forecast, to give another single valued forecast. Bias correction approaches typically make use of regression methods, which identify statistical relationships between inputs and outputs. Most well-known bias corrections approaches include the Model Output Statistics (MOS) approach pioneered by Glahn and Lowry [2], as well as Perfect Prog (PP) [3].

In today’s age where ensembles are leveraged on heavily for weather forecasting, one may suggest to utilise bias correction approaches on every ensemble member in order to obtain an improved set of outputs from the ensemble. However, it is argued that for skewed variables (like precipitation and wind speeds), bias corrections are not particularly helpful. Additive terms introduced by regression would affect all of the zero precipitation forecasts, while multiplicative terms only affect nonzero forecasts and do not help adjust the frequency of zero precipitation forecasts [4].

2.2 Ensemble calibration

Ensemble calibration refers to adjustments made to the forecasts probability distributions. Ensemble calibration, like bias corrections, typically also involve statistical regression against some training data. The difference is that in ensemble calibration, a distribution is fitted to the

[1]Actually, this probabilistic approach to weather forecasting came from a paradigm shift towards forecasting methodologies after Edward Lorenz’s treatment of chaos theory. This was also when Lorenz’s famous butterfly attractor became known as the symbol of chaos. Anyway, the probabilistic approach to forecasting is a clever strategy used to mitigate the effect of chaos theory onto the final output. We touch more on this in section 3.4.2

training data[2]. The aim of postprocessing then is to maximise the sharpness (concentration) of the probability distributions [1] as well as to account for biases or effects that are present in the local region but not captured in ECMWF ENS. Today, there are various techniques developed for postprocessing and the reader is advised to consult Vannitsem's work [5] for a thorough review of the postprocessing techniques of today. In general, the state of the art techniques can be classified into two distinct groups -

1. containing assumptions on the probability distribution
2. containing no assumptions on the probability distribution.

2.2.1 Containing assumptions on the probability distribution

Postprocessing techniques that make some assumptions on the probability distribution are typically referred to as 'parametric postprocessing'. They make assumptions on the form of the probability distribution and select a suitable family of probability distributions. Regression coefficients are then introduced and these coefficients are optimised via regression.

2.2.2 Containing no assumptions on the the probability distribution

Postprocessing techniques that avoid assumptions on the distributions usually construct approximations of the forecast distribution. An example of such a technique is ecPoint [6] (which we will touch on in more detail in section 4) where a decision tree takes in various forecasted quantities (wind speed at 700hPa altitude, 24 hour solar radiation, etc.) and outputs a calibrated probability distribution for precipitation.

2.3 RainForests' role in postprocessing

We have just explored the broad types of postprocessing techniques that have been developed and studied, which category, then, does RainForests belong in? Relevant facts about RainForests are that it takes in a single valued forecast from the control member as well as some ensemble statistics (ensemble mean and ensemble standard deviation of precipitation), outputs a cumulative probability density function and makes no assumption about the probability distribution whatsoever. Interestingly, it seems that RainForests cannot be fit into a single category alone from the list of categories laid out by Gneiting [1] and Vannitsem [5].

As mentioned above in section 2, raw output from the ECMWF ENS for the weather forecast over Singapore are not very skillful because they do not account for specific biases that only occur within the local region. With that, postprocessing of the raw outputs from ECMWF ENS is essential to extract more value from ECMWF ENS outputs.

In fact, to rely on ECMWF ENS outputs for weather forecasting in Singapore, significant post processing has to be done so that the forecasts are not out of trend or anomalous. This project of training RainForests to calibrate rainfall forecasts from ECMWF ENS is in essence, an attempt to incorporate machine learning into the postprocessing of the raw output from ECMWF ENS over Singapore, with hopes that the postprocessed ECMWF ENS outputs will perform even better. In particular, RainForests, like ecPoint, attacks a specific aspect of precipitation postprocessing - sub-grid variability (more in section 4.1)

[2] There are some subtle differences between 'fitting a forecast to training data' and 'fitting a distribution to training data' here. It is straightforward to do the former because training data is also of the same form (single valued). However, to fit a distribution to single valued training data requires the help of special loss functions like the Brier Score.

3 Numerical Weather Prediction

The atmospheric processes that give rise to the weather we experience is well-described by the Navier-Stokes equations, which are a set of partial differential equations. With the advent of super computers, we can in principal numerically solve the Navier-Stokes equations given the initial conditions of our atmosphere. In other words, we can calculate what the state of the weather will be some time in the future, given the state of the weather now. One could say that predicting the weather numerically amounts to solving an initial-value-problem with the Navier-Stokes equations.

However, some *practical* problems prevent us from obtaining a perfectly accurate calculation of the future weather. The main problems are

1. Imperfect representation of information and imperfect information
2. Assumptions in Navier-Stokes equations
3. Chaos theory

3.1 Imperfect representation and information

Numerically calculating the state of the weather $\mathbf{F}(t')$ for some future time t' involves solving a set of partial differential equations given an initial condition $\mathbf{F}(0)$. Two fundamental questions naturally arise when we proceed to perform a numerical calculation for the state of the weather.

1. How do we represent the state of the weather for any given time $\mathbf{F}(t)$ such that it is operable by supercomputers for the calculations? We explore this in 3.1.1
2. What appropriately constitutes an initial condition $\mathbf{F}(0)$? We tackle this problem in 3.1.2

3.1.1 Digital representations of reality

In fields like classical mechanics, we represent our real physical space as being a continuous space as we mathematically model dynamics in our real physical space. However, we cannot do so when we are performing numerical calculations because it would take an infinite memory for a computer to represent a continuous space. Instead, we discretise the space so that the computer has a chance to run calculations. This comes with a trade-off though, that is that we will not be able to perform exact calculations numerically - all calculations will be an approximate. This means that when we simulate a physical system numerically, we do not actually obtain an exact picture of what is happening, but rather a ‘pixelated' version of what is going on.

3.1.2 We are not omnipresent

What constitutes an appropriate initial condition $\mathbf{F}(0)$? To answer this question, we need to think about how we are representing the state of the weather in the first place. Mathematically, the form of the initial condition must be equivalent to that of our desired $\mathbf{F}(t')$. Therefore, if we want a useful description of $\mathbf{F}(t')$, which typically happens to be a multidimensional[3] array spanning across a discretised space ${\mathbb{R}_\square}^3$, then the form of our initial condition $\mathbf{F}(0)$ has to also be as such.

To construct $\mathbf{F}(0)$, the most direct way would be to turn towards weather stations that are actively collecting data. However, as mentioned in the previous paragraph, $\mathbf{F}(0)$ is technically

[3]Generally speaking, the dimensionality of this array depends on how many quantities makes up the description of the weather (a dimension for each quantity).

mathematically demanded to represent every point in the $\mathbb{R}_{\square}{}^3$ that we are forecasting for. This implies that we need access to information from *everywhere, all at once*. More precisely, this entails pulling data from at least one weather station in every point in $\mathbb{R}_{\square}{}^3$, which is impossible since we obviously do not have at least one weather station in every point in $\mathbb{R}_{\square}{}^3$. If we do indeed have at least one weather station in every point in $\mathbb{R}_{\square}{}^3$, one can imagine how potentially disruptive to infrastructure these weather stations would be, since this would mean there would be weather stations quite literally all over the place, at least within the domains of the space we want to forecast for. Therefore, one can identify a critical information gap for our representation of $\mathbf{F}(0)$ simply because *we are not omnipresent* and do not have weather stations everywhere to measure data for $\mathbf{F}(0)$. Realistically speaking, this means that however way we choose to represent $\mathbf{F}(0)$ in order to proceed with our calculations for $\mathbf{F}(t')$, we cannot strictly speak of $\mathbf{F}(0)$ as anything more than a clever approximation[4] of the state of the weather at $t = 0$.

3.2 Assumptions in Navier-Stokes equations

Even in the perfect, unrealistic situation where we have all the information we need, we will find that we still are unable to obtain a perfectly accurate weather forecast. This is because of some assumptions that were made in the derivation of the Navier-Stokes equations that render the equations unable to perfectly represent the behaviour of fluids. These assumptions are that

- fluids behave as if they were continuous and not constituted by discrete particles
- fluids are incompressible

However, it should be said that these assumptions were made to greatly simplify our derivation of a model that accurately captures fluid dynamics. For all we know, we may never arrive at a mathematical models that performs as well as the Navier-Stokes equations had we not made these simplifying assumptions.

3.3 Chaos theory

Chaos theory drastically grows any errors, however minor, we incur. This leaves us with little room for error which is a difficult situation for us because sources of error are abundant in the whole process of NWP.

Mathematically, a chaotic equation is one where small differences in the initial conditions can lead to vastly different outcomes. Unfortunately, the equations governing how our weather evolves are chaotic. This makes the incurrence of any form of error harshly punishable. From the rounding error of values in a supercomputer, to the imprecise measurement of the initial state by weather stations that reside in far from ideal conditions.

As previously mentioned, real time observations, with data assimilation techniques give us an idea of what the initial state of the weather. Therefore, it is clear that real time observations form an essential component of forecasting the future state of the weather since the business of forecasting the weather is solving an initial-value problem. With that being said, one of the main sources of error that significantly diminish the skill of forecasts is the construction of the initial conditions [5].

We know to a great extent, the processes in the atmosphere that drive the evolution of the weather, and we also know quite rigorously the equations that govern the evolution of the weather. We can model how the weather changes from a given state, but alluding to section 3.1.2, we cannot yet, however, systematically construct a perfect representation of the current

[4] In the study of *Data Assimilation*, numerous techniques are developed to effectively approximate $\mathbf{F}(0)$ to serve as a good enough substitute for the true $\mathbf{F}(0)$.

weather from real time observations. And any imperfections in a component as important as the initial condition (or for any component, for that matter) will surely contribute significant errors in our NWP output.

3.4 Mitigation and strategies

With that being said, we are not compeletely hopeless in accurately forecasting the weather. It turns out that with some clever strategies, we can minimise these sources of error to give us a shot at forecasting the weather accurately. We will take a look at some of these strategies in the next few paragraphs.

3.4.1 Data Assimilation: mitigating imperfect information

It was mentioned in section 3.1.2 that we are unable to construct an initial condition because we are not 'omnipresent'. However, we can still make a good guess as to how the initial condition looks like for points in space that we do not have data for, given the available data. In fact, there is a whole field dedicated to how we can deduce how the true initial condition really is and construct, from whatever data we have on hand, an emulation of the true initial condition to resemble the truth as much as possible. Introducing: Data Assimilation. It is a highly advanced and mathematical field that is essential in crafting one of the crucial ingredients of an accurate weather forecast - the initial condition.

3.4.2 Ensembles: keeping chaos at bay

As we know, the chaotic equations that govern the evolution of our weather are very sensitive to the errors in our initial condition. Therefore, any error in the initial condition that we fail to account/correct for will lead to significant error in the model output. This erroneous model output will then be interpreted by weather forecasters and the final weather forecast that gets released to the public will have a high chance of failing.

The catch is that we will never know if we managed to identify all sources of errors and also accommodated all of them appropriately. A cleverer way of doing things is to create N instances of the initial condition and to perturb them slightly so that they differ from each other slightly. With these N initial conditions, input them into the numerical weather prediction models and obtain N different outputs. The chart below visualises the N different trajectories of states as they evolve through time.

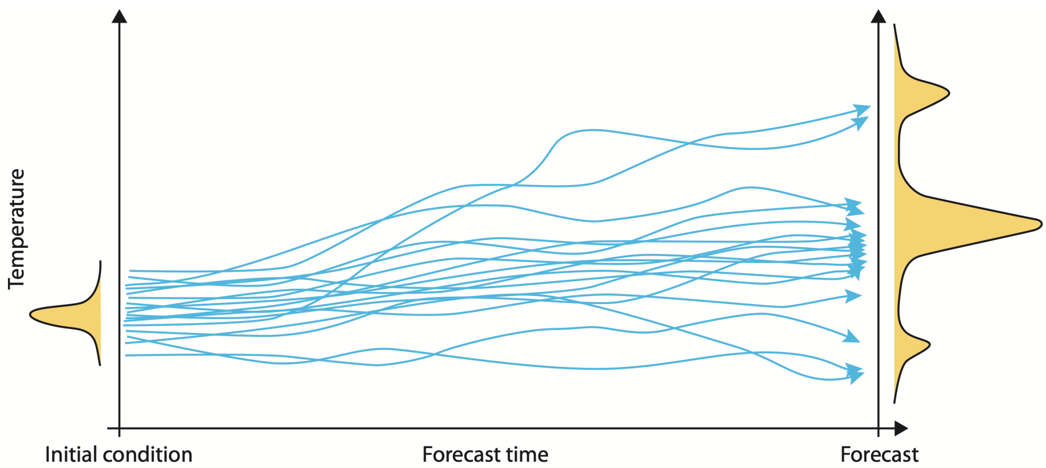


Figure 1: Trajectories of model runs with different initial conditions [7]

When the model outputs seemingly group themselves into 3 distinct outcomes like what is observed in figure 1, it is implied that there is a high chance that the true weather will develop into one of these 3 scenarios. One can imagine that this is not ideal for an operational weather forecaster because they have to make a judgement call on what exactly to forecast for the populace.

When done this way, we obtain a range of possible outcomes from a range of initial conditions and through this we get a glimpse of the different outcomes that could occur, assuming that the *true initial conditions* is captured in our range of initial conditions.

4 ecPoint

Before diving into the RainForests model, it is important to note that RainForests is actually based on the ecPoint [6] method developed by Hewson and Pillosu. The ecPoint method was developed to address the problem of statistically modelling sub-grid variability.

4.1 Sub-grid variability

Sub-grid variability is the variability of values seen within a single grid box (that typically covers an region with area in the order of magnitude of kilometers). Below is an illustration of subgrid variability within a single ECMWF ENS grid box.

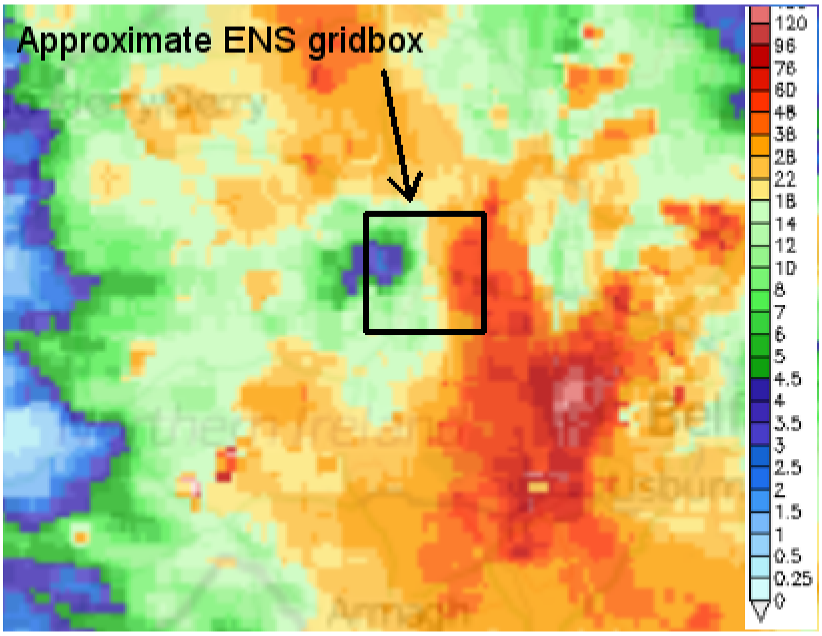


Figure 2: Radar-derived rainfall totals over Northern Ireland for a 12 hour period

Notice that within the same grid box, there exists regions of light rain as well as intense rain.

To model sub-grid variability statistically, the idea is that sub-grid variability of a grid box at a given time is dependent on the 'type of weather' in that grid box at that time. And this 'type of weather' or 'disposition of weather' are classifications of weather that can be captured by values of certain quantities. These quantities are

- wind speed at 700hPa altitude
- convective available potential energy

- 24-hour incident solar radiation
- total precipitation
- total precipitation from convection

Therefore, from the forecasted values of these 5 quantities over a given grid box, the 'type of weather' over that grid box can be inferred, and therefore, the probability distribution of precipitation can be known.

4.2 Calibration of ecPoint

As with any postprocessing technique, ecPoint also requires calibration. In essence, calibrating ecPoint entails the construction of the mapping between feature quantities or 'type of weather' to the corrections needed to represent this 'type of weather'. This mapping is done through a proxy dimensionless metric forecast error ratio (FER).

4.2.1 Forecast error ratio (FER)

Forecast error ratio (FER) is a dimensionless metric for one to characterise the sub-grid variability within a grid box. To calculate FER in a given grid box, there should ideally by at least one observational station that is capable of recording precipitation data within the grid box. FER takes on the form

$$\mathrm{FER} = \frac{r_0 - G}{G} \tag{1}$$

where r_0 is the observed precipitation within the grid box and G is the forecasted rainfall level for that grid box.

Evidently, the FER measures the disconnect between ground truth and the forecasted precipitation level. During the calibration process, the control member of the ECMWF ENS is run for a short term forecast and its output for the feature quantities and G_{control} are noted. With G_{control} as well as r_0 for the time period the short term forecast was run for, we can calculate

$$\mathrm{FER}_{\mathrm{control}} = \frac{r_0 - G_{\mathrm{control}}}{G_{\mathrm{control}}} \tag{2}$$

and construct a frequency distribution of $\mathrm{FER}_{\mathrm{control}}$ over this time period.

This distribution will eventually be interpolated to obtain a probability density function that will act as the correction term to G so as to obtain a more calibrated forecast. Because it was not known a priori that the 5 feature quantities that were identified were good indicators of the 'type of weather', statistical testing (Two-sample Kolmogorov-Smirnov test) was done on many quantities. As it turns out, the distribution of FER values are most dependent on the 5 feature quantities.

To summarise, the disconnect between the raw forecast and the observed precipitation within the grid box is captured within a dimensionless metric FER. Therefore, the frequency distribution of FER within a grid box can be viewed as the correction to the raw forecast G. As it turns out, the raw forecasts 5 feature quantities identified correlate very well with the probability distribution of FER. Therefore, from the raw forecasts of the 5 feature quantities, we know roughly the distribution of FER, and hence the necessary corrections to G to more closely match up with the actual observed precipitation.

5 RainForests

Developed by the Bureau of Meteorology in Melbourne, 'RainForests' [8] is a machine learning model that improves the precipitation forecast amount by inferring the distribution of forecasted precipitation within the grid box. The implementation of RainForests is compatible with either 24 hourly forecasts or 3 hourly forecasts of precipitation levels. In the case of my project, we implemented RainForests for 3 hourly forecasts.

The input of RainForests is the grid-scale forecast for precipitation from the ith member of the ensemble, $\vec{f}_i(t, \vec{r}, \vec{\alpha})$, which is a function of the time the forecast is for, t, position, $\vec{r}$, as well as a series of feature quantities represented as a vector, $\vec{\alpha}$.

The idea is that the probability distribution of precipitation within the grid depends on the 'disposition' of the weather or weather type at that time. Therefore, the training of RainForests will entail ingesting large amounts of data for the feature quantities (that are thought to be highly correlated with the 'disposition of the weather') and learning patterns between the 'disposition' of the weather and the probability distribution of precipitation it brings.

Feature quantities that are identified to be useful in this characterisation of weather are the exact same as the quantities presented in the ecPoint method back in section 4.1, with the addition of two more quantities that describe the ensemble statistics. As such, the feature quantities that are inputted into RainForests are

For ecPoint and RainForests	3 hour mean wind speed at 700hPa altitude 3 hour max convective available potential energy 24 hour incident solar radiation 3 hour total precipitation 3 hour total precipitation from convection
For RainForests	3 hour ensemble mean of total precipitation 3 hour ensemble standard deviation of total precipitation

Table 1: Table of feature quantities used in ecPoint and RainForests

After training, RainForests will be able to make a good inference on how precipitation will be distributed given the 'disposition' of precipitation, which is characterised by the values of the feature quantities.

5.1 Architecture of RainForests

As mentioned above, RainForests was implemented for 3 hourly forecasts and for maximum lead time of 36 hours. The structure of RainForests is that there exists an array of machine learning models for each valid lead time. In this case, there are 12 valid lead times (3h, 6h, 9h, ..., 30h, 33h, 36h). Therefore, there are 12 arrays of machine learning models and in each of the 12 machine learning models, 26 machine learning models are contained. Each machine learning model corresponds 1-to-1 to a precipitation threshold value[5]. In total, there are 12×26 machine learning models, with each machine learning model containing up to 400 gradient boosted decision trees.

These machine learning models are trained to predict probabilities of actual precipitation exceeding their associated threshold values (recall that there is a 1-to-1 correspondence between threshold value and machine learning model). In the end, the array of 26 machine learning

[5]The threshold values are (in mm): 0.01, 0.05, 0.1, 0.2, 0.4, 0.6, 1, 2, 5, 7, 10, 15, 25, 35, 50, 75, 100, 125, 150, 200, 250, 300, 350, 400, 450, 500. As a sanity check, we expect that for threshold value: 0.01 mm, the exceedance probability predicted by the associated GBDT should always be 1, if not close to 1.

models will produce a set of 26 probabilities like

$$S = \{P(X \geq \xi_1), P(X \geq \xi_2), ..., P(X \geq \xi_{26})\},$$

where X is the actual precipitation and ξ_j is the jth threshold precipitation value. For each member of the ensemble, a set of 26 probabilities are generated.

Suppose the ensemble contains 51 members (ECMWF ENS). The process of feeding the ith member of the ensemble's forecast, $\vec{f_i}(t, \vec{r}, \vec{\alpha})$, to RainForests is then repeated for every member in the ensemble, generating a set of probabilities for each ensemble member that we call S_i. From this set of probabilities, a continuous cumulative probability density function, $P_i(x_{\text{forecast}})$ can be generated by interpolation. The final output $\bar{P}(x_{\text{forecast}})$ would then be the average cumulative probability distribution across all 51 members

$$\bar{P}(x_{\text{forecast}}) = \frac{1}{51} \sum_{i=1}^{51} P_i(x_{\text{forecast}}) \tag{3}$$

where x is the forecasted precipitation value.

5.1.1 Gradient Boosted Decision Trees in each machine learning model

A given machine learning model consists of a series of gradient boosted decision trees (GBDTs). The function of this series of GBDTs is to predict via regression the probabilities of the precipitation equalling or exceeding the threshold value associated to the machine learning model $P(X \geq \xi_j)$. How the GBDTs work is almost exactly like a decision tree. A sample decision tree is plotted below

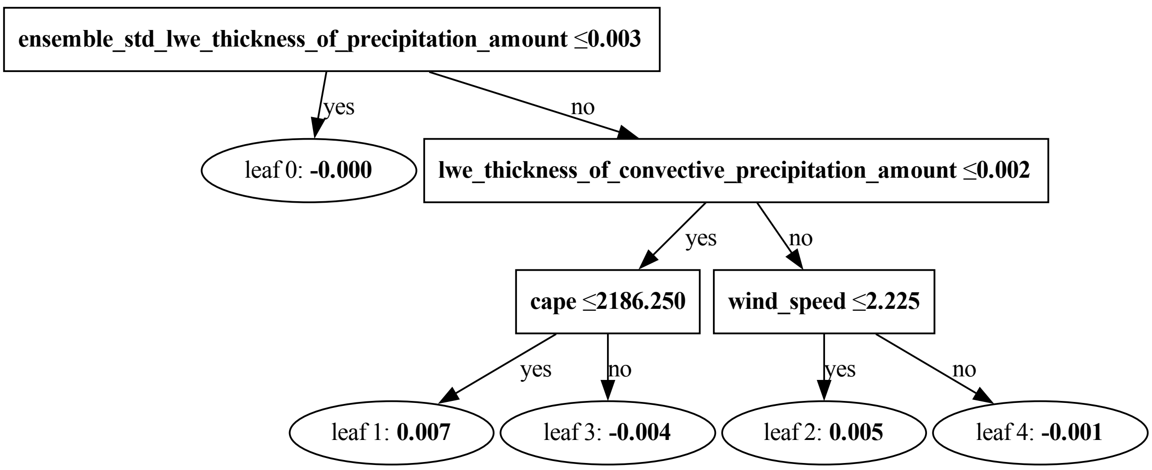


Figure 3: Sample decision tree (46th tree in machine learning model for threshold value 0.5mm for lead time of 6h)

The input $\vec{f_i}(t, \vec{r}, \vec{\alpha})$ of the GBDT enters the Root Node at the top which contains a condition. Depending on whether or not the input satisfies the condition or not, it gets passed onto either the Terminal Node (leaf 0) on the left or the Internal Node on the right. Depending on the outcomes of the conditions at Internal Nodes, a path is traced out, eventually terminating at a Leaf Node which contains an output value associated with it. All GBDTs in each machine learning model can only grow to have 5 Terminal Nodes. The final output of the machine learning model, the predicted probability of rainfall exceeding its associated threshold, is the result of the sum of outputs from all GBDTs within the machine learning model.

6 Training of GBDTs

Recall that the input of the RainForests is the grid-scale forecast for precipitation from a member of the ensemble $\vec{f}_i(t, \vec{r}, \vec{\alpha})$. The GBDT is supposed to take in the grid-scale forecast and output a calibrated precipitation value. Therefore, the ECMWF ENS forecasts from the control member were used as training data, along with actual precipitation data observed from automatic weather stations across Singapore which served as 'ground-truth' verification data. Both the training data and 'ground-truth' verification data spanned from 1st January 2020 to 31st December 2020.

6.1 Features

For RainForests, ECMWF ENS forecasts of 5 quantities and 2 quantities pertaining to ensemble statistics are inputted for the GBDTs to ingest and output the calibrated precipitation forecast.

6.2 Ground-truth

3-hourly precipitation data was pulled from rain gauges from 56 different automated weather stations in Singapore from 1st January 2020 to 31st December 2020. The approximate locations are plotted below.

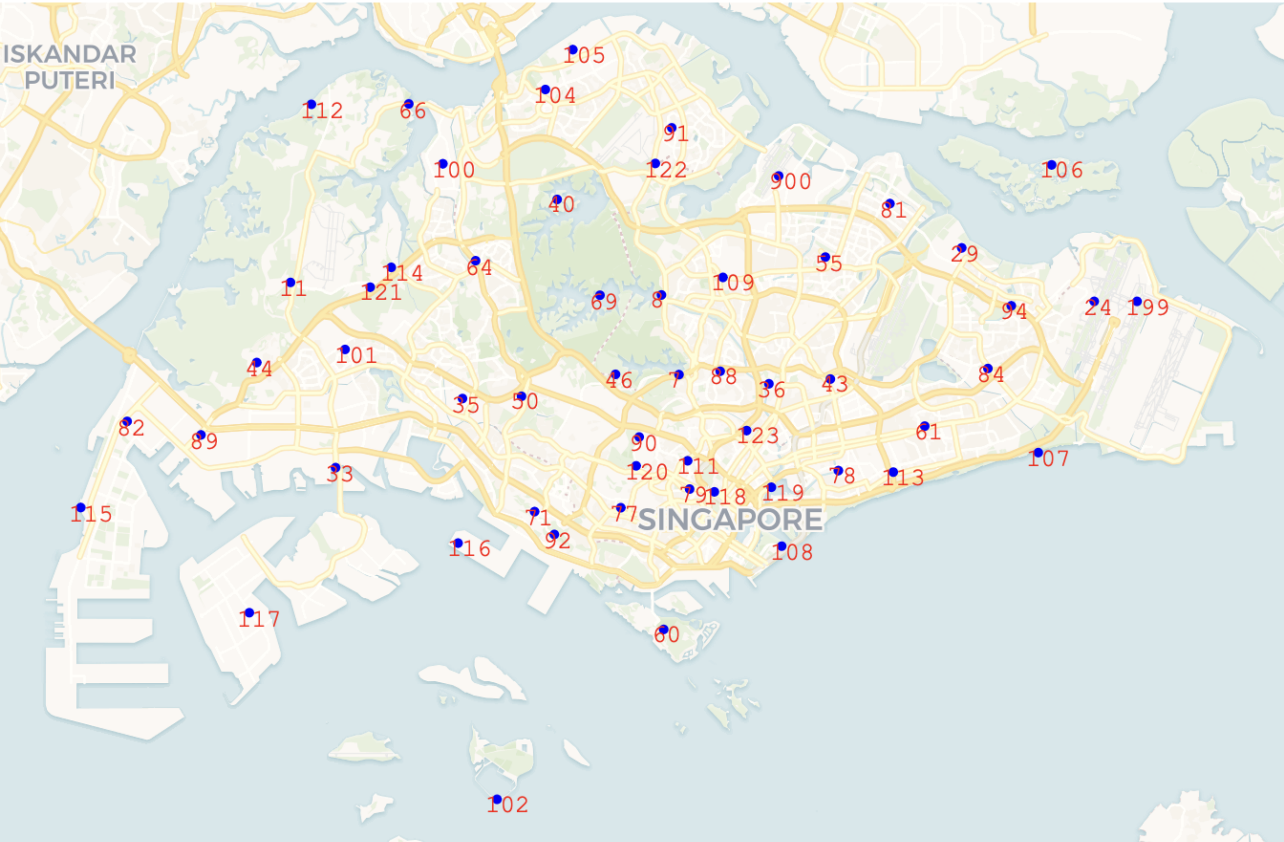


Figure 4: Positions of automated weather stations and their ID numbers

In a grid box, all 3-hourly precipitation data that were pulled from the rain gauges in the grid box are paired with every instance of the forecasted input, $\vec{f}_i(t, \vec{r}, \vec{\alpha})$, matching valid times. For example, every forecast for 9:00am 1st December 2020 in the grid box, regardless of lead time, will be paired with all the observed precipitations at 9:00am in the grid box. Then, with all these observation-input pairs, actual probabilities of observed rainfall exceeding a threshold given a forecasted input $\vec{f}_i(t, \vec{r}, \vec{\alpha})$ can be calculated. The actual probability of observed rainfall exceeding a threshold given the input will serve as 'ground-truth' in the training dataset.

6.3 Training process

The training process of is essentially to adjust the terminal values at leaf nodes and conditions at each non-leaf nodes where splits happen with the goal of minimising the error between the output (predicted probability of rainfall exceeding threshold associated with the GBDT) and the actual probability of observed rainfall actually exceeding the threshold.

The GBDTs that we train and use in RainForests are from the LightGBM python package that is open source and free.

7 Results

We applied RainForests to ECMWF ENS outputs of precipitation over Singapore and plotted the results with Satellite GPM observations on the top-left, ECMWF ENS raw model output on the top right, RainForests trained with Australian data on the bottom left, and RainForests trained with Singapore data (our model) on the bottom right.

7.1 9:00am 15th April 2020, 9h lead time

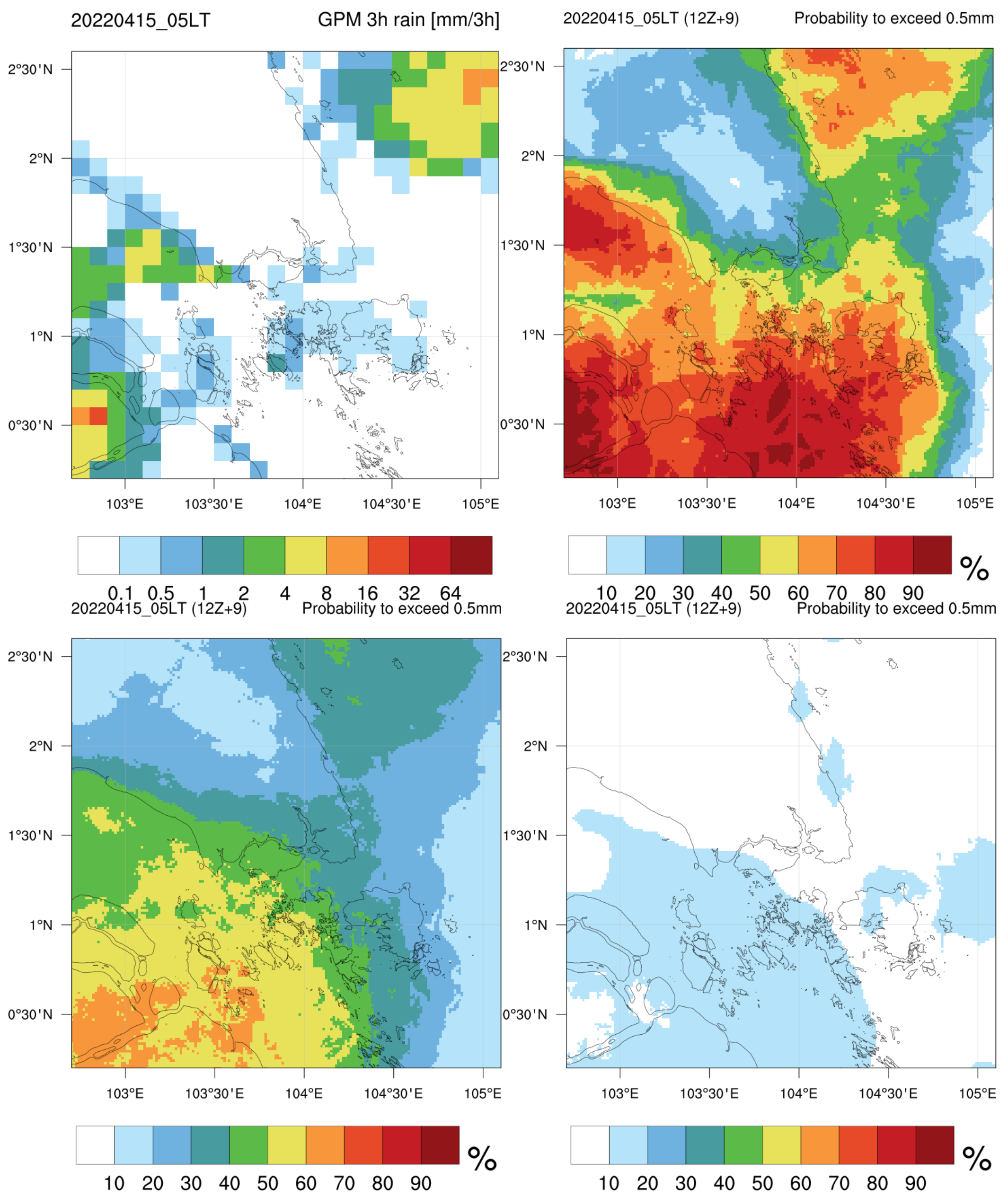


Figure 5: Outputs for 9:00am, 15th April 2020, 9h lead time, 0.5mm threshold

An underforecasting trend is observed with RainForests trained with Singapore data (SG-RainForests) when compared to RainForests trained with Australia data (Aus-RainForests). Worrying is also that SG-RainForests missed out the strong precipitation signal towards the top-right of the panel.

7.2 9:00am 20 March 2020, 9h lead time

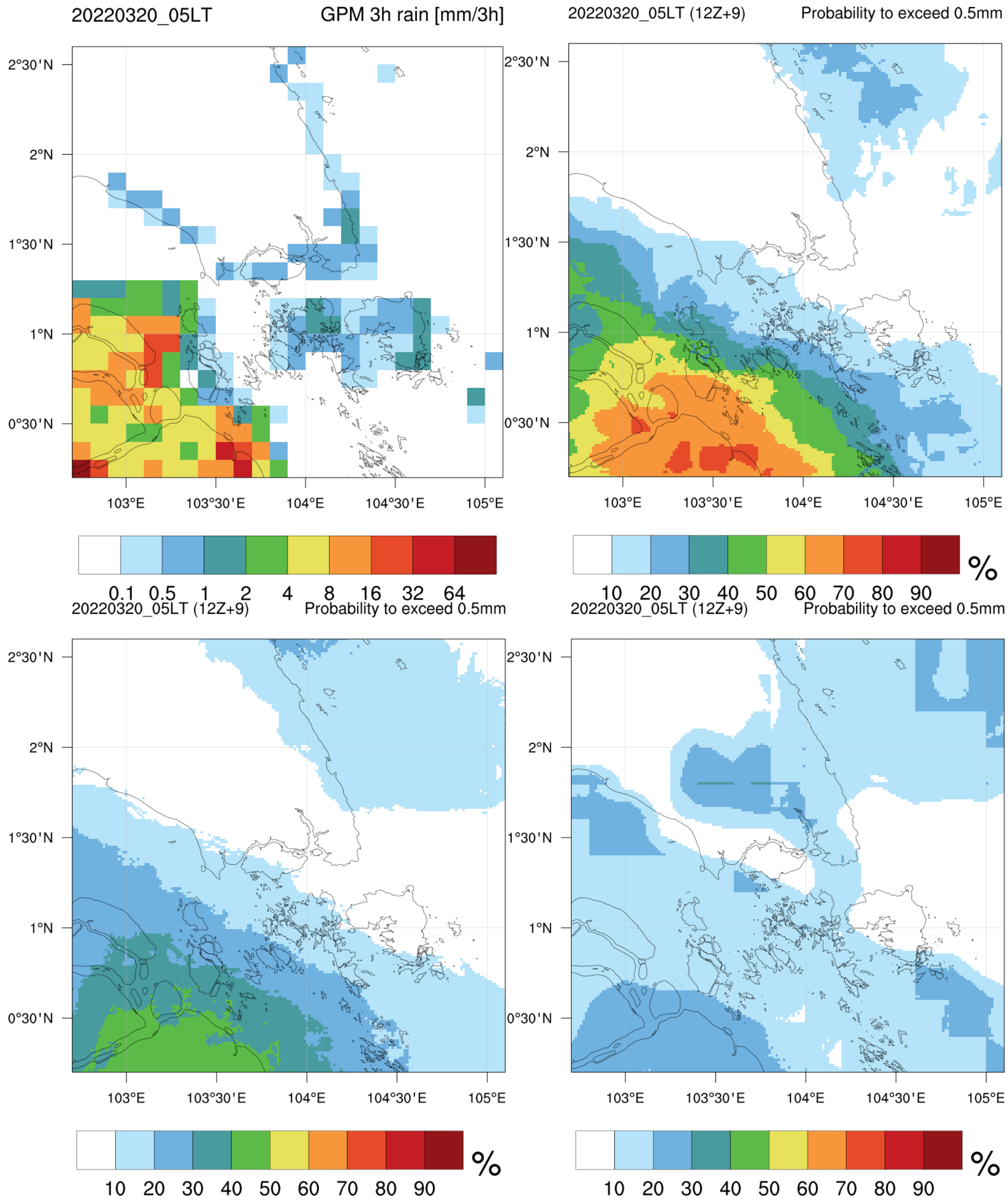


Figure 6: Outputs for 9:00am, 20th March 2020, 9h lead time, 0.5mm threshold

Notice that the SG-RainForests on the bottom-right panel is able to pick up the strong signal of precipitation from the raw ECMWF output (on the top-right panel) on the left half of the panel. However, it seems that the SG-RainForests is heavily underforecasting the probabilities of precipitation exceeding the 0.5mm threshold, relative to Aus-RainForests. A general lack of skill is also observed with a stray signal picked up by SG-RainForests near the centre of the panel.

7.3 ROC curve plot

The receiver-operating characteristic (ROC) curve shows the performance of a classification models. In particular, it plots the true positive rate and false positive rate. In this case, we plotted an ROC curve with a threshold of 0.5mm/3hr of observed precipitation for SG-RainForests, ECMWF ENS raw output and ECMWF ENS processed with Reliability Calibration [4] developed by Flowerdew.

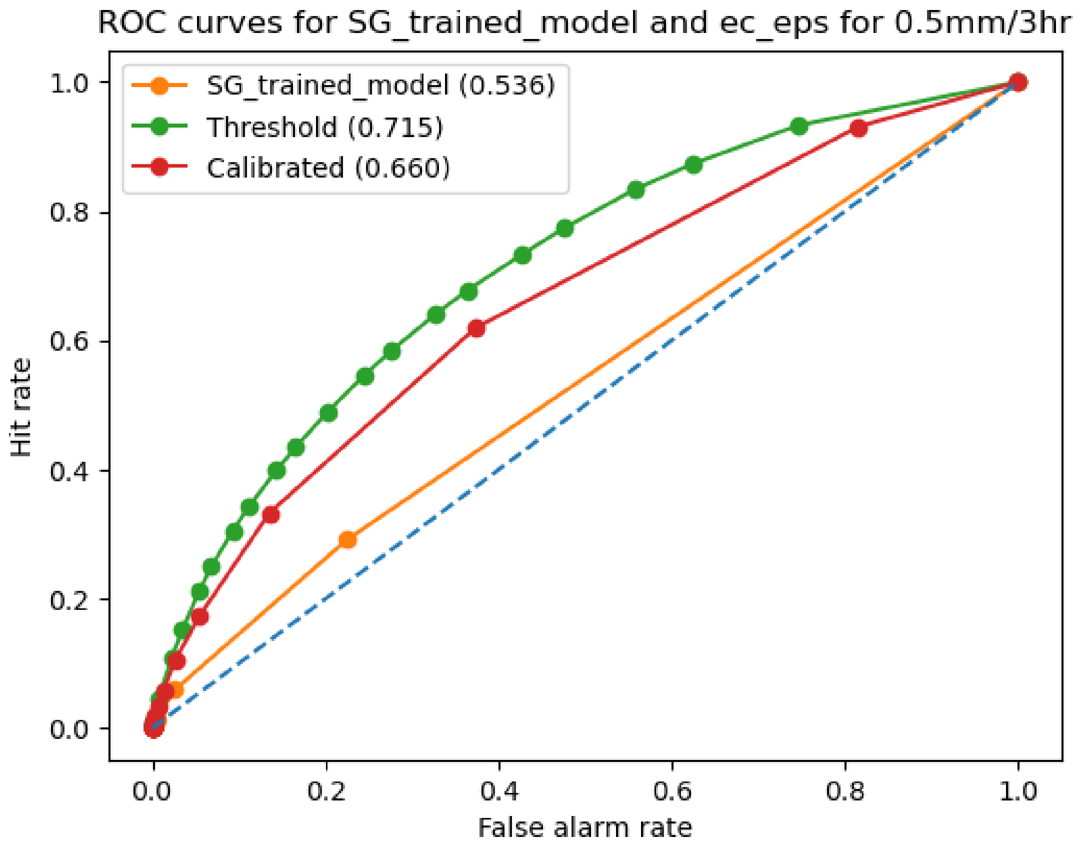


Figure 7: ROC curve for SG-RainForests (orange), Raw output (green), Reliability Calibration (red), baseline is represented by blue diagonal dotted line

A perfect model would show a vertical line on the far left which indicates that the model never ever raises false positives. We can observe that SG-RainForests does not perform outstandingly well (still near to baseline performance). However, it being on the left half of the plot indicates that it has slightly higher true positive events than false positive events when predicting evaluating SG-RainForests' non-zero probability predictions for precipitation exceeding the 0.5mm threshold.

7.4 Reliability plot

The reliability plot is an easy way for us to compare the predicted probabilities against the actual observed probabilities. Again, we plot for probabilities of precipitation exceeding 0.5mm.

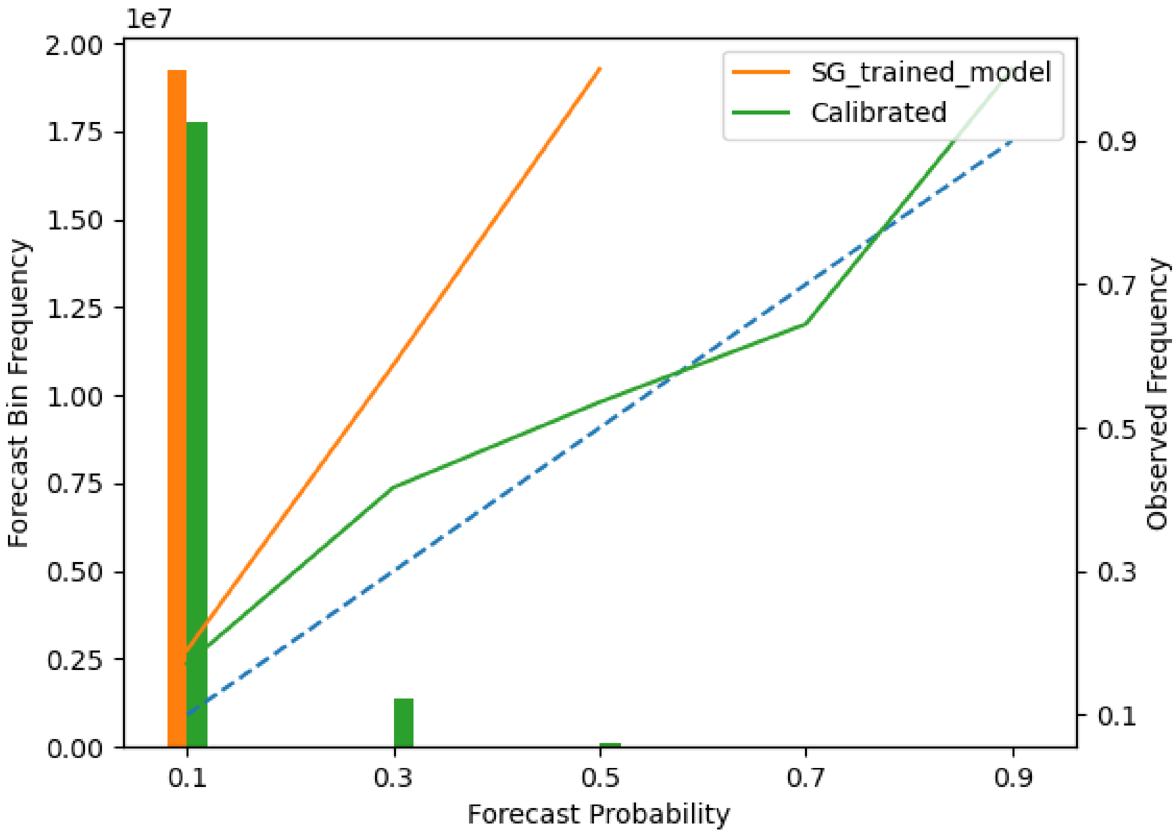


Figure 8: Reliability plot for SG-RainForests (orange) and Reliability Calibration (green)

A perfect model would align perfectly on blue diagonal dotted line. We can observe that the forecast probability of precipitation exceeding 0.5mm from SG-RainForests is always lower than the observed frequency of the same event. This corresponds to the underforecasting we observe in the probability plots in sections 7.2 and 7.1.

8 Discussion

Evidently, SG-RainForests exhibits a dry bias and general lack of skill. We present arguments that might explain these observations on the performance of SG-RainForests.

8.1 Dry bias

Recall that the training dataset ran from 1st January 2020 to 31st December 2020. As it turns out, 2020 was ranked 8th lowest in annual rainfall from 1990 to 2020 [9].

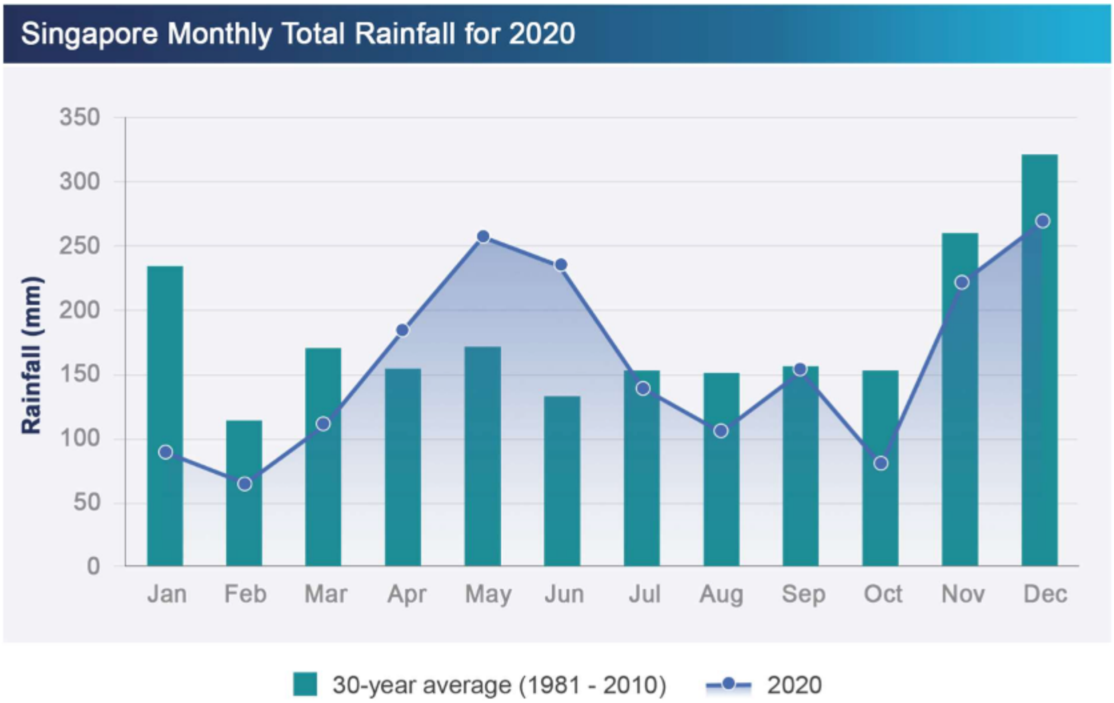


Figure 9: Singapore monthly total rainfall for 30-year average from Changi climate station (bars, 1981 – 2010) compared to 2020 (solid line) [9]

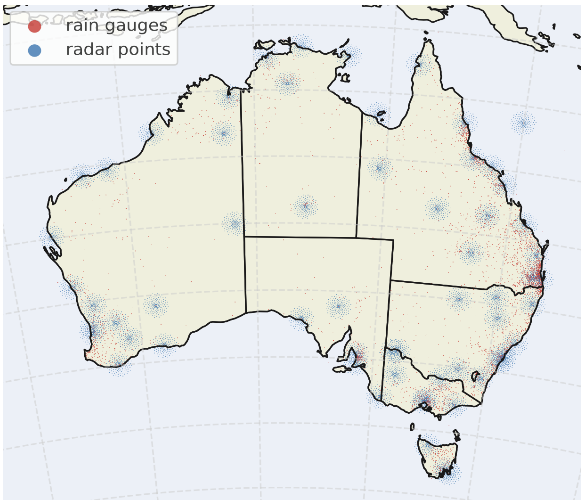
The annual total rainfall of 1886.6 mm for 2020 was 12.9% below the long-term annual average of 2165.9 mm. Perhaps this affected significantly the 'ground-truth' in the training dataset, causing a dry-bias behaviour to show up in SG-RainForests.

8.2 General lack of skill

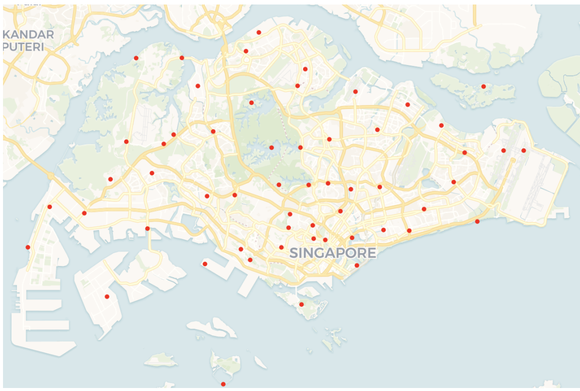
Taking Aus-RainForests as a point of reference, we could attribute the lack of skill to SG-RainForests to the weakness of the training dataset. The structure of the training dataset is such that there is a natural pairing between the forecasted 7 feature variables and the observational data at each time. As such, the availability of the observational data is the limiting factor in both the Australian dataset and Singaporean dataset in terms of the extent of their training datsets.

8.2.1 Number of observational data sources

The plot below compares the observational data for Aus-RainForests to the observational data for SG-RainForests.



(a) Aus-RainForests observational data sources



(b) SG-RainForests observational data sources

Figure 10: Data source positions for Australian dataset and Singaporean dataset

Evidently, the number of observational data sources in the Singaporean training dataset is much smaller than that of the Australian training dataset. This discrepancy in the number of observational data sources and type indicates that Aus-RainForests have learnt more general patterns than SG-RainForests. This could explain the discrepancy in skill between Aus-RainForests and SG-RainForests, even when evaluated over the region around Singapore.

8.2.2 Range and type of observational data

A noteworthy point is also that data from rain radars were used in the Australian training dataset while only rain gauges were used in the Singaporean training dataset.

However, more concerning is how dense the observational data sources are and yet, how little space they cover. In figures 7.2 and 7.1, where we can compare the performances of Aus-RainForests and SG-RainForests, the domain that they were both evaluated on was a 240km by 240km square around Singapore. Yet, the sources of observational data for the Singaporean training dataset are almost only within Singapore as shown in the figure below.

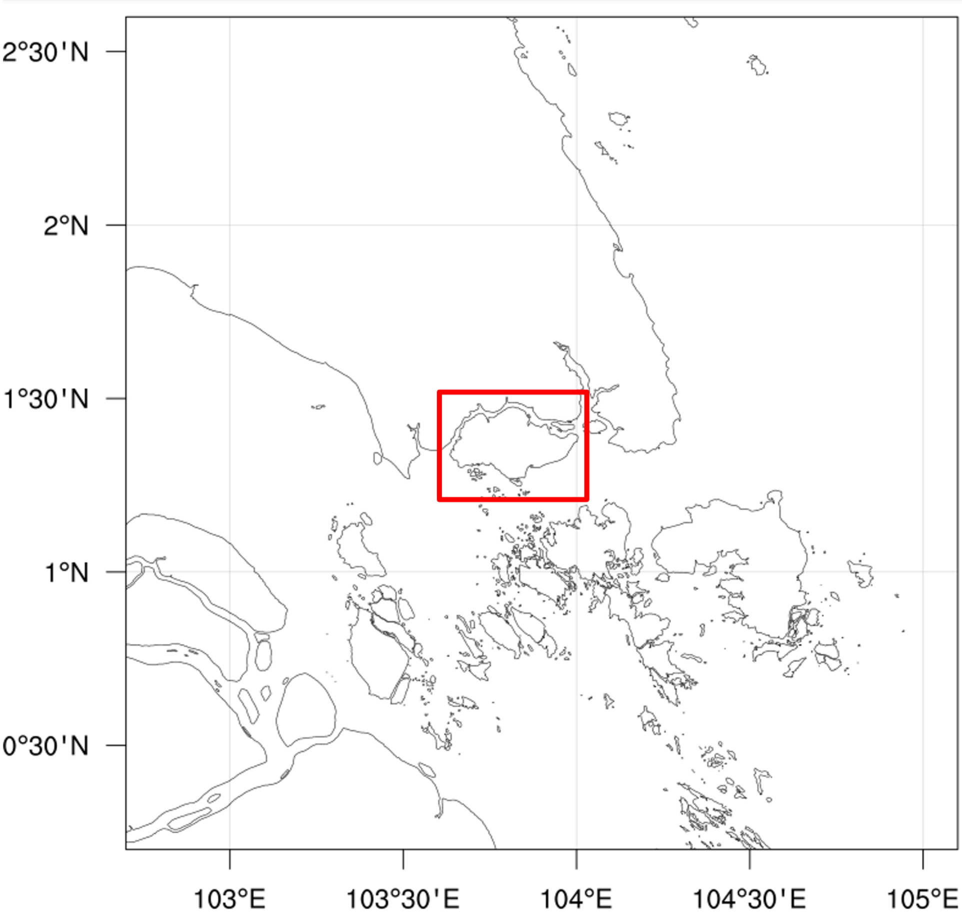


Figure 11: Red box encompasses all over observational data in SG training datset

The dimensions of the red box can be approximated to be 50km by 40km. Thus, we can calculated the ratio, R, between the the area of the red box and the area of the evaluation domain.

$$R = \frac{50 \times 40}{240 \times 240} = 0.03472 \tag{4}$$

This corresponds to the total area that SG-RainForests trained on being only 3.47% of the evaluated area.

Furthermore, in figure 10a, the distance between the eastern and western end of Australia is around 4000km. This indicates that the data in the Australian training dataset is much more general and diversified than that of Singapore's. In addition, because of the coarse resolution of ECMWF ENS (18km at that time), SG-RainForests have only really been trained over less than 10 ECMWF ENS grid boxes over a period of 1 year, which is too little training.

9 Conclusion

We have trained RainForests with ECMWF-ENS forecasts over Singapore as well as Singapore's observational data for precipitation over the period of 1 year (2020). Upon evaluation and comparison with RainForests trained with EMCWF-ENS forecasts over Australia as well as Autralia's observational data, we have found that the performance of SG-RainForests has paled in comparison with Aus-RainForests. We believe that the discrepancy in the performances between them can be attributed to the major weaknesses in the Singaporean training dataset. Moving forward, to see significant improvements in SG-RainForests, we need to expand the Singaporean training dataset.

The easiest way to do this would be to extend the temporal range of the training dataset, which will involve downloading of the ensemble statistics (ECMWF-ENS ensemble mean and standard deviation) for more years. However, this will not address the more severe issues identified in section 8 which have to do with the range and 'scope' of the training dataset.

The next step we can do to tackle the main issue would be to include data from rain radar as they are able to provide data on precipitation observation at very large distances. This will effectively boost the range and scope of the training dataset significantly. We could also tap onto Malaysia's network of rain radars and rain gauges to further boost the range and scope of our training dataset.

Another immediate step we could try is to modify the filtering process in the code for training RainForests such that each machine learning model in RainForests can be exposed to more data instead of it's initial allocated data (total dataset $\div$ (12×26)). However, this should not be done for the actual training of RainForests if it does become operational. This should only be done to evaluate RainForests' performance sensitivity to the volume of the training dataset.

References

[1] T. Gneiting, *Calibration of medium-range weather forecasts.* European Centre for Medium-Range Weather Forecasts Reading, UK, 2014.

[2] H. R. Glahn and D. A. Lowry, "The use of model output statistics (mos) in objective weather forecasting," *Journal of Applied Meteorology and Climatology*, vol. 11, no. 8, pp. 1203–1211, 1972.

[3] D. S. Wilks, *Statistical methods in the atmospheric sciences*, vol. 100. Academic press, 2011.

[4] J. Flowerdew, "Calibration and combination of medium-range ensemble precipitation forecasts," *Met Office Forecasting Research Tech. Rep*, vol. 567, p. 21, 2012.

[5] S. Vannitsem, J. B. Bremnes, J. Demaeyer, G. R. Evans, J. Flowerdew, S. Hemri, S. Lerch, N. Roberts, S. Theis, A. Atencia, *et al.*, "Statistical postprocessing for weather forecasts–review, challenges and avenues in a big data world," *Bulletin of the American Meteorological Society*, pp. 1–44, 2020.

[6] T. D. Hewson and F. M. Pillosu, "A new low-cost technique improves weather forecasts across the world," *arXiv preprint arXiv:2003.14397*, 2020.

[7] ECMWF, *The ECMWF Ensemble Prediction System.* European Centre for Medium-Range Weather Forecasts Reading, UK, 2012.

[8] B. Owen, B. Trotta, J. Liu, T. Gale, A. Schubert, and G. Weymouth, “Rainforests: A novel machine learning approach to calibrating rainfall forecasts,” tech. rep., Copernicus Meetings, 2022.

[9] M. S. Singapore, *2020 Annual Climate Assessment*. Meteorological Service Singapore, 2021.