

# TRANSLATING THE WORDS AND NUMBERS OF OFFICIAL FORECASTS INTO PREDICTIONS OF AMOUNT AND PROBABILITY OF PRECIPITATION

Harvey Stern<sup>1</sup>

## 1. INTRODUCTION

Australian Bureau of Meteorology day-to-day weather forecasts are presented in a format that comprises an extended worded description of the expected weather, a brief précis which summarises that description, estimates of the anticipated minimum temperature, maximum temperature, probability of precipitation, and amount of precipitation (Plate 1).

The latter weather element is expressed as a range - for example, 5 - 10 mm, unless no precipitation is considered possible, in which case, 0 mm is given (Plate 2).

## 2. METHODOLOGY

Using an 11-year data-set (Aug-05 to Jul-16) data base of official précis forecasts for Melbourne (Map 1), and a Sep-11 to Jul-16 data base of official estimates of the anticipated probability and amount of precipitation, statistical relationships are established between the words and numbers contained in the official forecasts and what eventuates in terms of the likelihood and amount of precipitation.



**Map 1** Location of Melbourne

Source: <http://www.ga.gov.au/placename>

-----  
<sup>1</sup>School of Earth Sciences,  
University of Melbourne,  
Parkville, Australia  
[hstern@unimelb.edu.au](mailto:hstern@unimelb.edu.au)

## 3. RESULTS

### 3.1 Amount of Precipitation

The image on the **left hand side** of Figure 1 illustrates the multiple linear relationship between the words used, and what eventuated in terms of the amount of precipitation observed (expressed as the square root of the precipitation amount).

The first column of the left hand side of Figure 1 lists the words, placed in order of the corresponding *t statistic*, which is given in the second column. It may be seen that the most positively related words (in order) are:

- RAIN;
- SHOWERS;
- SHOWER;
- HEAVY; and,
- THUNDER.

The most negatively related words (in order) are:

- LITTLE;
- FEW;
- CHANCE;
- CLEARING; and,
- LATE.

The third column of the left hand side of lists the probabilities that the sign of the corresponding *t statistic* (whether the sign was positive or negative) was not of the direction indicated.

These probabilities are rounded to three decimal places, and those significant at the 1% level or better (most of them) are highlighted with a reddish-pink colour. The third column lists the corresponding regression coefficients.

We shall now consider the relationship between what eventuated in terms of the  $\sqrt{\text{amount}}$  of precipitation that occurred and what was suggested by:

- the *upper* official estimate;
- the *lower* official estimate; and,
- $\sqrt{\text{amount}}$  suggested by the equation depicted in Figure 1 by the coefficients listed in the fourth column of the left hand side of the Figure (this equation relates  $\sqrt{\text{amount}}$  of precipitation to the words utilised in the corresponding official forecast).

This relationship is:

$$\sqrt{\text{amount}} = -0.024 + 0.220*\text{lower} + 0.459*\text{upper} + 0.262*\text{words} \dots(\text{Eqn 1})$$

### 3.2 Probability of Precipitation

The image on the *right hand side* of Figure 1 illustrates the multiple linear relationship between the words used, and what eventuated in terms of the likelihood (that is, *probability*) of precipitation occurring.

For this purpose, we shall firstly set:

- an observation of *Nil* at 0%;
- an observation of 0.2 mm at 50% (taking into account that 0.2 mm may reflect condensation on account of dew rather than rainfall); and,
- an observation of greater than 0.2 mm is set at 100%.

The first column of of the right hand side of Figure 1 lists the words, placed in order of the corresponding *t statistic*, which is given in the second column. It may be seen that the most positively related words (in order) are:

- RAIN;
- SHOWER;
- SHOWERS;
- DRIZZLE; and,
- THUNDER.

The most negatively related words (in order) are:

- LITTLE;
- CHANCE;
- CLEARING;
- LATE; and,
- FINE.

The third column of of the right hand side of Figure 1 lists the probabilities that the sign of the corresponding *t statistic* (whether the sign was positive or negative) was not of the direction indicated.

These probabilities are rounded to three decimal places, and those significant at the 1% level or better (most of them) are highlighted with a reddish-pink color.

The fourth column lists the corresponding regression coefficients.

We shall now consider the relationship between what eventuated in terms of the *probability* of precipitation and what was suggested by:

- the *official* estimate;
- the *probability* suggested by the equation depicted in Figure 1 by the coefficients listed in the fourth column of the right hand side of the Figure (this equation relates *probability* of precipitation to the *words* utilised in the corresponding official forecast).

This relationship is:

**probability =-**

$$6.634 + 0.765*\text{official} + 0.437*\text{words}$$

.....(Eqn 2)

### 3.3 Significance Levels

With regard to the most positively related words for amount and probability, the four words RAIN, SHOWER, SHOWERS and THUNDER are included among the five most significant words in both cases.

However, the word HEAVY only is included for amount, in which case the *t statistic* is 9.0, reflecting an extremely high level of significance. The corresponding *t statistic* for probability is only 1.0, reflecting negligible significance.

By contrast, with the word DRIZZLE, which only is included for probability, where the *t statistic* is 6.4, reflecting a very high level of significance. Even though DRIZZLE is not included for amount, its *t statistic* is, nevertheless, 5.6, which also suggests a very high level of significance.

With regard to the most negatively related words for amount and probability, the four words LITTLE, CHANCE, CLEARING and LATE are included among the five most significant words in both cases.

However, the word FEW only is included for amount, in which case the *t statistic* is -7.4, reflecting an extremely high level of significance. The corresponding *t statistic* for probability is -2.1, reflecting, nevertheless, a reasonable level of significance, albeit somewhat less.

The word FINE only is included for PROBABILITY, in which case the *t statistic* is -2.4, reflecting a moderate level of significance. Even though FINE is not included for amount, the corresponding *t statistic* for probability is -3.0, also reflecting moderate significance.

### 3.4 Trends in Accuracy

Figure 2(a), utilises the multiple linear relationship between the words used, and what eventuated in terms of the amount of precipitation observed (this is the relationship that is presented on the left hand side of Figure 1) in order to illustrate a measure of trends in the accuracy of how good an indication of precipitation amount is provided by the words in the official forecasts.

Similarly, Figure 2(b), utilises the multiple linear relationship between the words used, and what eventuated in terms of the amount of precipitation observed (this is the relationship that is presented on the right hand side of Figure 1) in order to illustrate a measure of trends in the accuracy of how good an indication of precipitation probability is provided by the words in the official forecasts.

Figure 3(a), utilises the multiple linear relationship between the words and numerical estimates used, and what eventuated in terms of the amount of precipitation observed (Eqn 1) in order to illustrate a measure of trends in the accuracy of how good an indication of precipitation amount is provided by the words in combination with the numerical estimates.

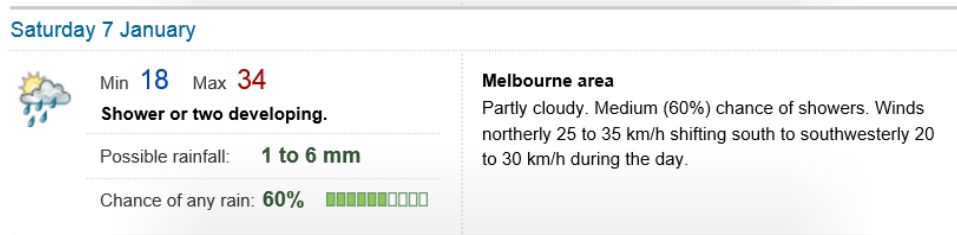
Figure 3(b), utilises the multiple linear relationship between the words and numerical estimates used, and what eventuated in terms of the amount of precipitation observed (Eqn 2) in order to illustrate a

measure of trends in the accuracy of how good an indication of precipitation probability is provided by the words in combination with the numerical estimates.

#### 4. SUMMARY

An extended data-base of official worded forecasts has been combined with corresponding data bases of observations, and also of official estimates of the anticipated probability and amount of precipitation. Statistical relationships have then been established between the various data bases:

- To indicate what words are most likely to suggest the likelihood and amount of precipitation;
- To provide a measure of trends in the accuracy of the worded forecasts taken separately; and,
- To provide a measure of trends in the accuracy of the worded forecasts taken in combination with the numerical estimates.



**Plate 1** Example of a forecast for Yarra Glen (a suburb of Melbourne located well to the east of the City) issued by the Bureau of Meteorology on 1st January 2017. <http://www.bom.gov.au/vic/forecasts/yarraglen.shtml>

#### What does the possible rainfall amount mean?

Possible rainfall: 5 to 10 mm

The possible rainfall amounts help to show how rainfall might vary according to the type of weather in a given time period. Both numbers relate directly to a chance of receiving at least that amount of rain.

The first number (5 mm in this example) represents a 50% chance of at least that amount of rain occurring.

The second number (10mm in this example) represents a 25% chance of at least that amount of rain occurring.

On days where we expect showers or thunderstorms, the possible rainfall amounts might be quite different, for example, 5 to 30 mm. When steady rainfall is expected over a wide area, the possible rainfall amounts might be similar, for example, 10 to 15 mm.

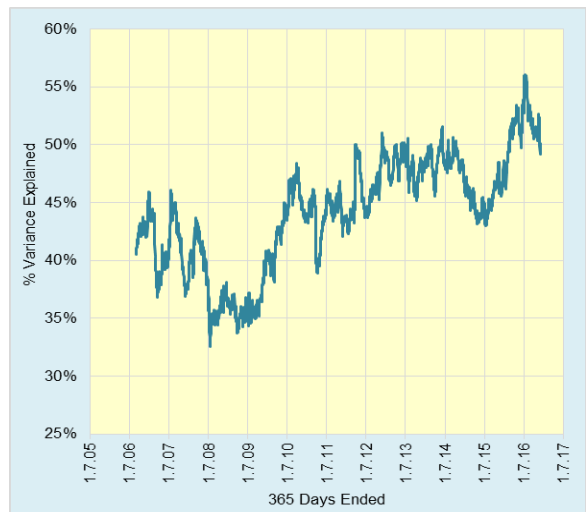
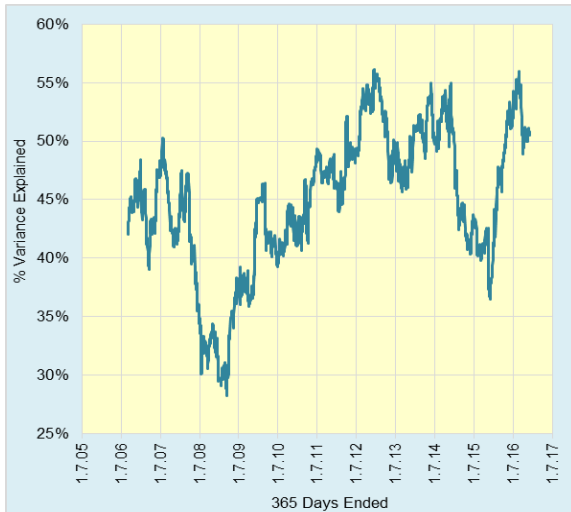
**Plate 2** A reference on the Bureau’s website explaining the meaning of the numbers utilised in the forecast possible rainfall. <http://www.bom.gov.au/NexGenFWS/rainfall-faq.shtml>

Word	t Statistic	Probability	COEFFICIENTS
RAIN	40.6	0.000%	2.106
SHOWERS	17.7	0.000%	0.947
SHOWER	13.1	0.000%	0.485
HEAVY	9.0	0.000%	2.847
THUNDER	7.3	0.000%	0.517
DRIZZLE	5.6	0.000%	0.520
EASING	4.4	0.000%	0.294
BECOMING	1.7	4.223%	0.097
FOG	-0.3	37.940%	-0.020
CLOUD	-1.7	4.776%	-0.073
FINE	-3.0	0.135%	-0.113
LATE	-3.6	0.018%	-0.144
CLEARING	-4.4	0.001%	-0.188
CHANCE	-5.0	0.000%	-0.267
FEW	-7.4	0.000%	-0.428
LITTLE	-12.2	0.000%	-0.931

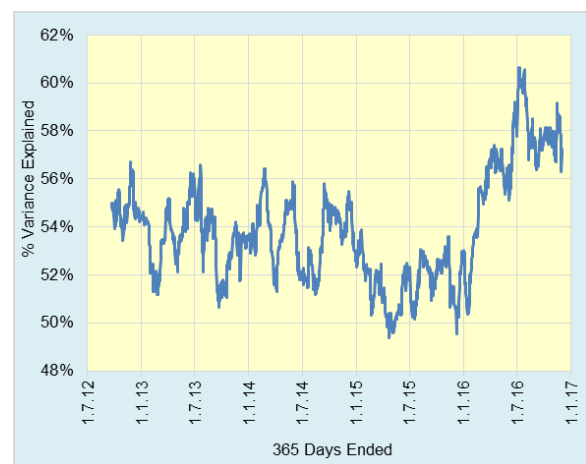
Word	t Statistic	Probability	COEFFICIENTS
RAIN	32.2	0.000%	74.807
SHOWER	20.7	0.000%	34.332
SHOWERS	14.1	0.000%	33.888
DRIZZLE	6.4	0.000%	26.354
THUNDER	4.9	0.000%	15.669
BECOMING	3.8	0.008%	9.541
EASING	1.3	10.103%	3.790
HEAVY	1.0	16.247%	14.004
CLOUD	-1.3	10.329%	-2.485
FOG	-1.6	5.181%	-4.728
FEW	-2.1	1.614%	-5.568
FINE	-2.4	0.793%	-4.073
LATE	-2.7	0.382%	-4.808
CLEARING	-2.9	0.190%	-5.544
CHANCE	-5.4	0.000%	-12.959
LITTLE	-7.9	0.000%	-26.815

**Figure 1** An illustration of the multiple linear relationship between the words used, and what eventuated in terms of the amount of precipitation (left hand side) and a corresponding illustration for the probability of precipitation (right hand side)



**Figure 2 (a)**...Left Hand Side: Trend in the % variance of the precipitation amount explained by the worded component of the official forecasts.

**Figure 2 (b)**...Right Hand Side: Trend in the % variance of the precipitation probability explained by the worded component of the official forecasts.



**Figure 3 (a)** Left Hand Side: Trend in the % variance of the precipitation amount explained by Eqn 1 through optimally combining upper and lower official quantitative estimates and the words utilized in the official worded forecasts.

**Figure 3 (b)** Right Hand Side: Trend in the % variance of the precipitation probability explained by Eqn 2 through optimally combining official quantitative estimates and the words utilized in the official worded forecasts.