Table of Contents

§1. Introduction

Future events such as weather forecasts and other meteorological predictions often have several possible prediction algorithms, each of which are likely to result in differing results (even if these results are close to each other, slight changes in future forecasts can have far-reaching implications). It is therefore necessary to take probabilities, or estimates, of the generated probabilities too.

Of course, the simplest way to achieve this would be rudimentary statistical measures such as the mean or the median. While these measures may work for small scale operations, for applications such as future forecasts in which the results generated can and will lead to widespread behavioral changes, more sophisticated tools should be used. The focus of this project is therefore the component resampling method described by Michael Dettinger.

When there are only a handful of data sets from which to draw conclusions, Dettinger's method provides a means of estimating probability distributions by expounding upon the set in a manner analogous to extrapolation and regression (note that this is not an exact description, but more accurately a tangentially related method mentioned for the purposes of easier illustration). The method makes use of principle component analysis (PCA) to divide the data into separate components, which are then redistributed in order to create a larger prediction set. This new set follows the trends of the original data but is now large enough to allow meaningful conclusions to be drawn, thus giving more detailed results than the simpler approaches such as taking the mean and standard deviations. Through a purely mathematical lens, this method also provides example applications of singular value decomposition (SVD) and generalised singular value decomposition (GSVD).

Now, predicting the likelihood of events involves creating their probability distribution functions (PDFs), which is simply the measure of the probabilities of obtaining certain outcomes from an outcome set. In order to expand the prediction set, Dettinger's method says to decompose each element of the set into component vectors, after which they can be recombined and solved using linear algebraic methods.

The purpose of Dettinger's method is to construct more data samples given a small set of data to obtain a more representative probability distribution function of the data. In other words, the procedures introduced by Dettinger serve to interpolate additional "forecasts" from a given ensemble of forecasts to predict the weather more accurately.

A set of forecasts about a measurable element of weather, such as temperature and humidity, can be represented by matrices. These matrices contain the time series data for given forecasts as columns. For example, the following 30×6 matrix $T_{m \times n}$ represents six different forecasts of the Maximum Daily Temperature in thirty days, where t_i^j represents the j^{th} forecast for the Max Daily Temperature in the i^{th} day.

$$T_{m \times n} = \begin{pmatrix} t_1^1 & t_1^2 & \dots & t_1^6 \\ t_2^1 & t_2^2 & \dots & t_2^6 \\ \vdots & \vdots & \ddots & \vdots \\ t_{30}^1 & t_{30}^2 & \cdots & t_{30}^6 \end{pmatrix}$$

Here, every row represents the temperature obtained each day, while every column represents different forecasts of the maximum temperature within a given timeframe of 30 days.

More generally, n different forecasts of the same set of m events can be characterized as an $m \times n$ matrix below.

$$X_{m\times n} = \begin{pmatrix} x_1^1 & x_1^2 & \cdots & x_1^n \\ x_2^1 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \ddots & \vdots \\ x_m^1 & x_m^2 & \cdots & x_m^n \end{pmatrix}$$

§1.1 A Simple Guide to the Dettinger's Method

1. Obtain the mean of values at each time i across all forecasts.

$$\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_i^j$$

2. Obtain the standard deviation of the given forecasts for each time i.

$$s_i = \left[\frac{1}{n-1}\sum_{j=1}^{n}(x_i^j - \bar{x}_i)^2\right]^{1/2}$$

   Standardize the given ensemble of forecasts to obtain a new matrix X' by subtracting the mean (to centralize the rows along zero) and dividing the difference by the standard deviation.

$$x'^j_i = \frac{(x_i^j - \bar{x}_i)}{s_i} \quad \text{for } 1 \le j \le n \text{ and } 1 \le i \le m$$

3. Compute the $m \times m$ Cross Correlation Matrix, $C = \frac{1}{n-1}X'X'^T$ such that:

$$c_r^s = \frac{1}{n-1}\sum_{j=1}^{n} x'^j_r x'^j_s \quad \text{for } 1 \le r \le m \text{ and } 1 \le s \le m$$

   This summarizes the covariance of each forecast with itself and other forecasts at each time step. Notice that this results in a symmetric matrix.

4. Construct the matrix E to contain columns that correspond to the eigenvectors of the columns of C.

   Construct coefficient matrix P by setting the columns of P equal to the projections of X' onto the vectors of E. The k^{th} column of p can be calculated by:

$$p^k = X'^T e^k \quad \text{for } 1 \le k \le m.$$

5. Construct additional "forecasts" by multiplying E with a random collection of corresponding coefficients in P.

$$r'^{l}_{i} = \sum_{s=1}^{m} e_i^s p_{random(j)}^s$$

6. Rescale each new “forecast” back by multiplying it with the standard deviation and adding the original mean. This is effectively the reverse of step 2.

$$r_i^l(rescaled) = r_i^l s_i + \bar{x}_i$$

By applying the following steps to a small sample of weather forecasts (Figure 1), a larger sample of interpolated forecasts is generated (Figure 2).

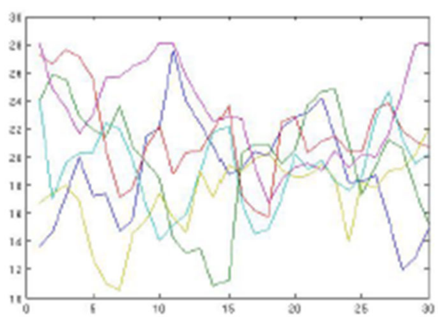

Figure 1

After Dettinger’s Method

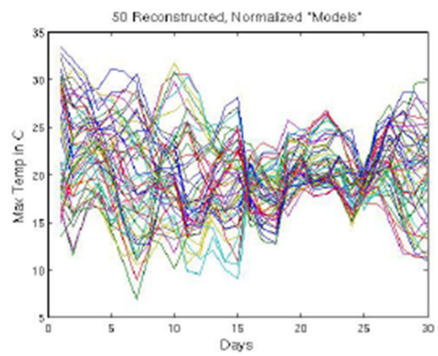

Figure 2

§2. Small Example

To illustrate the workings of the algorithm and help visualize Dettinger's method, a simple 5x3 matrix, X, will be used. This can be understood as a collection of temperatures collected by 3 different weather stations over 5 consecutive days. Using the function mean(X,2) in MATLAB will yield a 5x1 matrix, M, containing the mean temperatures for each of the days while the function std(X,0,2) will yield a 5x1 matrix, S, containing the standard deviation for each day.

Let $X = \begin{pmatrix} 30 & 32 & 31 \\ 29 & 27 & 30 \\ 32 & 33 & 33 \\ 29 & 29 & 30 \\ 31 & 32 & 31 \end{pmatrix}$. Then, $M = \begin{pmatrix} 31.0000 \\ 28.6667 \\ 32.6667 \\ 29.3333 \\ 31.3333 \end{pmatrix}$ and $S = \begin{pmatrix} 1.0000 \\ 1.5275 \\ 0.5774 \\ 0.5774 \\ 0.5774 \end{pmatrix}$

With X, a new matrix N can be found by subtracting the mean to produce rows of centered forecasts and dividing the result by the standard deviation. However, instead of having to calculate both the mean and standard deviation manually, the MATLAB function normalize(X,2) can be used to get N.

$$Thus, N = \begin{pmatrix} -1.000 & 1.000 & 0 \\ 0.2182 & -1.0911 & 0.8729 \\ -1.1547 & 0.5774 & 0.5774 \\ -0.5774 & -0.5774 & 1.1547 \\ -0.5774 & 1.1547 & -0.5774 \end{pmatrix}$$

After finding N, the cross-correlation matrix, C, can then be found with the formula: $C = \left(\frac{1}{n-1}\right) * N * N'$ where n is the number of columns in the matrix.

$$C = \begin{pmatrix} 1.0000 & -0.6547 & 0.8660 & 0 & 0.8660 \\ -0.6547 & 1.0000 & -0.1890 & 0.7559 & -0.9449 \\ 0.8660 & -1.890 & 1.0000 & 0.5000 & 0.5000 \\ 0 & 0.7559 & 0.5000 & 1.0000 & -0.5000 \\ 0.8660 & -0.9449 & 0.5000 & -0.5000 & 1.0000 \end{pmatrix}$$

This will be followed by using the following formula: $[E, D] = eig(C)$ to obtain E, the matrix of right eigenvectors of the columns and D, the diagonal of the corresponding eigenvalues.

$$E = \begin{pmatrix} 0.0966 & 0.6610 & 0.4449 & -0.3059 & 0.5121 \\ -0.5129 & 0.5362 & -0.3061 & -0.3059 & -0.5121 \\ 0.2890 & -0.1297 & -0.6583 & -0.5997 & 0.3265 \\ -0.1509 & -0.4546 & 0.5172 & -0.6696 & -0.2340 \\ -0.7883 & -0.2283 & -0.0867 & 0.0699 & 0.5605 \end{pmatrix}, D = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.8453 & 0 \\ 0 & 0 & 0 & 0 & 3.1547 \end{pmatrix}$$

With E, the coefficient matric can then be found with the following formula: $P = N' * E$

$$P = \begin{pmatrix} 0 & 0 & 0 & 1.2779 & -1.1894 \\ 0 & 0 & 0 & 0.1489 & 2.0416 \\ 0 & 0 & 0 & -1.4267 & -0.8522 \end{pmatrix}$$

To generate "new" data from the set of data, a loop can be used to get the coefficient matrix for any number of times. Here, the loop is run 5 times to get P2:

$$P2 = \begin{pmatrix} 0 & 0 & 0 & 1.2779 & -0.8522 \\ 0 & 0 & 0 & -1.4267 & -0.8522 \\ 0 & 0 & 0 & 0.1489 & -0.8522 \\ 0 & 0 & 0 & 0.1489 & -1.1894 \\ 0 & 0 & 0 & 0.1489 & 2.0416 \end{pmatrix}$$

Lastly, the formula $N2 = E * P2'$ is used to find a 5 x 5 matrix of the “new” normalized data, and each column can be regarded as a different weather station. This data can then be de-normalized with a simple for loop (code is attached in appendix) leading to the following matrix:

$$Final\ Matrix = \begin{pmatrix} 30.1727 & 31.000 & 30.5180 & 30.3453 & 32.0000 \\ 28.7362 & 30.000 & 29.2638 & 29.5275 & 27.0000 \\ 32.0636 & 33.000 & 32.4545 & 32.3909 & 33.0000 \\ 28.9545 & 30.000 & 29.3909 & 29.4364 & 29.0000 \\ 31.1091 & 31.000 & 31.0636 & 30.9545 & 32.0000 \end{pmatrix}$$

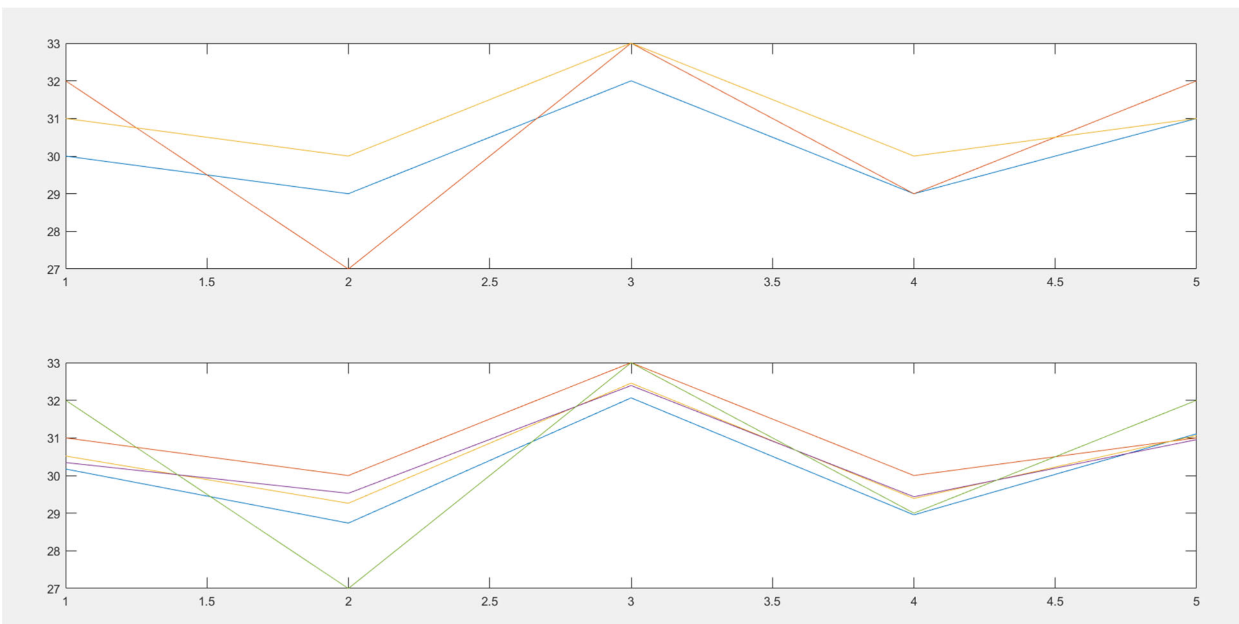

Figure 3: Before (top) and after (bottom) using Dettinger’s method on a small 5x3 matrix

As shown from the above graphs, the algorithm helps to “create” data around the average, allowing the extremes to carry less weight.

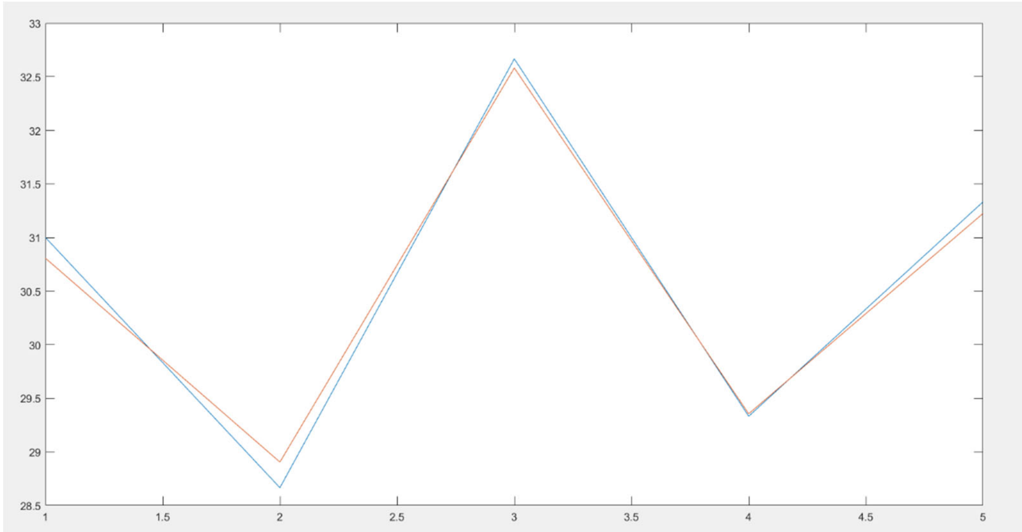

Figure 4: The means of the 2 different data sets

In Figure 4, the blue line traces out the mean temperatures over the 5 days. On the other hand, the orange line traces out the mean temperatures based on the data set after going through Dettinger's method. As shown in the diagram, just a small increase of 2 data sets allowed for the peaks to be smaller, signaling that this algorithm allows for anomalies to carry less weight even in situations where data sets are limited. In the case of weather prediction, this 'smoothening' of peaks allows for less fluctuations in the predicted weather even if anomalous behavior is present in the data sets.

§3. Real-Life Simulation

§3.1 Singapore Minimum Temperature Forecast for the next 7 days (as of March 31, 2021)

Table 1: Singapore Minimum Temperature Forecast (March 31 – April 6)

Source	March 31	April 1	April 2	April 3	April 4	April 5	April 6
NEA[1]	24	24	24	24	24	24	24
Weather Atlas[2]	26	26	26	25	25	26	26
Tides for Fishing[3]	27	28	21	26	26	27	27
MyWeather2[4]	29	27	26	26	26	27	26
Weather Zone[5]	23	24	24	24	24	25	26
AccuWeather[6]	26	26	26	25	26	26	26
CNA[7]	24	24	24	24	24	24	24
El Dorado Weather[8]	27	27	26	27	27	27	27

[1] https://www.nea.gov.sg/weather#weather-forecast2hr
[2] https://www.weather-atlas.com/en/singapore/singapore-long-term-weather-forecast
[3] https://tides4fishing.com/as/singapore/singapur/forecast/temperature
[4] https://www.myweather2.com/City-Town/Singapore/Singapore/14-Day-Forecast.aspx
[5] http://weather.themercury.com.au/world/southeast-asia/singapore/singapore
[6] https://www.accuweather.com/en/sg/singapore/300597/morning-weather-forecast/300597?day=1
[7] https://www.channelnewsasia.com/news/weather
[8] https://www.eldoradoweather.com/forecast/singapore/Singapore/Singapore.php

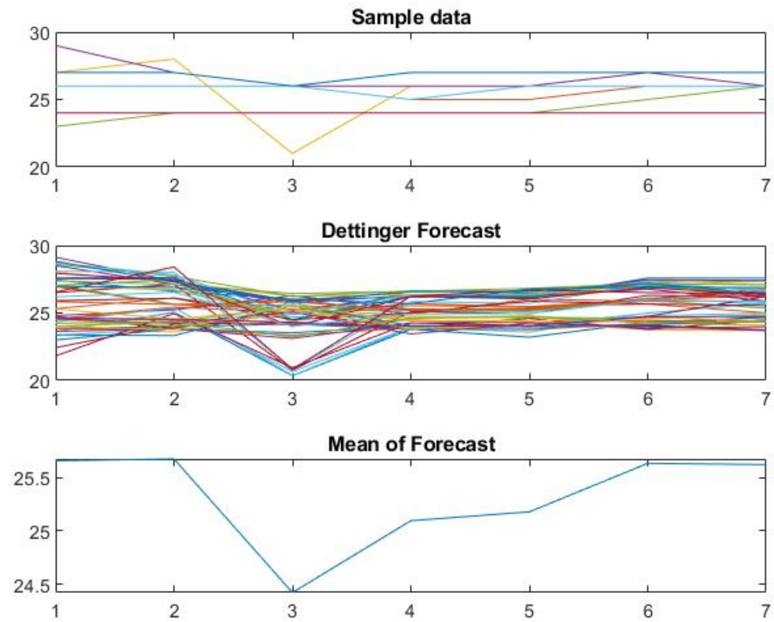

Figure 5: Graphs for Forecast of Singapore Minimum Temperature

Table 2: Our Prediction for Singapore Minimum Temperature (March 31 – April 6)

Date	March 31	April 1	April 2	April 3	April 4	April 5	April 6
Our Prediction (50 Predictions)	25.7	25.7	24.4	25.1	25.2	25.6	25.6

Table 3: Actual Singapore Minimum Temperature (March 31 – April 6)

Date	March 31	April 1	April 2	April 3	April 4	April 5	April 6
Actual Minimum Temperature [9]	24	26	25	25	25	24	25
Percentage Error	6.61	1.17	2.46	0.40	0.79	6.25	2.34

$$Average\ Percentage\ Error = 2.86\%$$

[9] https://www.wunderground.com/history/weekly/sg/singapore/WSSS/date/2021-3-31

§3.2 Singapore Maximum Temperature Forecast for the next 7 days (as of March 31, 2021)

Table 3: Singapore Maximum Temperature Forecast (March 31 – April 6)

Source	March 31	April 1	April 2	April 3	April 4	April 5	April 6
NEA[10]	34	34	33	33	34	34	34
Weather Atlas[11]	31	33	33	31	32	31	32
Tides for Fishing[12]	30	30	30	30	30	29	31
MyWeather2[13]	33	33	34	34	33	33	34
Weather Zone[14]	33	34	34	33	33	32	32
AccuWeather[15]	32	33	33	32	32	32	33
CNA[16]	33	34	34	33	33	34	34
El Dorado Weather[17]	30	31	31	31	30	30	31

[10] https://www.nea.gov.sg/weather#weather-forecast2hr
[11] https://www.weather-atlas.com/en/singapore/singapore-long-term-weather-forecast
[12] https://tides4fishing.com/as/singapore/singapur/forecast/temperature
[13] https://www.myweather2.com/City-Town/Singapore/Singapore/14-Day-Forecast.aspx
[14] http://weather.themercury.com.au/world/southeast-asia/singapore/singapore
[15] https://www.accuweather.com/en/sg/singapore/300597/morning-weather-forecast/300597?day=1
[16] https://www.channelnewsasia.com/news/weather
[17] https://www.eldoradoweather.com/forecast/singapore/Singapore/Singapore.php

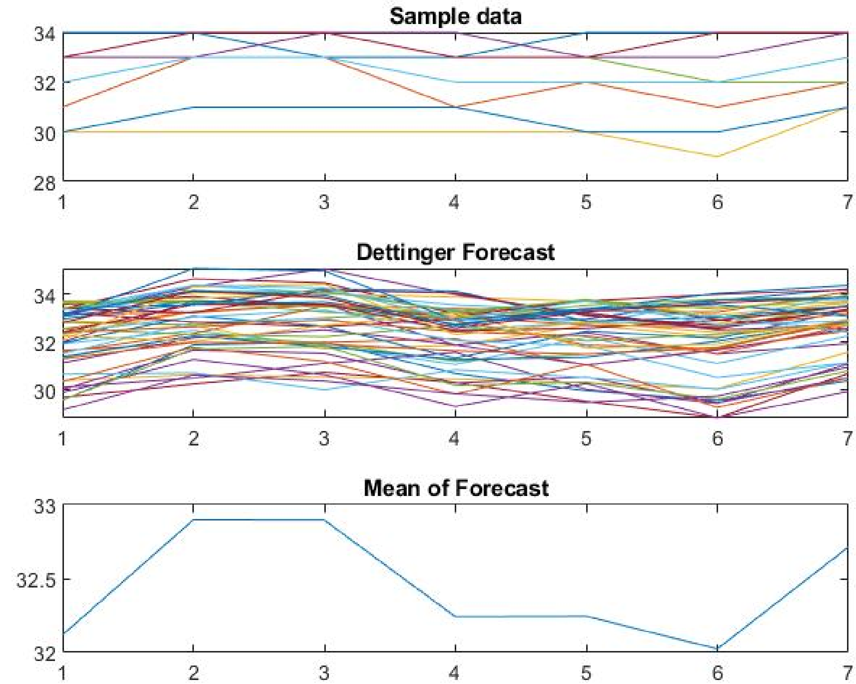

Figure 6: Graphs for Forecast of Singapore Maximum Temperature

Table 4: Our Prediction for Singapore Maximum Temperature (March 31 – April 6)

Date	March 31	April 1	April 2	April 3	April 4	April 5	April 6
Our Prediction (50 Predictions)	32.1	32.9	32.9	32.2	32.2	32.0	32.7

Table 5: Actual Data for Singapore Maximum Temperature (March 31 – April 6)

Date	March 31	April 1	April 2	April 3	April 4	April 5	April 6
Actual Maximum Temperature[18]	33	34	34	33	29	32	32
Percentage Error	2.80	3.34	3.34	2.48	9.94	0.00	2.14

$$Average\ Percentage\ Error = 3.44\%$$

[18] https://www.wunderground.com/history/weekly/sg/singapore/WSSS/date/2021-4-4

§3.3 Singapore Probability of Precipitation for the next 7 days (as of April 5, 2021)

Table 5: Singapore Precipitation Probability Forecast (April 5 – April 11)

Source	April 5	April 6	April 7	April 8	April 9	April 10	April 11
Tides 4 Fishing[19]	91	71	88	84	85	80	88
Weather Atlas[20]	89	76	73	74	75	73	72
The Weather Channel[21]	85	81	93	72	70	72	47
World Weather Online[22]	91	71	88	77	85	80	88
Weather Zone[23]	90	90	90	90	80	80	90
AccuWeather[24]	62	44	57	52	40	6	45
Time and Date[25]	75	67	66	64	67	61	60
El Dorado Weather[26]	70	17	40	17	17	17	17

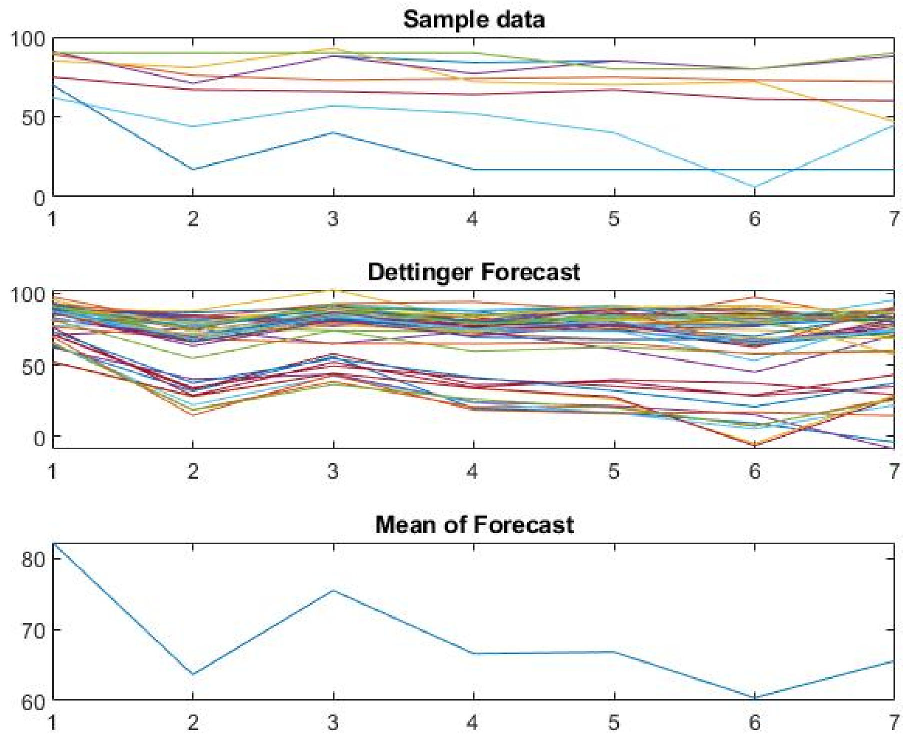

Figure 7: Graphs for Forecast of Singapore Maximum Temperature (April 5 – April 11)

[19] https://tides4fishing.com/as/singapore/singapur/forecast/chance-rain (Highest probability of precipitation)
[20] https://www.weather-atlas.com/en/singapore/singapore-long-term-weather-forecast
[21] https://weather.com/weather/tenday/l/eeda64fb7fd9ede722e5fd460307a636447fbf5ba1f226e8bf2ffa707e59dc53#detailIndex5
[22] https://www.worldweatheronline.com/singapore-weather/sg.aspx
[23] http://weather.themercury.com.au/world/southeast-asia/singapore/singapore
[24] https://www.accuweather.com/en/sg/singapore/300597/daily-weather-forecast/300597
[25] https://www.timeanddate.com/weather/singapore/singapore/ext
[26] https://www.eldoradoweather.com/forecast/singapore/Singapore/Singapore.php

Table 6: Our Prediction for Singapore Precipitation Probability Forecast (April 5 – April 11)

Date	April 5	April 6	April 7	April 8	April 9	April 10	April 11
Our Prediction (50 Predictions)	82.2	63.6	75.5	66.6	66.8	60.4	65.6
Actual Precipitation[27] (Rained for 2 hours or more)	Yes	Yes	Yes	Yes	Yes	No	No

$$Average\ Number\ of\ Rainy\ Days^{28} = 4.80\ Days$$

$$Actual\ Number\ of\ Rainy\ Days = 5\ Days$$

$$Percentage\ Error = 4.17\%$$

[27] https://www.timeanddate.com/weather/singapore/singapore/historic

[28] Calculated by taking the sum of probabilities of rainfall

§3.4 Singapore Humidity for the next 7 days (as of April 5, 2021)

Table 7: Singapore Humidity Forecast (April 5 – April 11)

Source	April 5	April 6	April 7	April 8	April 9	April 10	April 11
TimeandDate[29]	75	67	66	64	67	61	60
The Weather Channel[30]	78	74	76	76	75	74	72
Weather Atlas[31]	89	76	73	74	75	73	72
Weather-forecast.com[32]	77	73	72	71	73	70	71
Weather2[33]	81	76	78	72	77	76	76

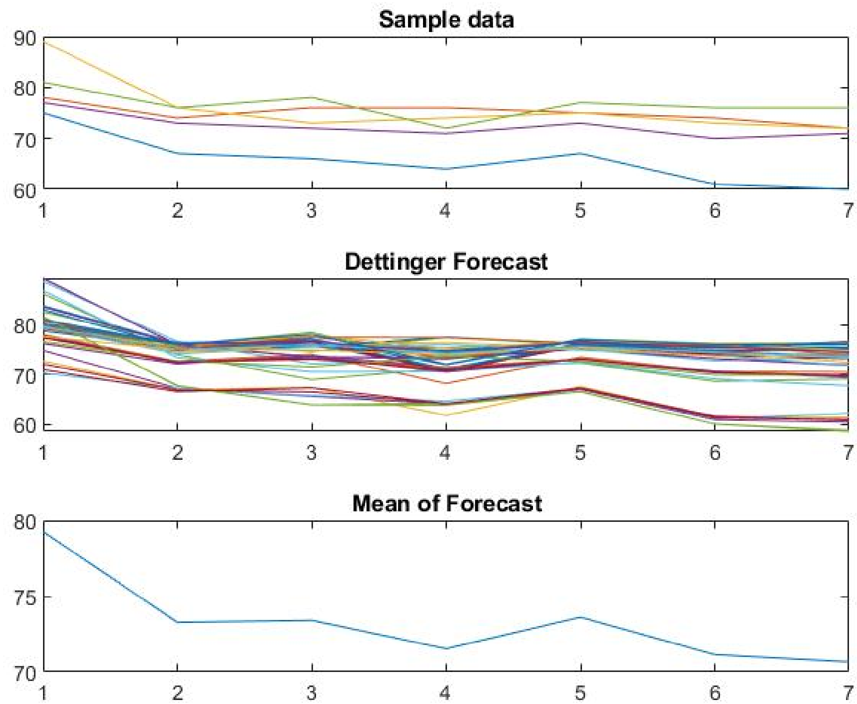

Figure 8: Graphs for Forecast of Singapore Humidity (April 5 – April 11)

29 https://www.timeanddate.com/weather/singapore/singapore/ext

30 https://weather.com/weather/tenday/l/eeda64fb7fd9ede722e5fd460307a636447fbf5ba1f226e8bf2ffa707e59dc53#detailIndex5

31 https://www.weather-atlas.com/en/singapore/singapore-long-term-weather-forecast

32 https://www.weather-forecast.com/locations/Singapore/forecasts/latest (Average Humidity)

33 https://www.myweather2.com/City-Town/Singapore/Singapore/14-Day-Forecast.aspx

Table 8: Our Prediction for Singapore Humidity Probability Forecast (April 5 – April 11)

Date	April 5	April 6	April 7	April 8	April 9	April 10	April 11
Our Prediction (50 Predictions)	79.2	73.2	73.4	71.5	73.6	71.1	70.7

Table 9: Actual Data for Singapore's Humidity Probability (April 5 – April 11)

Date	April 5	April 6	April 7	April 8	April 9	April 10	April 11
Actual Data[34] (Average Humidity)	88	83.75	71.5	77.5	76.3	71	73.3
Percentage Error	11.11	14.41	2.59	8.39	3.67	0.14	3.68

$$Average\ Percentage\ Error = 6.28\%$$

§3.5 Analysis of Real-Life Simulation Data

As seen from the average percentage errors of our predictions as compared to the actual weather recordings, the percentage errors are within 10% from the actual recordings which indicate that our predictions are relatively accurate over a period of 7 days.

However, it is important to note that the range for day-to-day percentage errors across all types of weather predictions is considerably large. For example, the percentage error for humidity ranges from 14.41% (April 6) to 0.14% (April 10).

[34] https://www.timeanddate.com/weather/singapore/singapore/historic?month=4&year=2021

§4. Limitations and Future Work

§4.1 Limitations

One of the main limitations of Dettinger's method is the accuracy or the lack thereof of each prediction set. Even though each source has a different accuracy rate when it comes to the prediction of the weather, this method fails to take the accuracy rate into account as none of the sets are weighted. A study by Boisramé (2010)[35] found that Dettinger's Method may not consistently provide reliable distribution estimates due to the noise present in the original data. Dettinger's method can amplify the effects of the existing noise (especially for heavily skewed and bimodal distributions) and possibly reduce the accuracy of the "new" forecasts formed. With the limited sources of weather prediction datasets, the issue of forming inaccurate forecasts could be further exacerbated. The limited number of values that can be chosen at random increases the possibility of choosing inaccurate data, causing the final prediction to be heavily influenced by the inaccurate data point. One way to mitigate this drawback is to apply the Singular Value Decomposition as explained in the *Future work* section.

Moreover, as the coefficients generated from the 'real-life data' values are randomly chosen to generate "new" data, some "new" forecasts may overlap as well. As such, one consideration to prevent overlapping is to limit the number of "new" forecasts generated or by removing repeated columns in randomly generated coefficient matrix P_{random}. Although in this report, we are unable to evaluate the statistical accuracy of the method in producing predictions, the method also effectively achieves its purpose: to create a large array of data given only a small sample size.

§4.2 Future work

More studies can be done to integrate the use of Singular Value Decomposition (SVD) on top of Dettinger's Method. In SVD, eigenvectors in matrix E, which contains eigenvectors of the cross-correlation matrix, are normalized and arranged in order based on the size of their corresponding eigenvalues. The new matrix U, formed with the normalized eigenvectors, provide a basis for the columns of data. We can limit the number of columns in matrix U to be less than the rank of the original data matrix to ensure that noise from the original data is removed. The highest-numbered columns are generally considered as noise.

As technology progresses however, it might be possible for the accuracy of various weather prediction sources to be quantified and weighted as well before using the Dettinger's Method. This might be achieved by duplicating the prediction sources with a higher accuracy rate in the original matrix, X. By doing so, the possibility of the more accurate value being chosen increases, allowing the final predicted values generated to be more accurate.

[35] Boisramé, G. F. S. (2010, May 12). The use of linear algebra in modeling the probabilities of predicted future occurrences. Arminda. https://arminda.whitman.edu/theses/57.

§5. Conclusion

Given the unpredictability of weather, which in a real-life situation, could be affected by a multitude of factors such as the unpredictable flow of air due to fluid dynamics, power storms, and movements between the layers of the cold fronts that could possibly result in the weather changing in unforeseen ways. Coupled with the difficulties in measuring and analysing large amounts of meteorological data, this could thus result in inaccurate weather predictions. Therefore, due to the fundamental limitations that plague weather prediction, the effectiveness of the Dettinger’s method in weather prediction is only as effective as the original predictions of the weather.

Appendix

MATLAB Code:

```
function [N2,A] = detting(X, n )
% Outputs the Dettinger's forecast
%   X is the input matrix, n is the number of forecast, N2 is the forecast
%   matrix and A is the mean of forcast.

M=mean(X,2); %matrix of means
S2=std(X,0,2); %Population Standard div
N=normalize(X, 2); %Normalizes Data
C=(1/(n-1))*N*(N'); %Cross Correlation Matrix
[E,D]=eig(C); %obtain E, a matrix of right eigenvectors of columns, and D the diagonal of
corresponding eigenvalues
P=(N')*E; %coefficient matrix
P2=[];

for x=1:n
   b=[];
   for i=1:(size(X,1))
      b=[b P(randi([1 (size(X,2))]),i)];
   end
   P2=[P2;b];
end % For Loop to get coefficient matrix for n new “forecasts”
N2 = E*(P2'); %matrix of new normalized “forecasts”
for i = 1:(size(X,1))
   N2(i,:)=arrayfun(@(x) x*S2(i,1)+M(i,1),N2(i,:));
end %For Loop to denormalize the matrix to original values centred around original mean and
distributed along an original standard deviation

%Plots the Forecast on a graph
A=mean(N2,2);
tiledlayout(3,1);
nexttile;
plot(X);
title('Sample data')
nexttile;
plot(N2);
title('Dettinger Forecast')
nexttile;
plot(A);
title('Mean of Forecast')


end
```