REPORT

A STATISTICAL ANALYSIS OF THE BUNDABERG ATLANT TRIAL

Prepared for: ARC
Prepared by: Stephen Beare, Ray Chambers
Date: 16 March 2009

Concept Economics: 27 Jardine Street, PO Box 5430
Kingston
ACT 2604 Australia

CSSM: University of Wollongong
NSW 2522 Australia
# Table of contents

**EXECUTIVE SUMMARY**

1. **INTRODUCTION**

2. **THE ANALYTICAL FRAMEWORK**
   2.1. PARTITIONING THE TARGET AREA

3. **STATISTICAL ANALYSIS**
   3.1. ANALYSIS OF THE CONTROL AREAS (STAGES 1 - 3)
   3.2. THE CONTROL MODEL (STAGE 4)
   3.3. ANALYSIS OF THE TARGET AREA (STAGE 5)
   3.4. ESTIMATING ATLANT EFFECT (STAGE 6)

4. **CONCLUSIONS**

**APPENDIX A**

A.1. STATISTICAL ANALYSIS

**GLOSSARY**
EXECUTIVE SUMMARY

From January to May of 2008, a trial was conducted near Bundaberg, Queensland, of the Atlant rainfall enhancement system. Atlant is a ground-based ion generator designed to increase the concentration of charged or larger aerosols that are carried into the atmosphere through natural uplift, atmospheric turbulence and prevailing winds. In turn, these aerosols are thought to act to enhance droplet formulation and ultimately rainfall.

A primary purpose of the Bundaberg trial was to measure the rate of formation of ions and their dispersion from the Atlant device. During the conduct of the trial it became clear that the actual process of ion formation and movement was complex, and demonstrating a causal link between the operation of the Atlant system and rainfall would require a major scientific undertaking beyond the scope of the trial per se. However, demonstrating a correlation between Atlant and rainfall would be possible through statistical analysis.

It seems prudent that in the first instance a link between the operation of the Atlant system and rainfall should be established on the basis of statistical inference. As a part of the trial, field data were collected by the University of Queensland from a network of rainfall gauges that supplemented rainfall data collected by the Australian Bureau of Meteorology. These measurements were made in the target area (a 70 km radius circle around the Atlant site) and in two control areas, one to the north near Gladstone and one to the south near Gympie. Both control areas were chosen on the basis of having similar weather and rainfall as the target area. As a result, spatio-temporal analysis of the data could be and was conducted using daily gauge and wind direction measurements. The analysis was exploratory in nature with the objectives of:

• Establishing whether a clear signal for the presence and extent of an Atlant effect could be detected and if so, whether an estimate could be made of the potential resulting contribution (negative, positive or none) to the observed rainfall, and

• Informing the choice of experimental design and statistical methods for subsequent trials of the Atlant system.

To emphasise the non-causal nature of any findings of the approach adopted in this analysis, and the lack of any proven causal link between Atlant and rainfall, any such potential Atlant effect on rainfall will be referred to subsequently as PAE (Potential Atlant Effect).

The Statistical Approach

A key aspect of the statistical approach was the dynamic partitioning of the overall target area on the basis of prevailing steering wind direction. Given the postulated way in which the Atlant system affects precipitation, it would be logical to expect that the main PAE or footprint of the system would be downwind of the operating site, although a working definition of downwind was not established prior to the analysis of the trial data. Much of the exploratory analysis was therefore centred on what might be a good definition of such a downwind target area.
It is conceptually possible that the Atlant system could enhance rainfall by increasing either the probability that a rainfall event occurs or the expected level of rainfall when a rainfall event has occurred, or both. Either of these effects could depend on prevailing meteorological conditions. Both of these effects were examined.

Findings

The analysis showed that under the general meteorological conditions prevailing during the trial, there was an enhanced probability of a rainfall event, defined as rainfall being recorded on a given day, in the target area compared with the two control areas. However, since during the trial the Atlant system was only switched on under conditions that appeared to be conducive for the operation of the system to enhance rainfall, it was felt that this enhanced probability could not be directly associated with the operation of Atlant, particularly since there was no evidence of an increased probability of a rainfall event downwind of the Atlant site. This inconclusive result may reflect the fact that there was not enough data to adequately examine whether the operation of the Atlant system could be associated with the probability of a downwind rainfall event. It also suggests that if there was an experimental bias it was in relation to the target area as a whole as opposed to downwind of the Atlant site.

In contrast, when a rainfall event occurred within the target area, the operation of the Atlant system was associated with a significant increase in expected rainfall downwind of the Atlant device. Within a 30-degree arc extending downwind 70 kilometres from the Atlant device, rainfall was more than 24 per cent higher than would be expected given the controls specified in the statistical model. No significant differences in rainfall were observed in the balance of the target area.

Given the structure of the model a simple reliability analysis was conducted on the basis of the sampling distribution of the estimated coefficient relating to downwind rainfall. The results indicated that there was roughly a 90 per cent probability that rainfall was at least 15.5 per cent higher in the downwind arc described above and a 75 per cent probability that rainfall was at least 19 per cent higher than would be expected given the controls specified in the statistical model.

The potential for experimental bias in the trial cannot be ignored, particularly since the Atlant system was only switched on during the trial when conditions were thought to be favourable for rainfall enhancement. While it does not correct for this potential bias in a formal sense, a conservative approach to assessing the PAE on downwind rainfall enhancement during the trial can be adopted, based on the fact that the operation of the Atlant system was also associated with an expected increase in rainfall in the overall target area, excluding the 30-degree downwind arc, of approximately 7.2 per cent. While this increase may or may not be wholly attributable to the potential experimental bias referred to above, a conservative approach to assessing the overall PAE is to deduct it from the downwind estimate.

This adjusted PAE for the downwind sector is then 17.6 per cent with a 75th percentile lower bound of 11.8 per cent and a 90th percentile bound of 8.3 per cent.

Effect versus Efficacy

In dealing with large data sets the problem is often less with establishing a significant effect in a statistical sense and more with whether the effect is sufficient to consider that the cost of development and implementation will be exceeded by the benefits.
The on-ground environmental effect of the PAE rainfall can be calculated. The elevated levels of rainfall downwind of the Atlant site associated with the operation of the system resulted in an increase in rain volume in the Bundaberg region of approximately 135 GL. Again allowing for experimental bias, the adjusted volume of additional rainfall associated with the operation of the Atlant system is 95.8 GL.

The analysis indicates that further Atlant trials are warranted. However, given the statistical approach is complex and potentially unfamiliar to those without a background in spatial statistics, peer review of the analysis is necessary to support this claim. Experimental design insights gained from the Bundaberg trial should be incorporated into further trials. These are highlighted in this report.

The highly directional nature of the identified PAE suggests that testing whether the system is associated with an increase in the probability of rainfall event occurring also needs to be examined relative to prevailing wind conditions.

More generally, the methodology adopted in the present analysis will allow more robust and timely evaluation of rainfall enhancement technologies than has previously been reported.
1. INTRODUCTION

From January to May 2008, a trial was conducted near Bundaberg, Queensland, of the Atlant rainfall enhancement system. Atlant is a ground-based ion generator designed to increase the concentration of charged or larger aerosols that are carried into the atmosphere through natural uplift, atmospheric turbulence and prevailing winds. In turn, these aerosols are thought to act to enhance droplet formulation and ultimately rainfall.

The trial was funded by the Commonwealth Government and Australian Rain Technologies.

A primary purpose of the Bundaberg trial was to measure the rate of formation of ions and their dispersion from the Atlant installation. Field data were also collected by the University of Queensland from a network of rainfall gauges that supplemented rainfall data collected by the Australian Bureau of Meteorology (BoM). Gauges were placed to provide uniform geographic coverage.

The analysis described in this report is based on daily rainfall measurements taken from 165 rainfall gauges over the duration of the trial. These included 24 gauges in a north control area near Gladstone and 24 gauges in a south control area near Gympie. The target area was defined by a circle with a radius of 70 km centred about the Atlant site at Paradise Dam south-east of Bundaberg. Within the target area there were a total of 117 gauges. A summary of the UQ trial rainfall report is at Appendix B and includes further definition of the target and control areas (see Chart B1).

The University of Queensland conducted a statistical analysis based on average rainfall over the trial period compared to historical records for the same months. The historical data were stratified according to the presence of La Niña conditions because such conditions prevailed through the trial period. While the results described by this initial statistical analysis indicated that rainfall levels were elevated in the target area over the trial period, they were not considered conclusive.

Even if the results had generated a highly significant increase in average rainfall in the target area over the trial period, a claim that the operation of the Atlant system was associated with a period of enhanced rainfall could not be reasonably justified. This is largely due to two reasons:

- First, within a five-month time frame there will be significant geographic variation in average rainfall relative to the longer term mean; and
- Second, even in the longer term, average rainfall is subject to longer term climatic cycles and induced climate change.

In the light of these issues, it was considered important to re-analyse the control and experimental data sets to reconstruct conditions that begin to approach a real-time controlled experiment. Central to this approach is the level of spatial correlation in rainfall gauge data.

Spatial Correlation

Spatial correlation is an important factor to consider because it will determine whether rainfall in one location will be a good predictor of rainfall in a neighbouring location. More specifically, the level of spatial correlation will determine whether rainfall in the control areas is a good
predictor of rainfall in the target area. Clearly, the level of spatial correlation between geographic areas is not revealed in long term averages. This requires the comparison of rainfall over limited time intervals.

Spatial correlation can occur over different time intervals. For example, at an hourly time level, the spatial correlation between observed rainfalls may be much lower than what might be observed over a monthly or annual time interval. However, the fact that there is spatial correlation between daily rainfall values implies that concurrent observations in a control area provide information about the expected level of rainfall in the target area. That is, we have a measure approaching a real-time experimental control so long as the target and control are well defined. However, the effectiveness of this control depends on the level of spatial correlation.

The Basic Approach

The approach uses the spatial correlation between daily observations of rainfall at different locations, along with other explanatory variables referred to as covariates to predict rainfall. Two prediction models are constructed:

- A Control Model that does not take into account any factors that relate to the operation of the Atlant system, and predicts rainfall in the target area on the basis of historical rainfall as well as contemporaneous rainfall in the control areas.

- An Effects Model that includes, in addition to the covariates included in the Control Model, factors that relate to the operation of the Atlant system. These could include the length of time the system is on or off, the location of a gauge relative to the downwind direction from the Atlant site or the distance between the Atlant site and a gauge.

The difference between the two model predictions is attributed to those factors that relate to the operation of the Atlant system and, by inference, is attributed to the system.

An advantage of comparing predicted rainfall estimates is that it doesn’t simply attribute an observed change in the level of rainfall to the enhancement technology. Rather it compares what can be predicted with and without the technology, the unexplained variation in rainfall not making any contribution to the estimate.

Experimental Design and Bias

Ideally the experimental design of a field trial will eliminate most forms of bias, particularly the generation of a spurious relationship between observed rainfall and the operation of the Atlant system. This typically arises when the variables that are used to represent the operating characteristics of the system are correlated with some other factor or factors that are in fact the source of the observed differences in rainfall.

Experimental design at the scale of the Atlant trial will almost always present difficulties due to the large areas involved and the lack of homogeneity in topography that may give rise to micro-climatic effects. Storm cells can often move along relatively narrow tracks. Put simply, we would expect considerable increases in variation in rainfall gauge data as distance between the gauges increases. This will clearly be associated with a reduction in the level of spatial correlation between rainfall gauges which points to an advantage of a spatial statistics
approach. Measured spatial correlations will therefore give us a good idea of how effective a given experimental design will be at least in terms location of target and control areas.

It is also recognised that a number of elements of the trial were not ideal in terms of potential experimental bias, most notably the fact that the system was activated and deactivated in response to a subjective evaluation of meteorological conditions. In particular, the system was not operated under high barometric pressure conditions when there was little cloud moisture. It was also switched off if weather warnings were issued.

One way to limit potential experimental and other sources of bias is through controls, in this case rainfall gauges that would reflect the same influences that might arise from one or more sources of experimental bias. For example, the bias that might result from:

- Locating the Atlant system in an area of above average or more variable rainfall;
- Only operating the system when there is a high level of cloud moisture.

This will tend to require the location of control gauges within a relatively close vicinity of the Atlant site, which in turn requires a reasonable degree of specificity of where the system is expected or postulated to have an effect on rainfall. That is, the more localised is the target area the easier it becomes to locate rainfall gauges that are both highly correlated with gauges in the target area and not likely to be influenced or influenced as strongly by the system under evaluation.

One approach is to partition the overall target area into different zones and compare the predictions generated by the Control and Effects models within these partitions. Importantly, these partitions do not need to be static; they may depend, for example, on wind direction and speed. Given the postulated mechanism through which Atlant operates, the natural partition of the overall target area would be relative to the prevailing wind direction. This approach is suited to exploratory analysis where there is uncertainty regarding the extent of the target area, as for example what might be considered the appropriate downwind arc. This is because the partitioning does not directly affect the predictions of the Control or the Effects models. The limitation of this approach is that the inferences that may be drawn depend on but do not exploit the spatial correlation between the gauges in the different partitions.

An alternative is to designate neighbouring gauges as controls and targets and use the control gauge information to specify the Control Model. Again the definition of targets and controls needn’t be static. However, the definition needs to be fixed before the data is analysed as the choice of targets and controls will alter the difference between the Control and Effects models. The approach is not open to exploratory analysis. However, it has the capacity to fully exploit spatial correlation.

Given there is a reasonably high degree of uncertainty with respect to area of effect, the initial trials will have to be exploratory in nature and the first approach described above will be taken. As a consequence, findings may be less robust and will need to be more qualified.

Ultimately, the objective of this statistical analysis is twofold. First is to show, subject to some important caveats, that a more powerful means of validating the efficacy of the Atlant system can be constructed from field trial data. Second is to design, evaluate and document improved experimental methodologies that will allow real-time evaluation of rainfall enhancement technologies more generally.
2. THE ANALYTICAL FRAMEWORK

The statistical analysis presented here differs from the original University of Queensland study in three important ways:

- It separates the potential rainfall enhancement effect into two components:
  - An increase in the probability that within a period, in this case 24 hours, a rainfall event will occur;
  - Given that a rainfall event does occur, the change in the expected level of rainfall over that period;
- It takes into account the spatial correlation between rainfall observations at different locations, again at a 24-hour frequency;
- It takes into account that an enhancement effect would be directed by meteorological conditions within the target area, specifically in line with the direction of prevailing winds.

Separating the enhancement effect into a probability that a rainfall event will occur and a change in the expected level of rainfall serves an experimental design as well as a statistical modelling purpose. In terms of statistical modelling, a two-stage approach can be used to take into account the fact that the distribution of rainfall has a probability mass at zero and (theoretically) a continuous distribution above zero; a potential source of estimation bias. From a design perspective it is more difficult to determine if the probability of a rainfall event has been enhanced:

- The definition of a rainfall event is somewhat subjective, for example, is rain falling in two successive 24-hour periods considered one or two rainfall events?
- There is greater potential for experimental design bias if the enhancement device is switched on under a subjective assessment of meteorological conditions, as was the case in the Paradise Dam trial. The system was switched off during severe weather warnings and in a period dominated by high pressure systems and little or no cloud moisture. A more costly and time consuming double-blind design would be an approach to eliminating this source of experimental bias in future trials.

For the purpose of this analysis it was assumed that Atlant had no effect on the probability that a rainfall event would be recorded within a 24-hour period in the target area. That is, the analysis is restricted to considering the conditional effect of the Atlant system on expected rainfall given a rainfall event has occurred. This is, in part, because of the potential experimental bias due to when the system was switched on and off. If the device was switched on when it was more likely to rain in the target area, this could result in a spurious correlation between the operation of Atlant and the probability of observing a rainfall event in the target area.

An empirical analysis of the probability of Atlant being associated with the occurrence of rainfall events in the trial area was done (see Appendix A). A relationship was found between the operation of the Atlant system and the probability of a rainfall event occurring in the general target area. There was no relationship found between the operation of the Atlant system and the probability of a rainfall event occurring in the downwind region of target area.
This is consistent with the assertion that the experimental bias associated with the switching on and off of the Atlant system affected the target area as a whole.

**Changes in the Expected Level of Rainfall**

It seems clear that the key to isolating any potential Atlant signal - in terms of increasing the expected level of rain as opposed to the probability of rainfall event - from naturally occurring noise and any experimental bias will depend on:

- Establishing the region affected or footprint of the Atlant system within the overall target area; and
- Having a reasonably high level of correlation between rainfall gauges within and outside the footprint.

To investigate the later of these dependencies, pairwise correlations of gauge specific values of LogRain (natural logarithm of daily rainfall) were calculated and plotted against the between gauge distances in km. Charts 1 and 2 below show this for all gauges (target and control areas), as well as for the target area only, respectively. LogRain was used instead of actual rainfall as this is the analysis variable in the study. The 4th order quadratic fit is also shown in each case. As would be expected, the correlation between the gauges is higher, the closer the gauges are. The target area gauges have a high correlation in rainfall at daily time scales that lend themselves to this type of study.

**Chart 1 Correlations between individual gauges in target and control areas**

![Chart 1](image1.png)

**Chart 2 Correlations between individual gauges in target area**

![Chart 2](image2.png)
Exploiting the directional aspect of the Atlant enhancement system within the target area is important. It can allow the target area to be portioned into segments or sub-areas where the enhancement effect should be greater or less, within relatively close proximity of the Atlant site.

Given that ions, in either a free state or bonded with aerosols, are carried upwards to mix with moisture in clouds, the direction they are carried and hence the enhancement effect will depend on:

- Meteorological conditions, such as wind direction and speed;
- Fixed effects of topography and land cover giving rise to orographic lifting and affecting turbulence.

Given that we can determine the principal direction in the simplest case, the enhancement effect would be expected to be found downwind of the Atlant site. As we move radially away from the principal direction in which the ions are moving, the enhancement effect should decline.

It should be noted that the term downwind is somewhat imprecise as the direction and speed of surface and upper level winds can vary. As a consequence the determination of the principal direction in which the ions are expected to be moving may introduce a degree of subjectivity. However, so long as there is no systematic bias, this should only serve to reduce the power of a statistical test. That is, to reduce the likelihood that the null hypothesis of no effect will be rejected when it is in fact false. A bias could arise, however, if information on rainfall was used to determine the direction downwind of the site.

For the Bundaberg trial data, the downwind direction was obtained by estimating a steering wind direction. Steering winds are those winds that are present in the cloud layer. The direction that appears to direct the general movement of airmass in the cloud layer is referred to as a steering wind direction. The method used to determine the steering wind direction is outlined in Boxes 1 and 2.
Box 1 Determining the Steering Wind Direction – Part I

The determination of primary steering wind flows were obtained from using radiosonde data, produced by the BoM in combination with local surface weather observations.

For each day of the trial, average steering wind speed and direction were estimated. When there was a large change in wind flow or airmass, that portion of the day that saw the most widespread cloud and/or precipitation throughout the trial area was used to average the primary wind flow.

The first step was to obtain daily weather observations at the Paradise Dam site, including wind speed and direction, temperature, pressure and humidity. The wind data were then compared to other trial area wind observations to determine a spatially and temporally averaged surface wind speed and direction for the day.

The next step was to determine a representative wind profile using the closest BoM radiosonde balloon weather observations. Two regional offices of the BoM, at Brisbane and at Rockhampton, Queensland, release radiosonde weather balloons twice daily and are 250kms south and 270kms north respectively away from the Paradise Dam site (see figure below). The sounding that most reflected the air mass characteristics as measured at the Paradise Dam site was chosen as the representative wind profile. Normally both locations were under tropical airmass types, but generally when weak frontal boundaries moved through the area, Brisbane was normally more representative of the Paradise Dam location, while when occasional tropical lows would move from north to south Rockhampton was considered more representative.

Paradise Dam site and Trial Area location with upper air sounding locations
Box 2  Determining the Steering Wind Direction – Part II

Aerological diagram from with Brisbane Airport showing the vertical temperature circled in yellow and wind profile circled in red.

Once the representative sounding was chosen, average wind speed and direction describing the steering flow were calculated by conducting a speed-weighted average of the wind data in the lower 6kms (20,000 feet) of the atmosphere or below the 500HPa level. The figure above is an example of a sounding with wind barbs (indicated direction and speed) located to the right and the temperature dewpoint trace in the centre. Time is given in Z (Zulu) or Greenwich Mean Time. On this day the average surface winds were from 250 degrees with the above ground average direction from 310 degrees and the overall primary steering flow given was from 290 degrees for this day.

2.1. PARTITIONING THE TARGET AREA

Partitioning the target area in relation to steering wind direction, or the principal direction of effect, serves to modulate the signal generated by the Atlant system. The fact that this is a dynamic process is important because the absolute direction of wind will vary. As a consequence, the fact that the Atlant system may be located near land features that enhance or limit rainfall becomes a small source of potential bias.

Locations within the target area were classified into wind flow sectors or segments. Segments were defined by a set of angles, increasing symmetrically in absolute terms about the principal direction of effect. Rainfall gauge locations were then associated with the minimum angle that contained that location. Note that day-to-day variation in wind direction meant that the same gauge was usually classified to different wind flow sectors on different days.
The wind flow sectors used in the analysis were:

- **Sector 1**: 0° ≤ ± angle < 5° (black ●)
- **Sector 2**: 5° ≤ ± angle < 10° (red ●)
- **Sector 3**: 10° ≤ ± angle < 20° (orange ●)
- **Sector 4**: 20° ≤ ± angle < 30° (lime ○)
- **Sector 5**: 30° ≤ ± angle < 45° (green ○)
- **Sector 6**: 45° ≤ ± angle < 90° (aqua □)
- **Sector 7**: 90° ≤ ± angle < 135° (blue ×)
- **Sector 8**: 135° ≤ ± angle ≤ 180° (pink +)

Two illustrations of how the wind flow sectors partitioned the target area are presented in charts 3 and 4.

Having winds from different directions over time reduces the likelihood that a given wind flow sector will be associated with any fixed orographic effects. For example, if rainfall was consistently associated with on-shore winds from the east then, for example, mountain ranges or hills could produce areas of elevated or reduced rainfall that were linked with a particular wind flow sector.

The distribution of steering wind directions over the trial period is shown in Chart 5. With the exception of north-northeast there is a wide distribution of steering wind directions. However, it is clear that the most prevalent direction is from the west.

The distribution of steering wind directions given that rain was recorded in the target area is shown in Chart 6. The most prevalent directions are from the east-southeast and west. There were very few days with winds from either the north or the south. While the distribution is clearly bi-modal, the fact that rainfall in the target area was most commonly associated with winds coming from opposite directions suggests there is little scope for orographic bias.
Chart 3  Target area 13 February 2008 – Steering wind direction is 20 degrees

Chart 4  Target area 15 March 2008 – Steering wind direction is 100 degrees
Chart 5  The distribution of steering wind directions from January through May 2008

Chart 6  The distribution of steering wind directions on days with rainfall in the target area from January through May 2008.
3. STATISTICAL ANALYSIS

As noted, the purpose of the statistical analysis was to determine whether the operation of the Atlant system was associated with increased rainfall in the target area, conditional on the fact that a rainfall event has occurred. A rainfall event was defined as having at least one recorded measurement of rainfall within the region of interest.

The distribution of daily rainfall observations in the study area is strongly right skewed. Raw observations were therefore transformed using the natural logarithm. Since the logarithm of zero is not defined, this automatically resulted in the analysis being confined, for each gauge, to days when rainfall was recorded. In what follows, this transformed value is referred to as LogRain when appropriate. However, the term rainfall itself is used generically - it may refer to rainfall levels or its natural logarithm.

The analysis was done in six stages:

1. Estimate and remove the trend in observed daily rainfall in the north and south control areas given that rainfall has occurred in at least one of the control areas;

2. Regress the daily deviation from trend at each gauge site in the north and south control areas as a function of latitude and longitude, as well as a fixed daily effect;

3. Calculate expected rainfall at the control and target gauges as the sum of the daily rainfall trend in the control areas and the regression estimate of the individual rainfall gauge deviations. The differences between the control and target area gauge deviations being attributable to location.

4. Estimate the Control Model for expected rainfall at the individual gauge sites in the control and target areas in the absence of variables related to the operation of the Atlant system, including the instrumental estimate of expected rainfall from stage 3 and historical differences in average rainfall in the control and target regions;

5. Estimate the Effects Model for expected rainfall at the individual gauge sites in the control and target areas including variables related to the operation of the Atlant system; and

6. Predict expected rainfall at the gauge sites in the target area using both the Control Model and the Effects Model, and then use the difference to calculate the estimated enhancement effect in each wind sector.

One of the limitations of the Bundaberg trial is the lack of gauge specific measurements on meteorological conditions such as barometric pressure and vertical wind profiles. As a consequence, the control modelling is based largely on fixed gauge effects.

3.1. ANALYSIS OF THE CONTROL AREAS (STAGES 1 - 3)

Rainfall data form the north and south control areas were used to construct an instrumental control variable for predicting rainfall in these areas as well as in the target area. The objective is to make the best possible prediction of rainfall in the target areas from concurrent observations of rainfall in the control areas. The underlying conceptual model is that similar weather patterns or trends would affect the control and target areas and those deviations can best be explained by locations of individual gauges. The instrumental approach used in this
analysis is a reflection of the need to identify weather patterns of trends using only data from the control areas.

The trend in expected daily rainfall in the north and south control areas was estimated using a penalized spline smooth, for days in which some rainfall was recorded in the control areas. The spline smooth was fitted to the value of LogRain at each site, and the penalty factor was subjectively chosen in order to provide a compromise between a straight-line trend (high penalty factor) that ignored day-to-day variation in rainfall at different sites and an excessively variable trend (low penalty factor) that was completely driven by daily average rainfall at sites in the control areas. It also removed some of the variability induced by days where only a few isolated rainfall measurements were recorded. The results are shown in Chart 7. The points in red correspond to rainfall at north control area locations and the points in green correspond to rainfall at south control area locations.

Chart 7  The spline estimate of daily expected log rainfall in the control areas

The deviation in rainfall from this trend was calculated and regressed against the locations of the gauges, along with a fixed day or rainfall event effect. The inclusion of the fixed effects was to account for short-term temporal variation about the trend. The estimates of the regression parameters for this model are presented in Table 1.

The explanatory power of the regression is around 28 per cent of the total spatio-temporal variation in observed daily rainfall in the control areas during the trial. The significance of the fixed day effects indicates that there is spatial correlation in daily rainfall within and between
the north and south control areas. The Latitude*Longitude interaction term indicates that there is a general drying trend toward the southwest.

Predicted rainfall was then calculated using the daily fitted trend values generated by the spline model and the predicted deviation at each gauge. These values were then used as an instrumental variable (denoted Expected Rain below) in the Control Model.

**Table 1  Stage 3 regression results**

<table>
<thead>
<tr>
<th>Effect</th>
<th>DF</th>
<th>F-Statistic</th>
<th>Significance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latitude</td>
<td>1</td>
<td>0.3096</td>
<td>0.5780</td>
</tr>
<tr>
<td>Longitude</td>
<td>1</td>
<td>1.0708</td>
<td>0.3007</td>
</tr>
<tr>
<td>Latitude*Longitude</td>
<td>1</td>
<td>6.4177</td>
<td>0.0114</td>
</tr>
<tr>
<td>Day</td>
<td>125</td>
<td>4.2642</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

**3.2. THE CONTROL MODEL (STAGE 4)**

The Control Model was fitted using rainfall observations from the control and target areas. In particular, these observations were regressed against values of Expected Rain for these areas derived from the regression model fitted at stage 3.

The covariates in the Control Model also included average monthly rainfall data for the two control areas as well as the target areas for the previous 30 years. In absolute terms these averages may not be reflective of the seasonal conditions that prevailed in the study regions. However, their relative levels may still explain regional differences in rainfall due to fixed effects such as topography. These values are provided in Table 2. The actual values used in the Control Model were the natural logarithms of the values in Table 2, and are denoted by Log 30 Average in Table 3.

The regression estimates for the Control Model are presented in Table 3.

**Table 2  Monthly average rainfall 1978 to 2007**

<table>
<thead>
<tr>
<th>Month</th>
<th>North Control</th>
<th>Target</th>
<th>South Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>112.22</td>
<td>100.59</td>
<td>100.19</td>
</tr>
<tr>
<td>February</td>
<td>109.66</td>
<td>105.54</td>
<td>110.19</td>
</tr>
<tr>
<td>March</td>
<td>65.56</td>
<td>77.24</td>
<td>77.72</td>
</tr>
<tr>
<td>April</td>
<td>43.38</td>
<td>49.82</td>
<td>69.12</td>
</tr>
<tr>
<td>May</td>
<td>47.27</td>
<td>55.16</td>
<td>61.08</td>
</tr>
</tbody>
</table>

*Source: Bureau of Meteorology*
### Table 3  The Control Model

<table>
<thead>
<tr>
<th>Effect</th>
<th>DF</th>
<th>t-Statistic</th>
<th>Significance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.1064</td>
<td>1.26</td>
<td>2.2069</td>
</tr>
<tr>
<td>Expected Rain</td>
<td>0.9085</td>
<td>48.11</td>
<td>0.0000</td>
</tr>
<tr>
<td>Log 30 Average</td>
<td>0.0971</td>
<td>1.21</td>
<td>0.2069</td>
</tr>
<tr>
<td>Log 30 * Expected Rain</td>
<td>0.3131</td>
<td>4.91</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Note: The Log 30 * Expected Rain term was mean corrected: (Log 30 Average – 1.09)(Expected Rain – 1.19).

The significance of the Expected Rain term is indication that there is strong spatial correlation between rainfall observations within the target area and between the target and control areas, as indicated in Charts 1 and 2 previously.

### 3.3. ANALYSIS OF THE TARGET AREA (STAGE 5)

The Effects Model included all the variables used in the Control Model plus the variables associated with the operation of the Atlant system. These included:

- Variables corresponding to the length of time (minutes) the Atlant system was switched on during the current day as well as the previous two days. These are denoted ATLANT On T, ATLANT On T-1 and ATLANT On T-2 respectively;

- Nine target area dummy variables made up of eight wind flow sector variables that indicated whether a target area gauge was in the relevant sector while the Atlant system was active, and a separate variable indicating when a rainfall reading is from a target area gauge on a day when Atlant was inactive.
  - Atlant was deemed to be active on any particular day if the system had operated that day or either of the previous two days.
  - The reference variable for the nine target area dummy variables is one that indicates whether a gauge is in the north or south control areas. The implied coefficient for this reference variable is the negative of the sum of the coefficients of the nine target area dummy variables.

The results are summarised in Table 4.
Table 4  Effects Model with 8 wind flow sectors

<table>
<thead>
<tr>
<th>Effect</th>
<th>Coefficient</th>
<th>t-Statistic</th>
<th>Significance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.0697</td>
<td>0.71</td>
<td>.4797</td>
</tr>
<tr>
<td>Expected Rain</td>
<td>0.9037</td>
<td>47.54</td>
<td>.0000</td>
</tr>
<tr>
<td>Log 30 Average</td>
<td>0.1760</td>
<td>2.25</td>
<td>.0314</td>
</tr>
<tr>
<td>Log 30 * Expected Rain</td>
<td>0.3576</td>
<td>5.55</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Wind Flow Sector 1</td>
<td>0.1795</td>
<td>1.69</td>
<td>.0914</td>
</tr>
<tr>
<td>Wind Flow Sector 2</td>
<td>0.1011</td>
<td>0.96</td>
<td>.3361</td>
</tr>
<tr>
<td>Wind Flow Sector 3</td>
<td>0.0399</td>
<td>0.52</td>
<td>.6064</td>
</tr>
<tr>
<td>Wind Flow Sector 4</td>
<td>-0.0745</td>
<td>-0.94</td>
<td>.3453</td>
</tr>
<tr>
<td>Wind Flow Sector 5</td>
<td>-0.0544</td>
<td>-0.82</td>
<td>.4127</td>
</tr>
<tr>
<td>Wind Flow Sector 6</td>
<td>-0.0147</td>
<td>-0.32</td>
<td>.7478</td>
</tr>
<tr>
<td>Wind Flow Sector 7</td>
<td>0.0156</td>
<td>0.35</td>
<td>.7288</td>
</tr>
<tr>
<td>Wind Flow Sector 8</td>
<td>0.0480</td>
<td>1.09</td>
<td>.2741</td>
</tr>
<tr>
<td>Atlant Inactive</td>
<td>-0.0606</td>
<td>-0.93</td>
<td>.3545</td>
</tr>
<tr>
<td>Atlant On T</td>
<td>0.0003</td>
<td>4.40</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Atlant On T-1</td>
<td>0.0002</td>
<td>2.17</td>
<td>.0302</td>
</tr>
<tr>
<td>Atlant On T-2</td>
<td>-0.0005</td>
<td>-6.29</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Note: The Log 30 * Expected Rain term was mean corrected: (Log 30 Average – 1.09)(Expected Rain – 1.19).

Again the most significant variable is Expected Rain as predicted from the Stage 3 regression. The operating time variables were also quite significant. As Atlant was activated according to different meteorological conditions the significance of the operational variables is not unexpected.

There were two principal reasons why the Atlant system was switched off:

* First, there was little or no cloud moisture in the target area due to high atmospheric pressure; and

* Second, there were severe storms or storm warnings for the Bundaberg region.

Their inclusion in the regression should reduce the likelihood that estimated wind flow sector effects will be biased due to the way in which the system was operated. However, in future trials it is recommended that the system be operated independently of the prevailing meteorological conditions.

The wind flow sector variables were collectively significant at the 2.5 per cent significance level (97.5 per cent confidence level). Sectors 1 through 3 were positive, sector 4 and 6 negative and sector 7 and 8 positive. The significance of these groupings is examined later.
3.4. ESTIMATING ATLANT EFFECT (STAGE 6)

To calculate the estimated Atlant effect, rainfall was predicted from the gauges in the target area using the Effects Model and the Control Model. Since these models are for LogRain, any difference in their predictions is on a logarithmic scale. On back-transformation, and after a nonparametric correction for the resulting transformation bias, such a log scale difference then corresponds to a ratio that estimates the relative change in expected rainfall with and without the operation of Atlant. In percentage terms, these relative changes are referred to as estimated rainfall enhancement effects below. The results are summarised in Chart 8 in which these estimated rainfall enhancement effects, averaged over the entire trial and by month, are presented. In interpreting this chart, it should be noted that the majority of the rainfall fell in the months of January and February.

Chart 8  The estimated rainfall enhancement effect by wind flow area

There is a clear pattern with the largest enhancement effect directly downwind, declining as we move at right angles to the steering wind and then increasing again. The pattern is evident in each month of the trial. For the entire trial the enhancement effect ranges from over 12 per cent in the third wind flow sector to over 27 per cent in the first wind flow sector.

While the pattern in the wind flow results is striking, the only individual effect that is associated with an estimated coefficient that is significant at or above the 10 per cent significance level is in the sector that is directly downwind of the Atlant site. This may be due to the limited number of gauges in each sector.

The pattern in the wind flow sector effects suggests that the wind flow areas could be aggregated into three sectors:

• Down-Wind – a 30° arc at 0° ≤ ±15°
• Cross-Wind – two 150° arcs at 90° ≤ ±75° and at 270° ≤ ±75°
• Up-Wind – a 30° arc at 180° ≤ ±15°
The upwind and downwind sectors were chosen to be a symmetric reflection along the direction of the prevailing wind. The results for this alternative specification are shown in Table 5.

Table 5  Effects Model with 3 wind flow sectors

<table>
<thead>
<tr>
<th>Effect</th>
<th>Coefficient</th>
<th>t-Statistic</th>
<th>Significance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.0437</td>
<td>0.45</td>
<td>0.6518</td>
</tr>
<tr>
<td>Expected Rain</td>
<td>0.9041</td>
<td>47.57</td>
<td>0</td>
</tr>
<tr>
<td>Log 30 Average</td>
<td>0.1782</td>
<td>2.18</td>
<td>0.0293</td>
</tr>
<tr>
<td>Log 30 * Expected Rain</td>
<td>0.3571</td>
<td>5.54</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Up-Wind</td>
<td>0.0212</td>
<td>0.34</td>
<td>0.7349</td>
</tr>
<tr>
<td>Cross-Wind</td>
<td>0.0200</td>
<td>0.59</td>
<td>0.5529</td>
</tr>
<tr>
<td>Down-Wind</td>
<td>0.1504</td>
<td>2.49</td>
<td>0.0126</td>
</tr>
<tr>
<td>Atlant Inactive</td>
<td>-0.0377</td>
<td>-0.67</td>
<td>0.5059</td>
</tr>
<tr>
<td>Atlant On T</td>
<td>0.0003</td>
<td>4.44</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Atlant On T-1</td>
<td>0.0002</td>
<td>2.25</td>
<td>0.0246</td>
</tr>
<tr>
<td>Atlant On T-2</td>
<td>-0.0005</td>
<td>-6.17</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Note: The Log 30 * Expected Rain term was mean corrected: (Log 30 Average – 1.09)(Expected Rain – 1.19).

Overall, the aggregation of the wind flow sectors did not substantially alter the regression estimates for the other variables in the model. The downwind sector estimate is significant at nearly the 1 per cent level (99 per cent confidence level). The crosswind and upwind sectors estimated are highly insignificant.

The enhancement effects were again calculated with reference to the Control Model as discussed previously. The results are summarised in Chart 9. It should be noted again that the majority of the rainfall fell in the months of January and February.
The average enhancement effects over the entire trial period were relative to the level of rainfall predicted by the Control Model:

- Downwind – +24.8 per cent;
- Crosswind – +7.2 per cent;
- Upwind – +7.8 per cent.

Given the wind flow sector effects in the model are purely additive a simple reliability analysis was conducted on the basis of the sampling distribution of the estimated coefficient relating to downwind rainfall. While this ignores the covariance between the coefficient estimates it does provide a general guide as to the reliability of the predicted estimates. The 10th and 25th percentiles of the sampling distribution, assuming normality, were used to predict the change in rainfall over the trial period.

The results indicated that there was roughly a 90 per cent probability that rainfall was at least 15.5 per cent higher in the downwind arc described above and a 75 percent probability that rainfall was at least 19 per cent higher, than would be expected given the controls specified in the statistical model.

Correcting the Estimates for Experimental Bias

Clearly it is not possible to correct for any potential experimental bias in a formal sense. However, there is an elevation in rainfall across the overall target area, excluding the downwind sector of around 7.2 per cent. Attributing this to the fact that the Atlant system was switched on under conditions which appear to be conducive for the operation the system to enhance rainfall may be both appropriate and conservative.

The corrected enhancement effect for the downwind sector would be 17.6 per cent with a 75th percentile lower bound of 11.8 per cent and a 90th percentile bound of 8.3 per cent.
**Dynamics**

The presence of significant lagged operating times for the Atlant system in the Effects Model is of potential interest. At face value it suggests that efficacy of the system may decline with its continuing operation. However, these lags may be a proxy for lagged meteorological effects that have not been included in the model, particularly since the system was switched on and off in response to changing meteorological conditions.

One and two-day lags were included in the Effects Model. As rainfall is expressed in log terms, observations for gauges that did not record rainfall were given a value of zero. The lags were significant but did not yield any substantial changes to the model in terms of results by wind flow sector, but this still leaves open the question as to whether these lagged Atlant effects are due to the operation of the system itself or whether they are due to the mode of operation of the system in the Bundaberg trial. This is an issue that will need to be revisited in the design of further trials of the system.
4. CONCLUSIONS

The original experimental design of the Atlant Bundaberg trial did not contemplate more than a simple statistical analysis of possible rainfall enhancement effects. However, a subsequent spatial-temporal analysis of the trial data showed that the operation of the Atlant system during the trial was indeed associated with a significant elevation in rainfall. A number of caveats have been made, largely with respect to the experimental design and the limited availability of local meteorological data, and will need to addressed in subsequent trials. These findings on their own are not sufficiently robust to establish the efficacy of the Atlant system but appear strong enough to warrant further investigation.

Another important point is the difference between statistical significance and substantive effect - that is, is any attributed effect sufficiently large to generate a return that would justify the costs of ongoing evaluations and ultimately implementing the technology on a larger scale.

The operation of the Atlant system was not associated with an increase in the probability that a rainfall event, defined as rainfall being recorded on a given day, in the target area under the general meteorological conditions prevailing during the trial (see Appendix 1). In this analysis the directional effects of Atlant were ignored. This is an area that should be addressed in future studies where issues of experimental bias are better addressed and a more extensive data set is available.

However, where a rainfall event occurred within the target area, the operation of the Atlant system was associated with a significant increase in rainfall downwind of the Atlant site. Within a 30-degree arc extending 70 km of the site, rainfall was 24.8 per cent higher relative to what would be expected given the control variable included in the statistical model. No significant differences in rainfall were observed in the balance of the target area. Allowing for the potential experimental bias associated with the trial, as discussed previously, and deducting the observed elevation in rainfall in the balance of the target area, the predicted change would be 17.8 per cent.

If, as the statistical results suggest, the elevated rainfall downwind of the Atlant site was attributable to its operation, then the facility contributed to increased rainfall in the Bundaberg region. The region downwind of the Atlant device covers an area of some 133,500 hectares. With an average rainfall of over 388mm from January to May, a 24.8 per cent increase in rainfall due to Atlant would equate to 99 mm. On this basis, the volumetric contribution of the Atlant operation over the five month trial was 135GL. Again allowing for experimental bias, the adjusted volume of additional attributed to the operation of the Atlant system is 95.8GL.

It is difficult to estimate an accurate value of say 100GL of additional rainfall in the target area. However it is quite clear that if such rainfall were directed into an area with a significant rainfall deficit, the field component of the trial would have been judged to have generated a substantial benefit. To cover the physical operating - as opposed to research and corporate - costs of the trial the break-even value of the additional rainfall would have to be in the order of $5 a megalitre or less than one cent a kilolitre. Taking the full costs of the trial into account would still leave the cost of estimated additional water at well under $100 a megalitre or 10
cents a kilolitre. These figures are only intended to highlight the fact that the cost of the trials themselves may be largely if not fully offset by even a modestly good result. A more accurate assessment of these benefits would have to take into account run-off and water transport calculations as well as the benefits to local landholders.

Should the technology reach a commercial stage, ongoing contracted pricing would also take commercial considerations of IP and its development and return on risk into account. Multiple-system operations would further reduce the cost of water generated. These are low figures against alternative means of incrementing water supplies and they will be tempered by the potential for other entrants who will benefit from the open exposure to the technology and the trials. Further, as highlighted by the trial, a successful application will generate benefits - beyond storage inflow for latter use - to landowners and the environment over a relatively large area.

Improvements in trial design, indicated by this present work, include:

• Operating the system independently of or at least in a fixed and objective way relative to prevailing meteorological conditions. This could utilise, for example, a double-blind design to avoid potential experimental bias, or a pre-determined sequence of days when the device is turned on, suitably balanced by days when it is turned off.

• Monitoring wind direction and wind velocity on a more frequent basis, such as hourly or 12-hourly.

• Locating subsequent trials where there are nearby meteorological measurements available on a concurrent and historical basis.

• A more detailed examination of the potential impact of the Atlant system on the probability that a rainfall event will occur.
APPENDIX A

A statistical analysis of the impact of operation of the Atlant system on the probability of a rainfall event is presented in this Appendix. In the main body of this report it was noted that the times of operation of the Atlant system during the Bundaberg trial were not independent of prevailing meteorological conditions. In particular, the system was activated when meteorological conditions were suitable for enhancing rainfall. As a consequence, if it was observed that there was a greater probability of a rainfall event in the target area during the trial then this may be simply due to an experimental bias. Conversely, if there was significant experimental bias then it would be reasonable to expect that the operation of the system was associated with a greater probability of observing a rainfall event in the target area.

While this does not preclude the possibility that the Atlant system increases the likelihood of a rainfall event, there are a number of reasons why this may be more difficult to detect than a change in the expected level of rainfall given a rainfall event has occurred. Notably, the meteorological conditions under which the Atlant system would increase the probability of a rainfall event may be reasonably rare. That is, there may be either too little of too much cloud moisture for Atlant to have a detectable effect. As a consequence the number of observations over time needed for such an analysis to yield robust results is likely to be quite substantial.

A secondary purpose of the analysis was to set out a methodology for use in future trials when the Atlant system is operated independently of meteorological conditions.

A.1. STATISTICAL ANALYSIS

The analysis was done in two stages. In the first stage the probability that a rainfall event would or would not occur within a given day at each gauge site in the north and south control areas was modelled via logistic regression. The target area was excluded from this process. The fitted logistic regression model was then evaluated by comparing actual rainfall events with predicted rainfall events for each day and gauge site for both the control and target areas. A rainfall event was predicted to occur at a particular location and on a particular day if the model specified that the probability of a rainfall event for that location and day was greater than 50 per cent.

In the second stage the linear predictor defining the estimated logistic model-based rainfall probabilities from the first stage was included in a more extensive logistic regression model that included data from gauges in the control and the target areas, and took account of whether target area gauges were downwind of the Atlant site on the day. Again the model was used to estimate the probability that a rainfall event would or would not occur. We then look to see if there was an increased likelihood that a rainfall event occurred downwind of the Atlant site as opposed to the balance of the target area.

A.1.1. STAGE 1

A binary variable was defined at each gauge site in the control areas for each day of the trial, taking on a value of 1 if no rain was recorded over a 24-hour period starting at 9:00am, and
zero otherwise. This was the dependent variable. There were a total of 4830 observations recorded over 144 days.

A logistic model was used to estimate the probability that a rainfall event would not occur. The explanatory variables included in the model were latitude and longitude as well as a day effect. That is, the probability of not observing a rainfall event at each gauge was adjusted according to the proportion of gauges that did not record a rainfall event across the control areas. The results are summarised in Table A1.

Table A1  Stage 1 logistic regression results

<table>
<thead>
<tr>
<th>Effect</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Significance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latitude</td>
<td>1</td>
<td>9.44</td>
<td>0.0021</td>
</tr>
<tr>
<td>Longitude</td>
<td>1</td>
<td>18.36</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Latitude*Longitude</td>
<td>1</td>
<td>5.74</td>
<td>0.0166</td>
</tr>
<tr>
<td>Day</td>
<td>143</td>
<td>2629.16</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

The latitude and longitude coefficients are significant at the 5 per cent level or less (confidence level of 95 per cent or more). However, collectively the day effects provide most of the model’s explanatory power.

Not surprisingly, this model predicts the occurrence of rainfall events in the control areas reasonably well. Given that a rainfall event was predicted for a gauge in a control area, this event was observed 75 per cent of the time. However, the model also predicts rainfall events in the target area with a similar level of accuracy. Given that a rainfall event was predicted for a gauge in the target area, this event was observed 70 per cent of the time.

A.1.2. STAGE 2

A binary dependent variable was defined at each gauge site in the control and target areas taking on a value of 1 if no rain was recorded over a 24-hour period starting at 9:00am, and zero otherwise. The linear predictor associated with the model fitted at the first stage, i.e. the logit of the predicted probability of not observing a rainfall event, denoted Stage1 PNR, was included as an explanatory variable. Dummy variables were created for gauges located in the north and south control areas, as well as one for gauges located in a downwind target area defined by the wind flow sector corresponding to an angle of 30 degrees either side of the steering wind vector and downwind of the Atlant site. This is analogous to the wind flow sectors reported in the main part of this paper.
Table A2  Stage 2 logistic regression results

<table>
<thead>
<tr>
<th>Effect</th>
<th>Coefficient</th>
<th>Chi-Square</th>
<th>Significance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.080</td>
<td>8.40</td>
<td>0.0038</td>
</tr>
<tr>
<td>North Control</td>
<td>-0.218</td>
<td>8.83</td>
<td>0.0030</td>
</tr>
<tr>
<td>South Control</td>
<td>-0.599</td>
<td>55.84</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Downwind Target</td>
<td>0.377</td>
<td>38.04</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Stage1 PNR</td>
<td>0.769</td>
<td>1872.80</td>
<td>0.0000</td>
</tr>
<tr>
<td>South * Stage1 PNR</td>
<td>0.126</td>
<td>13.52</td>
<td>0.0002</td>
</tr>
<tr>
<td>North * Stage1 PNR</td>
<td>0.342</td>
<td>82.05</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Downwind * Stage1 PNR</td>
<td>-0.210</td>
<td>56.98</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Note: The interaction terms were mean adjusted.

The explanatory variables are significant at the 1 per cent level (99 per cent confidence level). However, they are difficult to interpret directly due to the mixture of main and interaction effects. We therefore compare the distributions of predicted probabilities of observing a rainfall event in the different ‘drop areas’, i.e. the two control areas, downwind target area and the balance of the target area. Box and whisker plots are shown for each area in Chart A1. It is quite clear visually that the median predicted probabilities of a rainfall event (horizontal line in each box) are higher in the two parts of the target area. This again may be a reflection of the fact that Atlant was turned on and off in response to meteorological conditions as discussed in the main part of this paper.

The distribution of the predicted rainfall event probability in the downwind target area (Target-In) and balance of the target area (Target-Out) are nearly identical, indicating that the Atlant system does not appear to increase the likelihood of the occurrence of a rainfall event.
Chart A1  Box and whisker plots of the predicted probabilities of observing a rainfall event in the control, downwind target and balance of the target area.
APPENDIX B

B.1. UQ PARADISE DAM TRIAL: JANUARY – MAY 2008

B.1.1. TRIAL OVERVIEW

The Paradise Dam trial was evaluated by a team of UQ scientists between the period 01 January to 24 May 2008.

The target area for this trial was defined as a 70km radius circle, centred on the Atlant (25° 21’ 38.72”S, 151° 55’ 15.19”E), roughly 80km southwest of Bundaberg, as shown in Chart B1. This trial area encompassed the Wide Bay and Burnett River region, in South East Queensland which largely comprised the catchment of Paradise Dam. Two control areas were also established for the rainfall distribution analysis, and are shown in Chart B1 below to the north and south of the target area. The yellow dots indicate permanent Bureau of Meteorology rain gauges. Blue dots indicate UQ rain gauges installed for the trial.

Chart B1: Map of the Atlant position in southeast Queensland, along with the north and south control areas (blue circles) and the main Target trial area (red circle) used in the 2008 Paradise Dam Trial.
B.1.2. ASSESSMENT METHODOLOGY

The study built on an earlier trial conducted in May and June 2007 in the South East Queensland Wivenhoe Dam catchment. The study was divided into several specific sub-studies. A summary of the following sub-study is presented here.

- Historical rainfall analysis of the Central Queensland coast and spatial analysis of rainfall data collected during the trial from January – May 2008.

The UQ study components were intended to observe if Atlant produced ions and/or enhanced rainfall in a target area centred on Paradise Dam compared to that recorded in two control areas.

B.1.3. RAINFALL DISTRIBUTION

The rainfall distribution analysis was conducted in order to present spatial interpolations of the rainfall data for the target and control areas to examine the spatial distribution of precipitation from January to May 2008. Comparisons were made with historical rainfall data for the region and with two control sites, located outside the Atlant area of influence.

In order to conduct a spatial analysis of rainfall patterns, a network of automatic weather stations was installed across the target area and two control areas. A UQ network of weather stations augmented the existing BoM network in order to obtain a spatially representative record of rainfall. The target area was determined by creating a 70 km radius around the Atlant system. La Crosse WS2301 professional remote weather stations were distributed throughout the target and control areas with the distance between stations (UQ and BoM) not exceeding 15 km. Northern and southern control areas were designated according to their suitability in terms of geographical and climatologic similarity to the target area. A total of 73 UQ weather stations were positioned in the target area with 11 UQ stations in each of the control areas (see Chart B1 previously).

Weather stations were programmed to record precipitation at 12 hour intervals at 0900 EST and 2100 EST daily. The daily rainfall data for all BoM rain gauges within the target and control areas was also obtained in order to complement the UQ dataset. Quality control of data was performed by comparing UQ rainfall data with nearby UQ stations as well as BoM stations which are subjected to rigorous quality checks. The spatial distribution of precipitation in the target area was examined for the entire period and on a monthly basis and compared to climatology. Specific rainfall events were also examined. Spatial interpolations of the data were presented as rainfall maps, produced using ESRI ArcMap (see Chart B2 below). The linear krigging method of spatial interpolation was used in most cases except where data availability was poor, in which case the natural neighbour tool was used.
Rainfall patterns within the target and control areas were analysed and assessed against the historical rainfall trends in the region both in the last 100 years and during strong La Niña events in order to identify changes in rainfall distribution in the target region during January – May 2008. The strong La Niña years used in this report were determined by the BoM classification whereby years when the Southern Oscillation Index remained strongly positive for many months were considered classic strong events. While not a sufficiently large sample to make firm conclusions, the following points were noted.

- The total rainfall in Jan-May 2008 was generally higher than the 100-year average, but still considerably lower than the 11 strongest La Nina years.

- The rainfall in the Southern control area was only 1% higher than the long-term average, while the Northern control area had 22% more rainfall than the historic average. However, the target area recorded a 55% higher rainfall amount in the Jan-May 2008 period compared to the previous 100 years (same period).

- The rainfall difference between the control and target areas (averages of all available stations in both areas for Jan-May period) in 2008 was +112mm (Target - Control areas) compared to the long-term average difference of +15mm for the previous 100 years. This means that in the target area there was 26% more rainfall recorded than in the control areas in 2008 (see Chart B3), whereas the long-term average rainfall difference only represents 3% of the value recorded in the control areas. The last
times when such similar positive differences were recorded were in 1992 and 1981/82.

- This difference was in the 85th percentile of all values for the last 100 years, which means that only 14 other years had higher positive differences than 2008 (see Chart B3).

Chart B3: Average rainfall difference (mm) between control and target areas 1908-2008
GLOSSARY

Basis of statistical inference: Statistical inference or statistical induction comprises the use of statistics and random sampling to make inferences concerning some unknown aspect of a population. It is distinguished from descriptive statistics. Statistical inference is inference about a population from a random sample drawn from it or, more generally, about a random process from its observed behaviour during a finite period of time.

Bimodal: In statistics, a bimodal distribution is a continuous probability distribution with two modes. These appear as distinct peaks (local maxima) in the probability density function. In this report the distribution of wind directions over the period of the trial was bimodal since two wind directions occurred more frequently than others.

Causal link or Causality: Describes the relationship between causes and effects, is fundamental to all natural science, especially physics, and has a basis in logic. It is also studied from the perspectives of philosophy, computer science, and statistics. The human motivation for classifying some elements of experience (events) as "causes," and others as "effects", rests in the desire to explain why some events are fostered and others prevented.

Control Model: A statistical model for daily rainfall that does not include any factors that relate to the operation of the Atlant system, and predicts rainfall in the target area on the basis of both historical as well as contemporaneous rainfall in the control areas.

DF: Degrees of Freedom (DF). A measure of the number of unconstrained data values that contribute to the calculation of a statistic. Estimates of parameters of statistical models can be based upon different amounts of information or data. The number of independent pieces of information that are used in the estimate of a parameter is its degrees of freedom.

Dummy Variable: In regression analysis, a dummy variable (also known as binary variable) is one that takes the values 0 or 1 to indicate the absence or presence of a level of a categorical explanatory variable.

Effects Model: In addition to the explanatory variables included in the Control Model, this model includes factors that relate to the operation of the Atlant system. These could include the length of time the system is on or off, the location of a gauge relative to the downwind direction from the Atlant site or the distance from the Atlant site to a gauge.

Empirical analysis: The word empirical denotes information gained by means of observation, experience, or experiment. An empirical analysis is evidence-based since it is based on observed data.

Expected Rain: This is a variable whose values are generated from the model fitted in Stages 1 – 3 of the analysis. This fit is restricted to the rainfall for the control areas, but is then used to generate expected values for rainfall (i.e. values of Expected Rain) for both the control and target areas, with these values then used as an explanatory variable in the Control and Effects models. Since it is an independent estimate of the expected rainfall in the target area in the absence of any Atlant effect, Expected Rain corresponds to an instrumental-type explanatory variable in the analysis reported in Stages 4 – 6.
Explanatory variables: Explanatory, or independent, variables are variables that are part of a statistical model for another variable (typically referred to as the response or dependent variable of this model), and so can be used to predict the observed values of this response.

F Statistic: see Test Statistic.

Instrumental control variable: In statistics, the method of instrumental control variables is used to estimate causal relationships when controlled experiments are not feasible. In this report, rainfall data from the north and south control areas was used to develop an instrumental control variable that was used as a predictor of rainfall in the target area.

Logistic model: Binary variables are not suited to standard regression modelling since they can only take two values (0/1, Off/On, event does not occur/occurs). Since the probability of something occurring is just one minus the probability of it not occurring, variables of this type are usually analysed by modelling the probability of the event occurring. A standard model for such a probability is the logistic model. It is a generalized linear model where the logit of this probability is modelled as a linear combination of the values of a set of independent or explanatory variables that may be either numerical or categorical. The fitted values generated by a logistic model are probabilities (i.e. they lie between zero and one) and show how the chance of the event occurring is affected by the values of the explanatory variables. In the Atlant context, this type of model was used to investigate how different factors (including those relating to the operation of Atlant) affected the probability of rainfall occurring.

Logit: Logarithm of the probability that an event occurs divided by the logarithm of the probability that it does not occur.

LogRain: Natural logarithm of daily rainfall. The logarithm of daily rainfall was used as the dependant variable in the analysis described in the report. This was because the distribution of 'raw scale' daily rainfall over the trial period was highly right skewed, and therefore unsuited to regression modelling, while that of the logarithm of rainfall was much less extreme, and consequently much better suited to this type of modelling. A complication associated with regression modelling on the logarithmic scale is that any fitted values of 'raw scale' rainfall generated by applying an exponential transform to the fitted values for LogRain then need to be corrected for the resulting downward transformation bias. In the analysis set out in the report this was achieved by applying a positive bias correction to these values.

Null hypothesis: In statistical hypothesis testing, the null hypothesis formally describes some aspect of the statistical behaviour of a set of data; this description is treated as valid unless the actual behaviour of the data contradicts this assumption. Thus, the null hypothesis is contrasted against another hypothesis. Statistical hypothesis testing is used to make a decision about whether the data contradicts the null hypothesis: this is called significance testing. A null hypothesis is never proven by such methods, as the absence of evidence against the null hypothesis does not establish it. In other words, one may either reject, or not reject the null hypothesis; one cannot accept it. Failing to reject it gives no strong reason to change decisions predicated on its truth, but it also allows for the possibility of obtaining further data and then re-examining the same hypothesis.

Penalty Factors: A correction factor applied to an optimizing criterion in order to ensure that the optimal solution has desirable statistical properties. In the context of the nonparametric spline smoother described in the report, the penalty factor is used to increase the weight
given to that part of the optimization criterion used to calculate the spline fit that favours smoothness, at the expense of the part that favours it tracking the data as closely as possible. By varying this penalty factor the spline fit then becomes an optimal compromise between a highly erratic curve that 'overfits' the data and a very smooth curve that ignores most of its non-linearity and so 'underfits' the data.

**Pseudo R-Squared:** When analysing data using logistic regression, an equivalent statistic to the R-square goodness of fit statistic used in standard regression does not exist. The model estimates from a logistic regression are maximum likelihood estimates arrived at through an iterative process. They are not designed to minimize variance, so the ordinary regression approach to assessing goodness-of-fit via R-square does not apply. In order to evaluate the goodness-of-fit of a logistic model, several pseudo R-square values have been developed. These are "pseudo" R-square values because they are on a similar scale, ranging from 0 to 1 (though some pseudo R-squareds never achieve 0 or 1) with higher values indicating better model fit, but they cannot be interpreted as measures of goodness of fit in the same way that one would interpret an ordinary regression value of R-square.

**PNR:** Probability of no rainfall (PNR) in the control area as calculated in Stage 1.

**Reliability analysis:** In statistics, reliability is the consistency of a set of measurements or measuring instrument, and is often used to describe a test. This can either be whether the measurements of the same instrument give or are likely to give the same measurement (test-retest), or in the case of more subjective instruments, such as personality or trait inventories, whether two independent assessors give similar scores (inter-reliability). Reliability is inversely related to random error.

**Regression Analysis:** A statistical technique for measuring/quantifying the degree to which the values taken by a variable of interest (the dependent or response variable) can be explained or predicted by the values taken by one or more other variables (the independent or explanatory variables). A simple regression analysis has just one independent variable, while a multiple regression analysis can have many independent variables. Regression analysis is a necessary pre-requisite to causal analysis.

**Regression Estimates:** A regression estimate is the result of a regression analysis. A regression analysis refers to modelling and analysis of numerical data for a dependent variable in terms of one or more independent variables.

**R-Square:** A statistical measure of how well the fitted values generated by a regression model approximate the data values for the dependent variable; an R-square value of 1.0 (100%) indicates a perfect fit, while an R-square value of 0.0 indicates that the model has no explanatory/predictive power. It is calculated as the relative decrease in the variance of the regression model residuals (actual values minus fitted values) following the fitting of the regression model.

**Significance Level:** In the model output presented in this report, the Significance Level is stated for each independent variable. This is the 'P value' and indicates whether a variable has statistically significant predictive capability in the presence of the other variables, that is, whether it adds something to the model. In some circumstances, a non-significant P value might be used to determine whether to remove a variable from a model without significantly reducing the model's predictive capability. For example, if an independent variable has a non-
significant P value, we can say that it does not have predictive capability in the presence of the other independent variables, and so remove it and refit the model without it. One has to be cautious in doing this, however, since independent variables are generally not uncorrelated, and an independent variable that does not have predictive capability in the presence of the other independent variables may well have predictive capability when some of those variables are removed from the model.

**Spline smooth:** A nonparametric modelling method that is used to fit a smooth regression curve to data that are intrinsically non-linear. The spline model has no interpretable parameters (since it is nonparametric) and so is mainly used to smooth out noise in non-linear data.

**Spatio-temporal analysis:** A spatio-temporal analysis is a multivariable statistical analysis that allows for the inter-dependence of data across both space and time.

**T Statistic:** see Test Statistic.

**Test Statistic:** A test statistic is a numerical measure that allows one to assess whether a statistical hypothesis is consistent with the observed data. Typically, this is because the distribution of the test statistic is known under the hypothesis and so the actual value of the test statistic can be checked to see whether it is reasonable for this distribution. Two common test statistics used in regression analysis are the t-statistic and the F-statistic. The t-statistic measures the significance of the contribution of an individual independent variable to the fitted value for the dependent variable. It is desirable to have a large (positive or negative) value for a t-statistic for each independent variable. Generally, a t-statistic value that is greater than +2.0 or less than -2.0 is treated as significant, and corresponds to a P value that is less than or equal to .05. The F-statistic is a multivariate version of the t-statistic, in that it allows a group of independent variables to be simultaneously tested for significance. A large value of this statistic (or equivalently a P value less than .05) indicates that at least one variable in the group is significant. The F-statistic is typically used to test whether complex independent variables formed by the interaction of several explanatory variables, or a group of variables associated with different categories of a categorical independent variable, are jointly significant and so should be in the regression model.