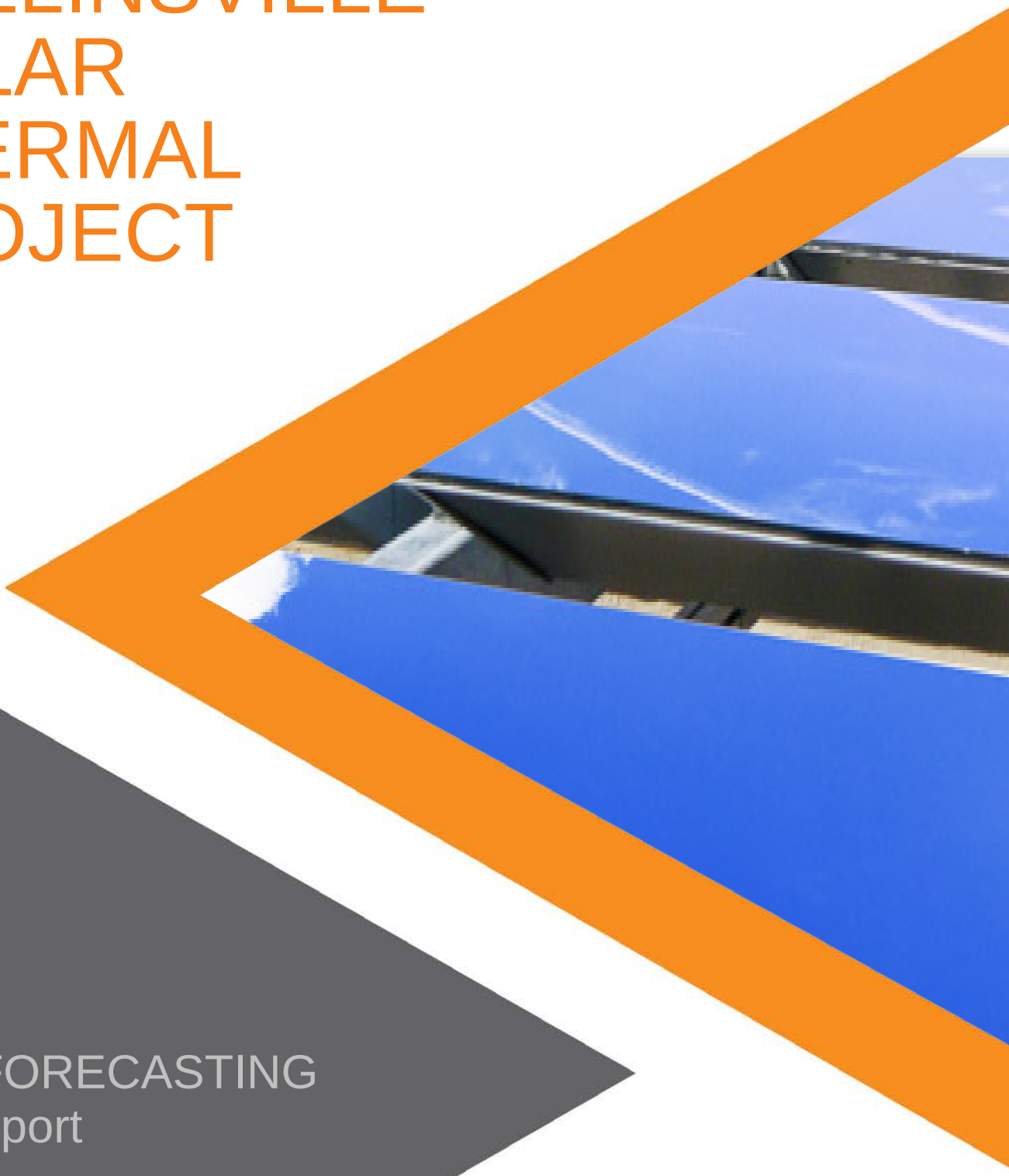
COLLINSVILLE SOLAR THERMAL PROJECT



YIELD FORECASTING
Draft Report

The University Of Queensland
AUSTRALIA

gciQ
Global Change Institute

Prepared for

RATCH-Australia Corporation

RAC
RATCH-Australia Corporation

Chief Investigators

Paul Meredith, Global Change Institute
Craig Froome, Global Change Institute
Hal Gurgenci, Queensland Geothermal Energy Centre of Excellence
John Foster, School of Economics
Tapan Saha, School of Information Technology and Electrical Engineering

Authors

William Paul Bell, p.bell2@uq.edu.au, Energy Economics and Management Group
Phillip Wild, p.wild@uq.edu.au, Energy Economics and Management Group
John Foster, j.foster@uq.edu.au, Energy Economics and Management Group

Energy Economics and Management Group

Postal address:	School of Economics The University of Queensland St Lucia, Brisbane QLD 4072, Australia
Phone:	+61 7 3346 0594 or +61 7 3365 6780
Fax:	+61 7 3365 7299
Website:	http://www.uq.edu.au/eemg/

Please cite this report as:

Bell, WP, Wild, P, Foster, J, 2014, *Collinsville solar thermal project: Yield forecasting - Draft report,* The University of Queensland, Brisbane, Australia.

This is a draft report and the final report is due for publication November 2014.

Draft report – version 7 – 15 July 2014

Copyright

Preface

This is a draft report and the final report is due for publication November 2014. This yield forecasting report is one of seven reports evaluating the feasibility of a hybrid gas-concentrated solar power (CSP) plant using Linear Fresnel Reflector (LFR) technology to replace the coal fired power station at Collinsville, Queensland, Australia. Table 1 shows the seven reports and the affiliation of the lead authors.

Table 1: Collinsville feasibility study reports and their lead researcher groups and authors

Report	Affiliation of the lead author
Yield forecasting (Bell, Wild & Foster 2014b)	EEMG
*Dispatch forecasting (Bell, Wild & Foster 2014a)	EEMG
*Energy economics (Bell, Wild & Foster 2014a)	EEMG
Solar mirror cleaning requirements (Guan, Yu & Gurgenci 2014)	SMME
Optimisation of operational regime (Singh & Gurgenci 2014b)	SMME
Fossil fuel boiler integration (Singh & Gurgenci 2014a)	SMME
Power system assessment (Shah, Yan & Saha 2014a)	PESG
Yield analysis of a LFR based CSP by long-term historical data (Shah, Yan & Saha 2014b)	PESG

*Combined report

These reports are part of a collaborative research agreement between RATCH Australia and the University of Queensland (UQ) funded by the Australian Renewable Energy Agency (ARENA) and administered by the Global Change Institute (GCI) at UQ. Three groups from different schools undertook the research: Energy Economics and Management Group (EEMG) from the School of Economics, a group from the School of Mechanical and Mining Engineering (SMME) and the Power and Energy Systems Group (PESG) from the School of Information Technology and Electrical Engineering (ITEE).

EEMG are the lead authors for three of the reports. Table 2 shows the “Collinsville Solar Thermal - Research Matrix” that was supplied by GCI to the researchers at EEMG for their reports. The suggested content for the three reports in the matrix was restructured to provide a more logical presentation for the reader that required combining the Energy Economics and Dispatch Forecasting reports.

Table 2: Collinsville Solar Thermal - Research Matrix – EEMG's components

Yield Forecasting
Modelling and analysis of the solar output in order that the financial feasibility of the plant may be determined using a long-term yield estimate together with the dispatch model and the modelled long-term spot price.
Dispatch Forecasting
Analysis of the expected dispatch of the plant at various times of day and various months would lead to better prediction of the output of the plant and would improve the ability to negotiate a satisfactory PPA for the electricity produced. Run value dispatch models (using pricing forecast to get $ values out). Output will inform decision about which hours the plant should run.
Energy Economics
Integration of the proposed system into the University of Queensland's Energy Economics Management Group's (EEMG) existing National Electricity Market (NEM) models to look at the interaction of the plant within the NEM to determine its effects on the power system considering the time of day and amount of power produced by the plant. Emphasis to be on future price forecasting.

The results from this yield report are used to inform the *'Energy economics and dispatch forecasting'* report (Bell, Wild & Foster 2014a).

Doctor William Paul Bell
Research Fellow
Energy Economics and Management Group
The School of Economics
The University of Queensland

Executive Summary

1 Introduction

This report's primary aim is to provide yield projections for the proposed Linear Fresnel Reflector (LFR) technology plant at Collinsville, Queensland, Australia. However, the techniques developed in this report to overcome inadequate datasets at Collinsville to produce the yield projections are of interest to a wider audience because inadequate datasets for renewable energy projects are commonplace. The subsequent report called *'Energy economics and dispatch forecasting'* (Bell, Wild & Foster 2014a) uses the yield projections from this report to produce long-term wholesale market price and dispatch forecasts for the plant.

2 Literature review

The literature review discusses the four drivers for yield for LFR technology:

- DNI (Direct Normal Irradiance)
- Temperature
- Humidity
- Pressure

Collinsville lacks complete historical datasets of the four drivers to develop yield projects but its three nearby neighbours do possess complete datasets, so could act as proxies for Collinsville. However, analysing the four drivers for Collinsville and its three nearby sites shows that there is considerable difference in their climates. This difference makes them unsuitable to act as proxies for yield calculations. Therefore, the review investigates modelling the four drivers for Collinsville.

We introduce the term "effective" DNI to help clarify and ameliorate concerns over the dust and dew effects on terrestrial DNI measurement and LFR technology.

We also introduce a modified TMY technique to overcome technology specific Typical Metrological Year (TMY). We discuss the effect of climate change and the El Nino Southern Oscillation (ENSO) on yield and their implications for a TMY.

2.1 Research questions

Research question arising from the literature review include:

The overarching research question:

> *Can modelling the weather with limited datasets produce greater yield predictive power than using the historically more complete datasets from nearby sites?*

This overarching question has a number of smaller supporting research questions:

- Is BoM's DNI satellite dataset adequately adjusted for cloud cover at Collinsville?
- Given the dust and dew effects, is using raw satellite data sufficient to model yield?
- Does elevation between Collinsville and nearby sites affect yield?
- How does the ENSO affect yield?

- Given the 2007-2012 constraint, will the TMY process provide a “Typical” year over the ENSO cycle?
- How does climate change affect yield?

A further research question arises in the methodology but is included here for completeness.

- What is the expected frequency of oversupply from the Linear Fresnel Novatec Solar Boiler?

3 Methodology

In the methodology section, we discuss the data preparation and the model selection process for the four drivers of yield.

4 Results and analysis

In the results section we present the four driver models selected and the process that was undertaken to arrive at the models.

5 Discussion

We analyse the extent to which the research questions are informed by the results.

6 Conclusion

In this report, we have identified the key research questions and established a methodology to address these questions. The models for the four drivers have been established allowing the calculation of the yield projections for Collinsville.

Contents

Figures

Tables

Equations

1 Introduction

The primary aim of this report is to produce hourly yield projections of electricity power for the proposed LFR plant at Collinsville, based on the environmental condition between 2007 and 2013. However, the techniques and methods used to overcome the inadequacies of the environmental, site-specific datasets provide a wider appeal for the report. The dataset inadequacies make accurate projections of future income streams and the subsequent securing of funding difficult (Cebecauer et al. 2011; Lovegrove, Franklin & Elliston 2013; Stoffel et al. 2010).

The hourly power yield projections from this report are used in a subsequent report called 'Energy Economics and Dispatch Forecasting' (Bell, Wild & Foster 2014a), to calculate the lifetime revenue of the proposed plant and perform sensitivity analysis on gas prices.

This report compares the yield from the proposed Collinsville LFR plant using two different calculation methods. One method simply uses complete historical datasets from three nearby sites: MacKay, Rockhampton and Townsville in Queensland. The other method uses datasets derived from a meteorological model developed from three sources:

- BoM's hourly solar satellite data
- BoM's Collinsville Post Office weather station
- Allen's (2013) datasets

The overarching research question for the report:

> *Can modelling the weather with limited datasets produce greater yield predictive power than using the historically more complete datasets from nearby sites?*

The executive summary provides an outline of the report.

2 Literature review

2.1 Introduction

This literature review helped us to develop the research question and informed the methodology to address the research question.

Linear Fresnel Reflector (LFR) technology provides at least three benefits:

- helping address climate change;
- providing a replacement for unsustainable fossil fuel dependency; and
- increasing diversity and resilience within the electricity systems.

Other renewable energy technologies such as solar PV and wind generation have successfully transitioned from the infant industry phase with numerous large-scale commercialisations of the technologies emerging in Australia. In contrast, LFR in Australia is very much in the infant industry stage with a few small booster projects. Furthermore, unlike solar PV and wind generation, LFR lacks the gradually increasing scale pathway from household units to large-scale units because LFR plants involve a minimum economy of scale consideration. This consideration makes the transition from the infant industry phase more problematic. Therefore, there is a requirement for a lager subsidy per venture and, consequently, less scope for experimentation and a risk of failure. The large scale investment requirements make a failure unacceptable, which means that research is essential to better inform investment decisions. This research has a public good aspect with benefits that go beyond those accruing to the individual firm willing to fund such research. The yield projections in this report are the first step in the process to help better inform investment decisions at Collinsville. However, the research is clearly useful to others considering such ventures.

Section 2 presents the four environmental drivers of yield and discusses driver data availability and contrasts the drivers in Collinsville with the three comparison sites. Section 3 introduces the concept of “effective” direct normal irradiance to address the dew effect and dust effect. Sections 4 and 5 discuss the effect of El Niño–Southern Oscillation and climate change on the four drivers and yield to scope sensitivity analysis. Section 6 discusses the format and technique “Typical Meteorological Year” and its implications for sensitivity and inter-year variation analysis and introduces modifications to the technique to overcome shortfalls. Section 7 concludes the literature review and presents the research questions that arise.

2.2 Four main drivers of yield

The US National Renewable Energy Laboratory’s (NREL) Systems Advisors Model (SAM) provides standard yield models for a range of renewable energy technologies, including a model specifically for the proposed LFR technology at Collinsville (Wagner 2012; Wagner & Zhu 2012). SAM calculates the kilowatts (kW) generated each hour using four environment variables.

- Direct normal irradiance (DNI)
- Temperature (Dry bulb)

- Humidity
- Pressure

Section 3.2 discusses the SAM methodology in more detail. Other environment variables also affect the amount of electricity produced but these four variables are considered the main drivers. Thus, they form the nucleus of the "complete meteorological dataset" in the following discussion where there is a choice between using complete historical meteorological datasets from nearby sites and using incomplete data from Collinsville to model the four environmental variables.

Shah, Yan and Saha (2014b) provide a detailed account of the yield calculated using the SAM model and the complete historical meteorological data from three nearby sites at MacKay, Rockhampton and Townsville. Table 3 aids in the inter-site comparison of the four drivers by grouping the annual daily-metrological means for the period 1981-2010. These means are by the four metrological drivers for yield for the four sites.

Table 3: Meteorological daily annual means 1981-2010 for Collinsville and neighbours

	Collinsville PO	MacKay Aero	Rockhampton Aero	Townsville Aero
DNI proxy				
daily sunshine (hours)	-	-	-	8.6
daily exposure (MJ/m2)	20.4	20.8	20.2	21.1
number clear days	121.3	-	120.6	116.3
number of cloudy days	78.2	-	93.0	100.9
9am cloud cover	2.9	-	3.7	4.2
3pm cloud cover	4.0	-	3.8	3.7
Temperature (Dry Bulb) (°C)				
max	30.4	27.4	28.6	29.2
min	16.8	17.9	17.2	20.2
9am	23.3	24.0	22.7	25.3
3pm	29.3	25.9	27.4	27.7
Relative Humidity (%)				
9am	66	72	67	65
3pm	43	64	46	57
Pressure proxy				
elevation (m)	196	5	10	4
Wind speed (km/h)				
9am	3.1	17.9	12.8	13.1
3pm	5.2	25.1	15.7	22.4
Dew point				
9am	16.3	18.2	15.8	17.8
3pm	14.2	18.3	13.6	17.9

(Source: BoM 2014a)

The interrelationship amongst the four drivers and other weather variables provides context to the following discussion and informs the methodology chosen. Radiant energy causes temperature changes, temperature changes cause pressure changes and pressure gradients cause winds. These direct relationships are interwoven and moderated within the hydrologic cycle whose indicators available at the Collinsville BoM weather station include relative humidity, cloud cover, evaporation, dew point and wet bulb temperatures. So, Table 3 also includes, wind speed, for discussion and the dew point. Table 3 presents annual

means, so masks any seasonal variation in the annual cycle that is present in each of the four drivers. Additionally, Table 3 only hints at the variation in the daily cycle.

This section discusses the four metrological drivers for SAM while considering two aspects for each driver. Firstly, is the use of alternative nearby sites suitable to provide weather proxies for Collinsville? Secondly, can the limited weather datasets available at Collinsville to model Collinsville's four drivers to inform the methodology in section 3?

2.2.1 Direct normal irradiance

DNI is the first of the four drivers for yield in the SAM model and is the primary component in driving yield from CSP (Stoffel et al. 2010, p. 101), such as LFR technology. *DNI is the instantaneous intensity of solar direct beam energy falling on a surface normal to the beam (BoM 2013).* BoM estimates DNI from hourly geostationary satellite images, starting in 1990. This contrasts with DNI data from Allen (2013) who produces minute ground based observations starting in Dec 2012.

2.2.1.1 Inter-site comparisons of DNI

The proxies for discussion of DNI in Table 3 include annual mean daily exposure, number of clear days and cloud cover. The daily exposure is derived from satellite data (BoM 2007). Allen (2013) sums BOM's hourly satellite data for the Collinsville Power Station site and finds the sum closely follows the BOM daily exposure at the Collinsville Post Office weather station, so a comparison using the daily exposure as a proxy for DNI is warranted.

In Table 3, the annual mean daily exposure for the four sites is similar, which implies that the yield at MacKay, Rockhampton and Townsville can provide a good approximation to the yield at Collinsville. However, there are two reservations. Firstly, the number of cloudy days at Rockhampton and Townsville are about 20% higher than at Collinsville, which calls into question the validity of the annual mean daily exposure derived from satellite data. Secondly, in Table 3, the 9 am and 3 pm cloud cover indicates a differing daily cycle of cloud cover between the inland high altitude Collinsville and the three coastal low altitude sites, which implies the profile of the daily yield cycle would differ.

MacKay, Rockhampton and Townsville are less than ideal sites for LFR because their low altitude and close proximity to the coast present higher concentrations of aerosols than would be found otherwise. Aerosols reduce DNI, which is a primary component in driving yield from CSP (Stoffel et al. 2010, p. 101). The higher aerosol concentration in the three coastal towns cause a larger yield deviation between satellite and ground station determined DNI than would be found at more ideal CSP sites. However, the BoM (2013) has adjusted the satellite data for atmospheric transmittance aka clearness index, which should ameliorate this concern.

In a further twist to the aerosol effect, sites destined for CSP could be subject to preliminary earthworks or demolishing of exiting power plant, such as in Collinsville. These activities increase the aerosol levels above those expected when the CSP plant is completed, so yield projections based on site based solar measurement underreport yield. This situation is discussed further in the dust effect section below.

2.2.1.2 Collinsville DNI data

Table 3 shows the satellite derived solar daily exposure but SAM requires hourly DNI. BoM (2013) provides hourly satellite DNI data but the previous section questioned the accuracy of the satellite data for terrestrial use when considering the cloud coverage. A solution to this issue is to adjust the satellite data for cloud coverage and other environment variables by calibrating against Allen's (2013) terrestrial DNI dataset for Collinsville. Section 3 discusses the methodology in more detail.

The solar altitude angle provides a way to approximate DNI without cloud cover. The solar altitude angle is the angle subtended between the sun and horizontal plane of the observer. The altitude can be calculated from the zenith or, more fully, the solar zenith angle, that is, the angle subtended between the sun and the normal to the horizontal plane of the observer. Reda and Andreas (2008) provide an algorithm to calculate the zenith angle and Roy (2004) implements the algorithm in computer code to calculate the zenith angle from the time and position by longitude, latitude and altitude.

As discussed, DNI is the primary driver for CSP. However two other measures of irradiance in common use are Global Horizontal Irradiance (GHI) and Diffused Horizontal Irradiance (DHI). *GHI is the instantaneous intensity of solar energy falling on a horizontal surface (BoM 2013).* BoM (2013) provides gridded satellite solar intensity dataset in W/m2 for both DNI and GHI but not DHI. Equation 1 shows how to calculate DHI from the GHI, DNI and zenith angle.

Equation 1: Three irradiances and zenith angle

$$DHI = GHI - DNI \cos(\text{zenith})$$

The BoM (2013) grids are produced for each hour starting in 1990; the grids consists of 839 columns by 679 rows where the grids' x and y corner corresponds to the longitude and latitude 112.025 and -43.975, respectively, and each cell size is 0.05 degrees or approximately 5km.

For the period of interest in this report, 2007 to 2013, two satellites sources were used: the Japanese Advanced Meteorological Imager (JAMI) and the Multi-Functional Transport Satellite (MTSAT) series operated by the Japan Meteorological Agency. Table 3 shows the coverage dates of the two satellites. A grid is produced for each hour but the satellite takes time to traverse Australia hence the minutes past the hour the satellite image was produced is related to the latitude. The latitudes for the proposed Collinsville LFR plant and Collinsville Post Office and Allen's (2013) weather stations are -20.5344, -20.5533 and -20.5418, respectively. These latitudes are between 48 to 49 minutes past the hour for satellite MTSAT-1R and between 46.8 and 47.7 minutes past the hour for the satellite MTSAT-2.

Table 4: Satellite's minutes past the hour by latitude

Start date	2005-11-01	2010-07-01
End date	2010-06-30	Ongoing
Latitude	MTSAT-1R	MTSAT-2
-10.0	46.2	44.7
-15.0	47.2	45.7
-20.0	48.3	46.8
-25.0	49.2	47.7
-30.0	50.1	48.6
-35.0	51.0	49.5
-40.0	51.7	50.2
-44.0	52.3	50.8

(Source: BoM 2013)

Section 3 discusses further calculating the altitude and zenith and adjusting the satellite derived DNI, GHI and DHI for the minutes past the hour. Table 4 presents the minutes past the hour. These derivations provide a means to produce a modified satellite DNI that better matches terrestrial conditions at Collinsville.

2.2.2 Temperature (Dry bulb)

Dry bulb temperature is the second of four drivers for yield in the SAM model. In this report "temperature" means "dry bulb temperature". In contrast, wet bulb and dew point temperatures are referred to explicitly.

2.2.2.1 Inter-site comparisons of dry bulb temperature

Table 3 shows a wider range of temperatures that is the difference between maximum and minimum temperatures in Collinsville than in the three coastal towns. The higher maximum temperatures that usually occur during mid-afternoon temperature, and the lower minimum temperatures that usually occur during early morning, are a consequence of the higher altitude compared to coastal locations. The sea breeze cools the coast sites during the day and land breeze moderates the loss of heat during the night. Consistent with these differences in climate, Collinsville has fewer cloudy nights and heavier dew. Section 2.3.6 discusses the dew effect further.

There is a relationship between elevation and temperature but this relation is complex. Table 3 contrasts the elevations of Collinsville at 197 m with three nearby comparison sites whose elevations range from 4 m to 10 m. Complexity stems, in part, from three different lapse rates that are changes in temperature per change in elevation. These lapse rates help explain cloud dynamics. The National Oceanic and Atmospheric Administration (NOAA 2014) provides a dry adiabatic temperature lapse rate (DALR) near 9.6 °C /km and a saturated adiabatic lapse rate (SALR) near 6 °C /km. The adiabatic condition provides the rate of loss of temperature of a parcel of air that does not swap energy with its surroundings, such as an idealised cloud. The environmental lapse rate (ELR), that is, for the air outside the parcel, is about 6.5 °C /km (Fovell 2010). These lapse rates vary from place to place and over time but they provide some guidance for a temperature sensitivity analysis on yield between the Collinsville and the three comparison sites.

The climatic differences between Collinsville and its comparison sites, has implications for temperature and thus yield. Comparatively, Collinsville has a cold wet start in the morning

but Collinsville's temperatures are close to the other sites by 9 am and surpass them by 3 pm. So, even if the daily yield from the comparison sites were the same, this shift in yield from early morning to late afternoon has implications, as the prices for electricity in Queensland is usually higher in the late afternoon than in the early morning.

The implications for temperature and, thus, yield for the different climates, calls into question the suitability of using the complete historical meteorological data from the three coastal towns in yield calculations for Collinsville. Relatively higher electricity price in the late afternoon compound this climate issue.

2.2.2.2 Collinsville temperature data

At the BoM's Collinsville Post Office weather station, in operation since 1939, there are currently three daily temperature measurements taken at 6 am, 9 am and 3 pm. This BoM coverage is far short of the hourly input required for the SAM model but at least the 3 measurements are taken during the daylight hours when yield is modelled. The BoM also provides daily maximum and minimum dry bulb temperatures measured daily at 9 am for the previous 24 hours. In contrast, Allen (2013) provides temperature readings each minute but coverage only starts in December 2012. This is far short of the 2007-2013 yield projection requirements of the subsequent report called 'Energy Economics and dispatch forecasting' (Bell, Wild & Foster 2014a).

As discussed, radiant energy causes temperature changes; temperature changes cause pressure changes and pressure gradients cause winds. Therefore, this relationship provides additional variables to model temperature. Radiant energy indicators are the BoM's hourly DNI, GHI and DHI and daily total solar exposure derived from satellite images discussed in the previous section. Wind direction and speed are taken thrice daily at 6 am, 9 am and 3 pm at the BoM's Collinsville Post Office weather station. Atmospheric pressure lacks coverage at the BoM's Collinsville weather station. The following sections discuss wind speed, wind direction and alternative indicators for atmospheric pressure.

In addition to the direct relationships just discussed there is the hydrologic cycle, which acts to ameliorate temperature differences and whose available indicators include relative humidity, cloud cover, precipitation, evaporation, dew point and wet bulb temperatures. Therefore, these indicators provide additional variables to model temperature and are measured thrice daily at BoM's Collinsville weather station, excepting evaporation and precipitation are measured once daily.

Section 3 discusses further the use of these indicators in modelling temperature.

2.2.3 Relative humidity

Relative humidity (RH) is the third of the four drivers for yield in the SAM model. To inform the discussion below, a brief description of the relationship amongst RH and the three temperature dry bulb, wet bulb and dew point is provided. RH is the ratio between vapour supply and vapour capacity. The dew point temperature indicates vapour supply as it is the lowest air temperature at which the current vapour supply remains unchanged that is before saturation is reached. The vapour capacity is a function of dry bulb temperature. Wet bulb temperature indicates the coolest air temperature achievable by evaporation (Fovell 2010, p. 21).

2.2.3.1 Inter-site comparisons of relative humidity

Table 3 shows the relative humidity at MacKay is the least comparable to Collinsville. The 9 am relative humidity at Collinsville, Townsville and Rockhampton is comparable. However, the 3 pm relative humidity at Collinsville is much lower than the RH at the comparison sites. The decrease in Collinsville's RH from 66% at 9 am to 43% at 3 pm is explained by both the water vapour supply decreasing indicated by the dew point temperature decreasing from 16.3°C to 14.2°C and the water vapour capacity indicated by the dry bulb temperature increasing from 23.2°C to 29.3°C. Exacerbating this effect is the feebleness or absence of a cooling sea breeze at Collinsville to moderate the afternoon rising temperatures and provide further moisture. This situation contrasts to the coastal comparison sites.

2.2.3.2 Collinsville relative humidity data

BoM's Collinsville weather station provides thrice-daily RH data. As for the related variables, the weather station also provides thrice-daily measurements for three temperatures: dry bulb, wet bulb and dew point and daily measurements for evaporation, precipitation and solar exposure. Another consideration is wind direction as a sea breeze could moderate temperature and increase the supply of water vapour in contrast a land breeze could exacerbate the rising afternoon temperatures and reduce the supply of water vapour. The weather station provides thrice-daily wind direction data.

2.2.4 Pressure

Pressure is the last of the four drivers for yield in the SAM mode. As there is an absence of atmospheric BoM data for Collinsville, the use of the ideal gas law becomes invaluable to the following discussion. The ideal gas laws in Equation 2 stipulates that pressure, temperature and density are dependent on one another, meaning a change in one cause a change in one or more of the others.

Equation 2: Ideal gas law

$$p = \rho r t$$

Where p = pressure (Pascals)
ρ = density
r = proportionality constant
t = temperature (Kelvin scale)

2.2.4.1 Inter-site comparisons of pressure

Table 3 contrasts the elevations of Collinsville at 197 m with three nearby comparison sites whose elevations range from 4 m to 10 m. As elevation increases, the proportion of atmosphere bearing down decreases, so reducing air density. The ideal gas law indicates that there is a corresponding decrease in temperature and/or pressure with an increase in elevation. This is indeed the case within the troposphere where the ELR for temperature is 6.5 °C/km and pressure is 1.2 kPas/100 m (Fovell 2010). The sensitivity of yield to elevation via the associated changes in temperature and pressure is an issue when using the nearby sites as proxies for yield at Collinsville. This implies a sensitivity analysis is warranted to investigate the relationship.

2.2.4.2 Collinsville pressure data

The absence of BoM atmospheric data requires consideration of other variables to model pressure, such as those variables in the direct relationships of irradiance causing temperature change, and temperature change causing pressure change and pressure gradient causing wind. Furthermore, the hydrological cycle moderates temperature change, so these hydrological variables also require consideration. Previous sections discuss the availability of these variables.

Atmospheric tides are regular cyclic changes in the atmospheric pressure over periods of 12 or 24 hours. These are small oscillations at low elevations but are amplified at high elevations. The tides are driven largely by solar irradiance and to a lesser extent the lunar cycle. These atmospheric tides are mostly driven by solar irradiance, so both a daily and annual cycle in the pressure is expected. There is an extensive literature on atmospheric tides. However, the National Oceanic and Atmospheric Administration (NOAA 2012) consider the most basic change in pressure occurs twice daily with maximums at 10 am and 10 pm and minimums at 4 pm and 4 am. Section 3 discusses implementing this basic cycle to represent solar tides and modelling pressure.

2.2.5 Why not use wind speed as a fifth driver?

SAM fails to include wind in its calculation of yield to allow for a chill factor. This could be considered a major oversight but building a linear Fresnel technology plant in a site with high winds is unlikely as the plant would be subject to damage. A consideration in the optimal positioning of linear Fresnel technology plant is low wind speed and at low wind speeds, the chill factor can be ignored in modelling.

There is considerably lower wind speed in Collinsville than in the three comparison sites in Table 3. This is consistent with the three comparison sites being subject to the sea breeze cycle and Collinsville being situated inland at higher elevation and within a valley.

The higher wind speed at Collinsville's three coastal neighbours makes both SAM unsuitable to model the yield from these sites and the sites unsuitable to build linear Fresnel technology plants. However, the exclusion of wind speed from SAM's calculation of yield does make the yield calculated from these comparison sites more comparable with the yield from the Collinsville.

We disregard wind speed as a driver in the calculation of yield but wind speed is present in the direct relationships flowing from solar irradiance, temperature, pressure to wind. Therefore, we consider wind's suitability as a variable to model the four drivers. Like temperature and humidity, wind speed is measured thrice daily by BoM, but unlike temperature and humidity, whose change is slow, wind speed can vary greatly. This makes wind speed less amenable to interpolation using three measurements. However, wind direction is more consistent so more amendable to interpolation. Section 3 discusses these issues further.

2.3 Effective Direct Normal Irradiance

The previous section discussed DNI as the first of the four drivers of yield in the SAM model but there is a requirement to introduce the concept of "effective DNI" that is the component of DNI that a CSP plant can use. We frame the concept within two effects: the dew and dust

effects. The discussion of the effects both simultaneously help crystallise the concept of effective DNI and ameliorates concerns about the discrepancy between the measurement of DNI and effective DNI.

2.3.1 Dew effect and effective DNI

The dew effect involves dew collecting on LFR or DNI measuring instruments where both require warming and evaporation by the sun before the suns energy can be harnessed to produce electricity, whether by the LFR or the instrument measuring DNI. The dew effect is considerable in places with clear nights low wind speed, which describes the weather at Collinsville in Table 3. The clear cloudless nights allow cooling of the earth surface via reradiating heat into outer space and low wind speed allows the cooling of the air close to the ground, so the air precipitates its moisture. Collinsville has the lowest average minimum annual dry bulb temperature of the four sites. However, the dew effect can be ignored because both the measuring instrument for DNI and the LFR plant are subjected to the dew effect. The DNI that is measured by the instrument is an "effective" DNI.

The automatic adjustment for the dew effect on ground based measurement to read effective DNI is absent in satellite data. The dew effect makes the unmodified use of satellite DNI data questionable.

2.3.2 Dust effect and effective DNI

This section discusses the dust effect with the following hierarchy

- Dust-in the atmosphere
- Dust-on
    - the LFR
    - the measuring instrument

The "dust-in" the atmosphere that attenuates DNI can be modelled along with other aerosols in the atmosphere. This modelling assumes that the surrounding natural or manmade dust producing activities remain consistent between model calibration and projection periods. However, a change in coal mining intensity or methodology could affect dust levels or a change in wind patterns. Wind patterns can change because of El Niño–Southern Oscillation or climate change, which the next section discusses.

"Dust-on" the LFR reduces the effectiveness of DNI to heat water. The Mechanical and Mining Engineering (MME) School at UQ (Guan, Yu & Gurgenci 2014) reports on the cleaning requirements to address dust-on the LFR.

Similarly, "dust-on" Allen's (2013) measuring instruments reduces the amount of DNI measured, so only effective DNI is measured. Allen (2013) discusses the dust and cleaning of the measuring instruments.

As with the dew effect, both the measuring instrument for DNI and the LFR are subject to dust effects, so in a simplifying assumption the measured DNI can be considered the "effective" DNI for the LFR. However, if the MME group provides an economical solution to cleaning dust from the LFR more frequently than the cleaning of dust from the measuring instrument, the yield calculated from the dusty measuring instrument will slightly underreport the true yield. The converse will hold if Allen cleans the measuring instrument more often

than the LFR are cleaned. This section acknowledges the dust effect but ignores the effect as being self-compensating via the “effective” DNI reasoning and too complex to model exactly given the uncertainty over the future actions of Allen, the MME group, El Niño–Southern Oscillation and climate change.

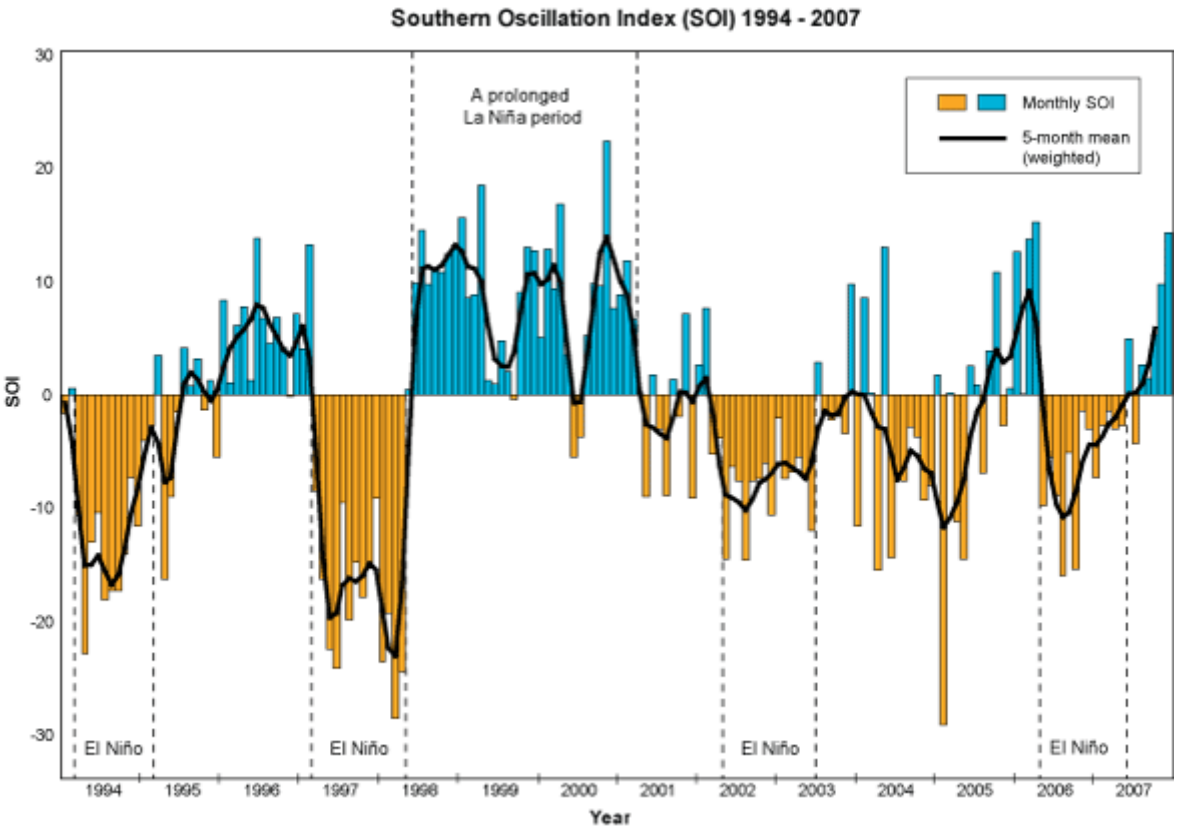
Satellite DNI data may be adjusted for the first component of the dust effect that is dust-in the atmosphere, but the self-compensating “effective” DNI reasoning in the dust-on effect is absent in satellite data. Consideration for the dust-on effect makes the unmodified use of satellite DNI data questionable.

2.4 The effect of El Niño–Southern Oscillation on yield

This section discusses the effect of El Niño – Southern Oscillation (ENSO) on the four drivers for yield. The previous sections discuss the four drivers within the context of regular daily or annual cycles. In contrast, the ENSO is irregular and can spans more than a year. Therefore, there is an expectation of many ENSO cycles during the lifetime of the proposed plant at Collinsville.

BoM (2005) discusses ENSO within a worldwide context. In contrast, this section discusses ENSO implications for Collinsville, Queensland. The ENSO spans the Pacific and consists of two main phases: the La Niña and El Niño phase. La Niña is considered normal weather phase within the cycle whereas El Niño is considered an abnormal weather phase. During La Niña, there is an atmospheric convection current established between Queensland and Central America by the warmer waters off Queensland and cooler waters of Central America. This convection cycle is called the Walker Circulation and produces the trade winds blowing east to west. The trade winds crossing the Pacific are high in moisture when they reach Queensland and the Walker circulation causes the trade wind to ascend over Queensland encouraging precipitation from the moisture-laden air. Another consequence of the Walker circulation is the relatively low pressure over Queensland compared to the mid Pacific. The Southern Oscillation Index (SOI) in Figure 1 shows La Niña and El Niño phases indicated by this pressure difference. However, the SOI uses the difference in pressure between Tahiti and Darwin.

Figure 1: Southern Oscillation Index 1994-2007



(Source: BoM 2005)

In the El Niño phase, the water surface temperature in the Eastern and mid Pacific warms disturbing the Walker Circulation. For Queensland, the atmospheric convection current now runs counter to the trade wind and high-pressure forms over Queensland relative to mid and Eastern Pacific. The air arriving in Queensland is now dry.

In summary, relative to the La Niña phase, the El Niño phase brings higher pressures, weaker winds and less water vapour, which results in lower humidity, fewer clouds and rain. Fewer clouds and rain improve DNI. The El Niño phase also brings higher day time temperatures and lower night time temperature because the moderating effects of moisture is reduced.

Consequently, the El Niño phase produces higher yield. Additionally, the higher temperatures drive higher prices for electricity in Queensland. So, El Niño events could prove a profitable time for CSP plants. This comes with the caveat that the El Niño induced increase in bush fires fail to attenuate DNI.

The ENSO cycle has implications for finding a “Typical” representative year for a TMY. The TMY section discusses this issue further.

2.5 The effect of climate change on yield

This section discusses the effect of climate change on the four drivers for yield. The previous sections discuss the effect of weather cycles on the four drivers. In contrast this

section discusses the effect of the more permanent and gradual change in climate over the expected lifetime of the proposed plant.

Climate change is a global phenomenon whose focus is on the average rise in global temperatures but the main driver for CSP is DNI. Nevertheless, studying the temperature change and the associate changes in other variables provides a useful background to the issue. Additionally, Climate change focuses on global temperature change but the local effects can run counter to the global effect, as seen with the ENSO, a rise in temperature in one area can cause disruptions to normal weather patterns whose effects can be uneven.

Consequently, there requires some discretion in selecting Global Climate Models (GCM) that report the most likely, hottest or coolest cases for the geographic area of interest and provide the range of variables required for analysis. For this report, there is a tension over the selection of the geographic area because selecting the National Electricity Market (NEM) as a the geographic area will best reflect the demand and price for the electricity produced but selecting GCMs for Collinsville will best reflect the yield. Foster et al. (2013) have already conducted an analysis for the NEM for five variables, including three of the four drivers, but their focus is temperature rather than DNI. Their choice of the carbon emission scenario is SRES A1FI, which best reflects the high carbon emissions trajectory currently occurring around the world. Clarke and Webb (2011) select three GCMs from 23 GCMs reflecting two extremes and an average case for Foster et al. (2013):

- Most likely case – MRI-CGCM2.3.2
- Hottest case – CSIRO-Mk3.5
- Coolest case – MIROC3.2

For the five environment variables:

- solar radiation;
- temperature;
- relative humidity;
- wind speed;
- rainfall.

The hottest case is the worst case from a climate change perspective but the hottest case could be the best case from an LFR perspective because higher temperatures help provide more yield and increase electricity demand in Queensland.

Table 5 shows the projected change in climate from 1990 to 2040 for the location at latitude and longitude (-20.5, 148) from the ozClim projection series (CSIRO 2011; Page & Jones 2001). This location is the closest to the proposed plant at (-20.5344, 147.8072). Notable is the ordering of the projected mean temperature change where the most likely case is smaller than both the coolest and hottest cases. As discussed earlier, the local effect can run counter to the global effect. The fourth driver, pressure, is omitted from the table as ozClim (CSIRO 2011; Page & Jones 2001) lacks pressure projections.

Table 5: projected change in climate from 1990 to 2040

	Coolest case MIROC3.2-Medres	Most likely case MRI-CGCM2.3.2	Hottest case CSIRO-Mk3.5
Solar radiation (%)	-1	0.1	0.8
Temperature Mean (°C)	1.21	1.04	1.33
Relative humidity Mean (%)	0.8	-0.7	-0.9

(Source: CSIRO 2011; Page & Jones 2001)

Therefore, the most likely expected percentage change in solar radiation, the main driver for yield, from 1990 to 2040 is 0.1 percent. The change in temperature is just over 1 °C and a decrease in humidity is 0.7 percent. These three changes taken together would increase yield but only by a tiny amount. Similarly, in the hottest case, the changes would act to increase yield slightly. In coolest case the changes may slightly decrease yield. Sensitivity analysis could provide a more exact estimate.

2.6 Typical Meteorological Year and Wholesale Spot Prices

This section discusses the use of the Typical Meteorological Year (TMY) with consideration to matching electricity demand data for the given metrological conditions and the ensuing wholesale spot price and dispatch calculations. The ensuing 'Energy Economics and dispatch Forecasting report' (Bell, Wild & Foster 2014a) uses the TMY yield projects from this report to help forecast wholesale prices and dispatch.

2.6.1 TMY both a technique and format

Marion and Urban (1995) and Wilcox and Marion (2008) provide user manuals for the collection and processing of data to produce TMY2 and TMY3 data files that are TMY version 2 and 3. TMY is both a format and a technique. SAM can use both TMY3 and TMY2. This report uses the TM3 format and we introduce a modified TMY technique. As a format, the TMY files are an hourly record of selected weather variables for an entire year for a specific location. Importantly, TMY's hourly data represents the average of the weather variable for the previous hour. This representation contrasts with BoM's data that usually records the instantaneous reading.

Originally, the TMY technique was developed to calculate a hypothetical year that could represent a number of years ranging from 15 to 30 years to estimate the typical heating and cooling costs for buildings. However, the results of the TMY technique were extended for use within the renewable energy generation sector. The TMY technique involves finding the 12 most typical meteorological months (TMMs) from a range of years. The meteorological variables of interest are weighted according to their importance and the weighted average used to select the TMMs.

The advantages of the TMY technique include the simplicity of the technique, simplifying ensuing calculations, such as providing a single baseline year in sensitivity analysis. These factors in turn provide easy to explain results. The disadvantages include lacking analysis of the variability between years, so lacking P90 analysis, and subjectivity of assigning weights to each weather variable and the technology dependency of the weights. For instance, appropriate weights for a LFR and wind generator would differ considerably.

To address the subjectivity of the weights and their technology dependence, this report introduces a modified TMY technique that compares average monthly yield within the range

of years to determine the 12 TMMs. This simultaneously avoids the explicit assignment of weights to each of the four drivers: DNI, temperature, humidity and pressure and ensures the four drivers are implicitly weighted in a technologically appropriate way.

To address the lack of analysis of variability for CSP yield, Stoffel et al. (2010, p. 101) suggests using several years in the analysis rather than a single year or a TMY to assess the effects of inter-year variability. However, analysis of each year carries a large computational overhead that becomes excessive for any sensitivity analysis.

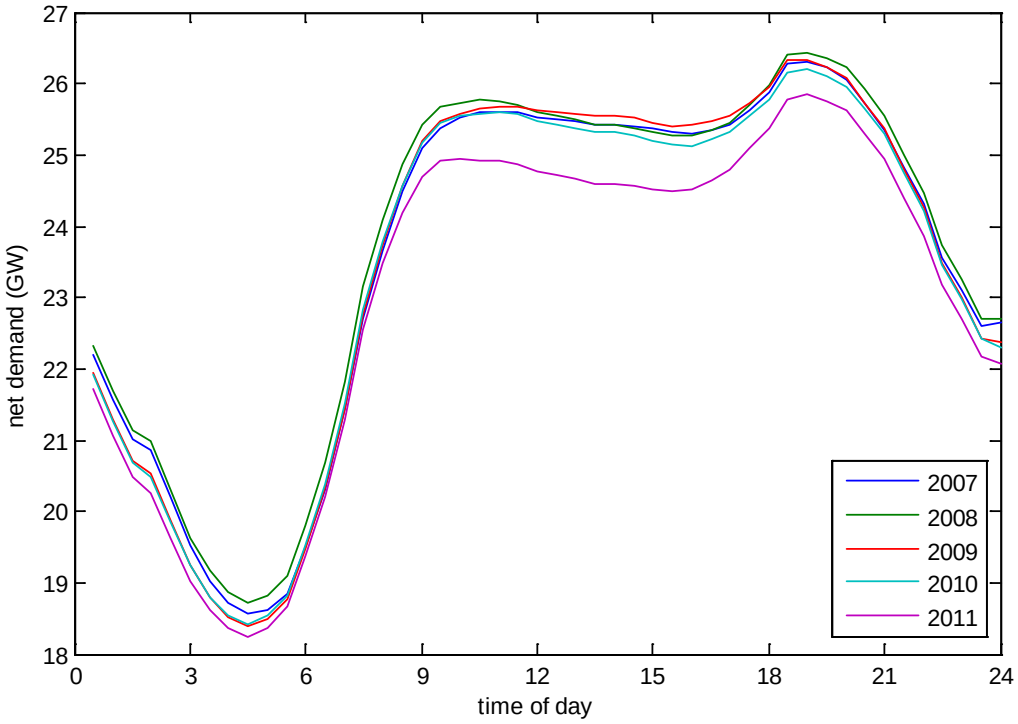
Whether to conduct sensitivity analysis or variability analysis requires an assessment of priorities. Section 7.1 discusses variability analysis for further research rather than the TMY method used in this report and the two subsequent reports.

2.6.2 TMY implications for demand and supply in the NEM

A long been established relationship is the dependence of electricity demand on weather but the production of electricity in a predominantly fossil fuel generation fleet is relatively independent of weather. However, with the introduction of more renewable energy, the production of electricity is becoming more dependent on the weather and since the marginal cost of the renewable segment is nearly zero, weather can now have an even more dramatic effect on spot prices.

Figure 2 shows the average demand across the NEM for the years 2007 to 2011 by time of day. Bell, Wild and Foster (2013) calculate that the increasing midday depression in net demand can be explained largely by the introduction of solar PV. This reduction in net midday demand is expected to continue with further solar PV installations. In contrast, the reduction in demand in the early hours of the morning can be explained by the introduction of solar water heaters replacing electrical water heaters that use off-peak power. This transformation of the net demand curve requires that the same TMMs calculated for Collinsville's LFR are consistently applied across the NEM to determine generation mix and net demand to calculate realistic wholesale spot prices. Bell, Wild and Foster (2014a) discuss in more detail the implications for net demand, wholesale prices and dispatch the ensuing 'Energy Economics and dispatch forecasting'.

Figure 2: NEM's net average demand for 2007 to 2011



(Source: Bell, Wild & Foster 2013)

2.6.3 ENSO implications for TMY selection

The requirement of the subsequent reports to select a TMY from the years 2007-12 and ENSO cycle have implications for finding a "Typical" representative year for a TMY. Selecting a TMY from a larger number of years would average out the ENSO cycle to find a more representative TMY but the constraints of the subsequent reports eliminate this possibility. However, a comparative analysis of yield from years 2007-2012 with earlier years could ameliorate this concern.

2.7 Conclusion

The literature review has both established the research questions and provided direction for the methodology to address these questions.

Motivating the research question is the questionability of using yield projections from nearby sites at MacKay, Rockhampton and Townsville as proxies for yield from Collinsville. The appeal of using these three comparison sites is their complete historical environmental datasets of the four drivers for yield: DNI, temperature, humidity and pressure. However, the literature review has established considerable differences in climate between Collinsville and the comparison sites. The comparison sites have costal climates moderated by the daily alternating cycle of the sea and land breeze. In contrast, Collinsville has a colder wetter start in the early morning but lacking the moderating sea breeze, temperatures surpass those of the coastal comparison site in the mid-afternoon. In Queensland, the price of electricity is generally higher in the late afternoon, so this climate engendered shift in yield production makes the yield more profitable.

Furthermore, the efficacy of using the raw hourly BoM (2013) DNI data derived from satellite images for Collinsville is questionable when comparing the daily total solar intensity derived from satellite images and cloud coverage at Collinsville and its three comparison sites. Additionally, the dew and dust effects make the use of raw BoM DNI satellite data questionable.

The BoM Collinsville Post Office weather station, in operation since 1939, provides thrice daily measurements for temperature and humidity but lacks any pressure data. This coverage is far short of the hourly coverage required by SAM to calculate yield. However, Allen (2013) provides one-minute data for all four drivers starting in December 2012 but this coverage is far short of the 2007-2013 yield projection period requirements of the subsequent reports.

The review introduces the concept of "effective" DNI to help ameliorate dew and dust effect concerns and a modified TMY technique to eliminate the need for technology specific weighting of environment variables. The methodology further develops these two items.

2.7.1 Research questions

The research questions arising from the literature review.

The overarching research question:

> *Can modelling the weather with limited datasets produce greater yield predictive power than using the historically more complete datasets from nearby sites?*

This overarching question has a number of smaller supporting research questions:

- Is BoM's DNI satellite dataset adequately adjusted for cloud cover at Collinsville?
- Given the dust and dew effects, is using raw satellite data sufficient to model yield?
- Does elevation between Collinsville and nearby sites affect yield?
- How does the ENSO affect yield?
- Given the 2007-2012 constraint, will the TMY process provide a "Typical" year over the ENSO cycle?
- How does climate change affect yield?

A further research question arises in the methodology but is included here for completeness.

- What is the expected frequency of oversupply from the Linear Fresnel Novatec Solar Boiler?

3 Methodology

3.1 Introduction

This section describes the methods used to address the research questions arising from the literature review in the previous section. The overarching research question:

> *Can modelling the weather with limited datasets produce greater yield predictive power than using the historically more complete datasets from nearby sites?*

This report uses the Systems Advisor Model (SAM) to calculate yield from the proposed LFR at Collinsville from the four drivers.

- DNI
- Temperature
- Humidity
- Pressure

The previous section established the questionability of using complete historical datasets of the four drivers from nearby sites to calculate yield as a proxy for yield at Collinsville. This questionability necessitated using datasets from Collinsville for the four drivers.

The hourly BoM DNI dataset starting in 1990 derived from satellite imagery for Collinsville meets both temporal requirements for this report. These requirements are an hourly dataset for SAM and the range of years, 2007-2012, for two subsequent reports. However, as discussed in the literature review, the dew and dust effects and ambiguity over cloud cover make the use of this DNI dataset questionable without modification for the aforementioned.

The BoM weather station at the Collinsville post office in operation since 1939 provides both temperature and humidity datasets but these datasets contain only thrice daily readings taken at 6 am, 9 am and 3 pm. This thrice daily reading is insufficient to meet the hourly requirement for SAM. In addition the weather station lacks any datasets for the fourth driver, pressure.

The above inadequacies of the BoM datasets require that models of the four drivers are developed to form the appropriate projections. These projections must satisfy the hourly requirement for SAM and the range of years, 2007-2012, for the subsequent reports.

Allen (2013) provides one-minute resolution terrestrial based measurements taken at Collinsville for all four drivers. Allen (2013) converted these one-minute datasets into hourly datasets to meet SAM's requirements. But Allen's (2013) datasets start in December 2012, which fails to meet the 2007-2012 requirement of the subsequent reports. However, Allen's datasets are suitable to calibrate models of the four drivers with the inadequate BoM datasets, as the Allen's (2013) datasets are derived from terrestrial based measurement and in an appropriate format for SAM.

Modelling the four drivers requires considering their explanatory variables for inclusion in a model. As discussed in the literature review, there are a set of direct relationships moderated by the hydrological cycle. The direct relationships include: solar irradiance causes temperature rise, temperature change causes pressure change and pressure gradients cause wind. There is considerable interrelation between the four drivers and their

explanatory variables, which is unremarkable, since we are dealing with weather cycles. There are 22 variables available to explain the four drivers. This presents two problems: highly correlated environmental variables and the curse of dimensionality.

There is a great possibility that the environmental variables are highly correlated or synchronised, so a subset of the variables, that are the most uncorrelated, could be selected to model the four drivers. The procedure adopted is known as principle component analysis. For instance, the fourth driver, pressure, may be adequately modelled with just three of the variables: temperature, wind direction, and month.

Regarding the 'curse of dimensionality', neural networks are used to develop the models as they are a standard tool within the electricity industry to analysis weather-demand relationships and are well suited to modelling non-linear systems, such as the weather (Deoras 2010; Hippert, Pedreira & Souza 2001). However, neural networks are non-communicative, that is the order of the explanatory variables affects the results of the fitted model. So, there are 22! (= 1.124×10^{21}) ways to order 22 variables. This simple factorial fails to account for all combinatorial possibilities with fewer than 22 explanatory variables that are potential models for the four drivers.

Using the Akaike Information Criteria (AIC) (Akaike 1974) within a pragmatic search routine to find a minimal set of explanatory variables eliminates the need to calculate every combination of explanatory variables. The AIC value helps to select between models and provides a trade-off between goodness of fit and model complexity. The number of variables *k* in the model indicates the level of complexity. For example, the fourth driver, pressure, could be modelled with either a simpler two-variable model (temperature and wind direction) or a more complex three-variable model (temperature, wind direction, and month). The first line in Equation 3 shows the generalised AIC form and the last line shows the residual sum of the squares (*RSS)* form (Burnham & Anderson 2002, p. 342) used in this report. The *RSS* form assumes that the errors are normally distributed with a mean of zero and independent. In that case, the likelihood function *L* is the residual sum of the squares divided by the number of observations *RSS/n* for large values of n. In model selection, the model with the smallest AIC is preferred. The *2k* provides a penalty for model complexity and the natural log of the likelihood function *2ln(L)* provides a measure of goodness of fit. In model comparison, the constant *c* can be ignored.

Equation 3: Akaike Information Criteria

$$AIC = 2k - 2\ln(L) \quad \text{- general form}$$
$$AIC = n \ln(RSS/n) + 2k + c \quad \text{- } RSS \text{ form use in this report}$$

Where

L = likelihood function
k = number of variables
RSS = residual sum of the squares
c = constant

The selected models are used to produce projections of the four drivers for the years 2007-2013, which SAM uses to calculate yield for 2007-2013. As discussed in the literature review, there is the option whether to use a single Typical Meteorological Year (TMY) in a

sensitivity analysis on gas prices or analyse individual years to calculate inter-year variability and develop a P90 for lifetime revenue. However, obstacles to acquiring data in a timely fashion for this report precluded the possibility of the variability analysis. Therefore, this report calculates a single TMY to enable a sensitivity analysis on gas prices in the subsequent report (Bell, Wild & Foster 2014a). The TMY is constructed from the years 2007-2012. These years correspond to datasets of electricity demand and renewable energy generation in the subsequent reports. The Typical meteorological month (TMM) is the month in the years 2007-2012 that has the yield closest to the mean of the months from these years. The TMY is made from the 12 TMMs. For the year 2013, the yields calculated from the datasets of the four drivers from the driver models and from Allen (2013) are compared to help validate the models.

The following sections elaborate on the process outlined above where necessary. Section 2 discusses preparation of the datasets. Section 3 discusses selecting the best models for the four drivers. Section 4 discusses modelling yield from the four drivers. Shah, Yan and Saha (2014b) present the methodology for calculating yield from the historical datasets.

3.2 Preparing the data

This section discusses the preparation the datasets for use in this report. The outline below shows a hierarchy of the datasets and their functional use.

- Target or dependent variable
    - Allen's (2013) datasets
- Input or explanatory variables
    - BoM's hourly solar satellite data
    - BoM's Collinsville Post Office weather station
    - Other

There are 4 target variables and 22 explanatory variables available.

3.2.1 Allen's datasets: Target or dependent variables

Allen (2013) collects one-minute data from a terrestrial weather station at Collinsville for the four drivers of yield. Allen (2013) converts the one-minute data into an hourly form specifically to meet the requirements of SAM. SAM's requirement for hourly data is the average of the instantaneous values of the previous hour.

3.2.2 BoM's Collinsville Post Office datasets: Input or explanatory variables

BoM observes data at three different frequencies at Collinsville:

- Once daily
- Thrice daily
- Six times daily

The frequency of measurement of the environmental variable determines their preparation, so the following discussion groups the datasets or variables by frequency. Interpolation of missing values was performed using the average of the previous and following day. The thrice or six times daily measurements were interpolated using the measurements taken at the same time the previous day and next day.

3.2.2.1 Once daily data

The outline below shows the datasets with a single daily data reading.

- Temperature dry bulb
  - Maximum
  - Minimum
- Evaporation
- Solar exposure

BoM takes daily data readings at 9 am for the previous 24 hours, except for the total daily solar exposure, which is estimated from satellite images

This report assumes that minimum temperatures occur in the early hours of the morning and maximum temperatures occur in the remaining hours of the day. So, the same minimum dry bulb temperature is assigned to the hours 00 to 09 for that day and the same maximum dry bulb temperature is assigned to all hours 10 to 23 for the previous day.

Evaporation is a daily rate, so the value was simply assign to the hours 00 to 09 that day and the hours 10 to 23 for the previous day.

The solar exposure was simply assigned to every hour of the day. The modelling is only required for the daylight hour, so the paradox of assigning solar exposure to the night time hours is a non-issue.

3.2.2.2 Thrice daily data

Thrice daily readings are taken at 6 am, 9 am and 3 pm for the following variables.

- Temperature
  - Dry bulb
  - Wet bulb
  - Dew point
- Relative humidity
- Wind
  - Speed
  - Direction
- Cloud cover
- Visibility
- Precipitation

As discussed in the literature review, wind speed and direction are fickle and unsuitable for interpolation with such low-resolution datasets. However, the other environment variables are slower changing, so are more amenable to interpolation and modelling. Additionally, only the daylight hours require modelling, which in effect doubles the resolution of the thrice-daily readings. The thrice-daily readings were simply interpolated with the following exceptions.

Precipitation is a cumulative measurement. In contrast, the other variables are instantaneous measurements. Therefore, precipitation is converted into a rate and the rate simply assigned to the relevant hours.

An improvement in the interpolation for wind direction should be considered in further research for two reasons. Firstly, the wind direction 360° represents north and 0° represents no-wind or calm, so a simple interpolation between 0° and a positive value is misleading. Secondly, the wind may simply switch directions, for instance from a land breeze to a sea breeze, which also makes interpolation misleading.

3.2.2.3 Six times daily data – weather present and past

The BoM uses the codes in Table 6 to record weather phenomena present at 6 am, 9 am and 3 pm and in the past hours. These thrice-daily past and present weather recordings in effect give six daily readings. The codes for the present times are simply assigned to the respective hour of the day and the codes for the past times are assigned to the previous intervening hours. For example, the code for the past weather reading at 9 am is assigned to 7 am and 8 am.

Table 6: BoM's past and present weather phenomena types and codes

4	Smoke	38-39	Blowing snow
5	Haze	40	Distant fog
6-7	Dust	42-49	Fog
8	Dust whirls	50-55	Drizzle
9	Dust storm	56-57	Freezing drizzle
10	Mist	58-59	Drizzle
11,41	Fog patches	60-65	Rain
12	Shallow fog	66-67	Freezing rain
13	Lightning	68-69	Sleet
14	Distant/nearby virga	70-75	Snow
15-16	Distant precipitation	76	Ice prisms
17	Thunder	77	Snow grains
18	Squall	78	Starlike crystals
19	Funnel cloud	79	Ice pellets
20	Recent drizzle	80-81	Shower
21	Recent rain	82	Violent shower
22,26	Recent snow	83-84	Sleet
23	Recent rain and snow	85-86	Snow shower
24	Recent precipitation	87-88	Soft hail shower
25	Recent shower	89-90	Hail shower
27	Recent hail	91-95	Thunderstorm
28	Recent fog	96,99	Thunderstorm and hail
29	Recent thunderstorm	97	Heavy thunderstorm
30-32	Dust storm	98	Thunderstorm and dust
33-35	Severe dust storm		

(Source: BoM 2011)

However, there are issues with interpolating the missing codes. For instance averaging code 4 for smoke with code 98 for "Thunderstorm and dust" gives a code 51 for drizzle. In further research, a more sophisticated algorithm is required to handle the missing readings. Additionally, the weather phenomena are a composite of existing variables such as visibility, relative humidity, precipitation, temperature, cloud cover, and evaporation. Therefore, the codes in Table 6 are an index of components already modelled.

3.2.3 BoM's Satellite datasets: Input or explanatory variables

As discussed in the literature review, the BoM's hourly satellite data for DNI and GHI requires adjustment for minutes past the hour according to the latitude of the observation, see Table 4. The times of the BoM's DNI and GHI readings are adjusted for the minutes past the hour then interpolated to provide the readings on the hour to match those datasets from BoM's Collinsville post office weather station.

3.2.4 Other input or explanatory variables

This section discusses other variables derived from the BoM datasets or otherwise. The following outline groups these variables by type.

- Astrological angles
    - Azimuth
    - Zenith
    - Altitude
- DHI
- Pressure represented as a sine wave
- Time
    - Month
    - Hour

The astrological angles are calculated for every hour of the year for Collinsville, using the algorithm described in section 2.2.1.2.

Equation 1 calculates DHI using the original DNI and GHI datasets from the previous section. DHI is then adjusted for the satellite latitude / minute past the hour deviation as described in the previous section.

Pressure is modelled as a sine wave with maximums at 10 am and 10 pm and minimums at 4 pm and 4 am (NOAA 2012). This is intended to capture the atmospheric tide as discussed in section 2.2.4.

Lastly, any other daily and seasonal effects can be captured using the hour of the day and month of the year as variables.

3.3 Selecting the best model for the four drivers

As discussed in section 3.1, AIC is used in a pragmatic search routine to select a minimal set of explanatory variables for each of the four drivers. The routine addresses two problems: highly correlated environmental variables and the curse of dimensionality. This section discusses the steps in the search routine.

3.3.1 Step 1 – finding the besting fitting one-variable models

The first step involves finding the first explanatory variable for each of the four drivers to provide the best fitting single-variable model. This involves simply calculating R-squared (R^2) values for each of the 22 explanatory variables and selecting the explanatory variable with the highest R^2 value. AIC is unsuitable for this first step because AIC fails to convey information in an easily interpretable way about goodness of fit of the model whereas R^2 does. Equation 4 shows the calculation of R^2. In this report the total sum of squares SS_{tot} is

the variability in Allen's (2013) dataset and the residual sum of squares SS_{res} is the square residuals between the fitted model and Allen's (2013) datasets. So, an R^2 value closer to 1 denotes a better fit as the SS_{res} approach zero.

Equation 4: R-squared a measure of a model's goodness of fit

$$R^2 = 1 - SS_{res} / SS_{tot}$$

Where

SS_{res} = Residual sum of squares
SS_{tot} = Total sum of squares

3.3.2 Step 2 – selecting two-variables models

In the second step the best fitting explanatory variable for each of the four drivers is used to form 21 two-variable models by appending one of the remaining 21 explanatory variables. For instance if the explanatory variable, *month*, provides the best fitting single-variable model for the driver, *pressure*, then the driver, *pressure*, is modelled with the following two-variables models: (month, temperature), (month, humidity), (month, hour), and so forth.

3.3.3 Step 3 – selecting three-variable models

The two-variable model with the lowest AIC is selected and the remaining 20 explanatory variable are used to form 20 three-variable models. For instance if the two-variable model, (month, hour), provides the lowest AIC value for the driver, pressure, then the driver, pressure, is modelled with the following three-variable models: (month, hour, temperature), (month, hour, humidity), (month, hour, DNI), and so forth.

3.3.4 Step N – iterating through N-variable models until information is exhausted

The above routine is iterated until there lacks any significant decrease in AIC. At this point the information value in the remaining explanatory variable has been exhausted and adding further explanatory variables to the model just introduces noise into the results.

3.3.5 Neural network internal weights affecting the number of variables in value k

All the neural networks in this report have 10 internal weights that are optimised to provide model fit. These weights in effect add extra variables to the models by increasing the value of *k* in Equation 3. This increases the value of AIC but the effect of these weights on AIC can be ignored because a constant number of weights were used throughout the report for all models and the value of AIC is only used in model comparison where such constants can be ignored.

However, the weight effect in Equation 5 adjusted-R^2 (adj-R^2) cannot be ignored. The adj-R^2 extends the R^2 for single-variable models in Equation 4 for use with multi-variable models. Failing to allow for the weights will slightly over report the adj-R^2 value.

Equation 5: Adjust R-squared a measure of goodness of fit for multi-variable models

$$\text{Adj-}R^2 = R^2 - (1 - R^2)\, k / (n - k - 1)$$

Where

k = number of variables in the model

n = number of observations used in the model

Nevertheless, the effect of the 10 weights on the adj-R^2 values in Equation 5 is trivial as looking at the worst case scenario can show. Calculating the term $k / (n - k - 1)$ for a single-variable model with and without the weights that is $k = 1$ & 11 and with the number of observation as the number of daytime hours in year that is n = 4380, which gives the following results *0.000228 & 0.002518,* respectively. The overall effect on adj-R^2 is less than 0.25% as the term $(1 - R^2)$ is always less than 1. Ameliorating the effect even further is the degrees of freedom of the 10 weights requiring only adding 9 weights to the term *k*. Furthermore, only a fraction of each weight may require reflecting in the term *k*. This issue is left for further research but for this report the issue can be ignored.

3.3.6 Neural network and variability of AIC and adj-R^2

In the above steps, the mean adj-R^2 or AIC of10 simulations is used because the goodness-of-fit of each simulation of a neural networks can differ slightly. This variation arises because there is a random assignment of the data into segments for specific purposes: training (70%), validation (15%) and test (15%) (MathWorks 2014a). Where the training set provides the data to find the best fit; the validation set provides data to prevent over fitting the training data (MathWorks 2014b); and the test set provides data that is independent of both training and validation. This test set independence offers predictive falsifiability of the fitted model. Running the neural network over a number of simulations and averaging the adj-R^2 or AIC values helps improve the veracity of the results because in each simulation the data is randomly assigned into the training, validation and testing sets.

3.4 Calculating yield with the Systems Advisor Model from four drivers

The NREL's SAM model (Wagner 2012; Wagner & Zhu 2012) calculates the hourly yield for LFR given hourly values for the four drivers in TMY format. The company Novatec Solar will provide the LFR technology for Collinsville. SAM has a sample file for a "Linear Fresnel Novatec Solar Boiler". This file contains all the default parameters for a standard Novatec Solar installation. Table 7 shows the changes from the default setting advised by Novatec Solar.

Table 7: Advised default setting changes to SAM's 'Linear Fresnel Novatec Solar Boiler'

Field groupings	Input fieldnames	Advised value	Default value
Solar Field parameters	Number of modules in boiler section:	17	12
	Number of modules in superheater section:	6	6
	Collector azimuth angle:	-10°	0
Steam Conditions at design	Field outlet temperature:	500°C	500
	Turbine inlet pressure:	120bar	90
Plant Design	Design turbine gross output:	30.07 MWe	49.998

(Source: Glaenzel 2013)

The increase in the number of modules in the boiler section is consistent with an increase in 'turbine inlet pressure'. However both these changes are expected to increase the default 'design turbine gross output' but the default has been decreased from 49.998 to 30.07MW. This implies the plant will exceed the 30 MW AEMO imposed dispatch limit under ideal climatic conditions. The oversizing of the boiler allows the plant to contribute to the grid closer to its 30 MW limit under less than ideal climatic conditions. However, this may involve some spillage of excess supply but the amount is uncertain. The frequency of exceeding the dispatch limit is added to the research questions.

The collector azimuth angle *-10*° means a 10° inclination to the west, which allows the plant to maximise output during the later afternoon when electricity prices are typically higher in Queensland.

3.5 Conclusion

This section building on the literature review has discussed the methodologies ready to apply to the research questions to provide the results in the next section.

The overarching research question:

> *Can modelling the weather with limited Collinsville datasets produce greater yield predictive power than the more extensive datasets from nearby sites?*

Has a supplementary question:

> *What is the expected frequency of oversupply from the Linear Fresnel Novatec Solar Boiler?*

4 Results and analysis

4.1 Introduction

This section presents the results from running the simulations described in the methodology in Chapter 3 to address the research questions arising in the literature review:

> *Can modelling the weather with limited datasets produce greater yield predictive power than using the historically more complete datasets from nearby sites?*

Section 2 presents the results from modelling the four environment variables that are drivers for yield in the Systems Advisor Model (SAM). The model is calibrated using the year 2013 dataset from Allen (2013).

4.2 Selecting the best models for the four drivers

Sections 3.2 and 3.3 discuss the preparation of the data and methodology for this section. The four drivers for yield calculations in SAM are:

- DNI
- Temperature (Dry bulb)
- Humidity
- Pressure

Allen (2013) provides the datasets for the four drivers from his observations at Collinsville. These four drivers are the target or dependent variables. BoM provides most of the 22 input or explanatory variables. Section 3.2 provides details.

4.2.1 Step 1 – selecting the one-variable models

Table 8 shows the mean adj-R^2 values for the four drivers against the 22 input or explanatory variables ranked by descending mean adj-R^2. The mean adj-R^2 of 10 simulations is used because the results from each simulation of a neural network can differ slightly. Section 3.3.6 discusses this issue.

The selection of the first explanatory variable for DNI, temperature, and relative humidity is unsurprising. Selecting the first explanatory variable for pressure is more vexing but Section 2.2.4 discusses the moderating effect of the hydrological cycle and the direct relationships: temperature causes pressure changes and pressure gradients cause wind. Consistent with these relationships, Table 8 (d) shows that four forms of temperature measurement rank within the six highest mean R^2 explanatory variables. Month and Azimuth also feature in the highest six, which would reflect the annual atmospheric tide discussed in Section 2.2.4. However the mean R^2 values for wind speed and direction indicate no fit. As discussed in the literature review the three daily observations for wind is insufficient for such a fickle variable.

Table 8: Step 1 - selecting the one-variable models for the four drivers using R^2

Rank	(a) DNI		(b) Temperature		(c) Humidity		(d) Pressure	
	Explanatory Variables	mean adj-R^2		mean adj-R^2		mean adj-R^2		mean adj-R^2
1	dni	0.81	temp	0.93	hum	0.81	wet	0.57
2	ghi	0.50	maxMin	0.85	hour	0.55	mon	0.55
3	dhi	0.39	wet	0.62	maxMin	0.51	dew	0.40
4	cloud	0.33	hum	0.52	dni	0.42	azimuth	0.37
5	zenith	0.30	azimuth	0.45	azimuth	0.42	temp	0.35
6	altitude	0.30	hour	0.44	ghi	0.40	maxMin	0.34
7	hour	0.29	ghi	0.36	temp	0.38	evap	0.21
8	solar	0.27	altitude	0.35	speed	0.38	rain	0.14
9	hum	0.22	zenith	0.35	zenith	0.27	hour	0.12
10	speed	0.14	mon	0.33	altitude	0.27	weather	0.11
11	pressure	0.13	speed	0.24	solar	0.21	solar	0.10
12	direct	0.12	evap	0.22	direct	0.19	hum	0.10
13	rain	0.12	solar	0.21	cloud	0.18	direct	0.09
14	azimuth	0.11	dni	0.21	dew	0.17	pressure	0.08
15	evap	0.11	dhi	0.21	evap	0.16	cloud	0.07
16	weather	0.11	direct	0.17	weather	0.16	altitude	0.06
17	dew	0.08	dew	0.13	dhi	0.15	vis	0.06
18	maxMin	0.07	cloud	0.11	rain	0.13	speed	0.06
19	mon	0.07	pressure	0.08	vis	0.12	zenith	0.05
20	vis	0.06	weather	0.05	mon	0.11	dni	0.05
21	temp	0.05	vis	0.03	pressure	0.09	dhi	0.03
22	wet	0.04	rain	0.01	wet	0.06	ghi	0.02

Equation 6 shows the one-variable models from the Table 8 for step 1.

Equation 6: The best fitting one-variable models and their mean adj-R^2

$dni_a = f(dni_b)$ mean adj-$R^2 = 0.81$ (a)
$temp_a = f(temp_b)$ mean adj-$R^2 = 0.93$ (b)
$hum_a = f(hum_b)$ mean adj-$R^2 = 0.81$ (c)
$pres_a = f(wet_b)$ mean adj-$R^2 = 0.57$ (d)

Where

a = Alan's (2013) dataset
b = Bureau of Meteorology's dataset
temp = Dry bulb temperature (°C)
hum = Relative humidity (%)
wet = wet bulb temperature (°C)
pres = Atmospheric Pressure (mbar)

The results from the one-variable model selection are used in step 2.

4.2.2 Step 2 – Selecting the two-variable models

The unshaded rows in Table 9 show the mean AIC and adj-R^2 values for the four drivers against the 21 two-variable models input or explanatory variables ranked by ascending order of mean AIC. The first row in the Table 9 is shaded grey to indicate this is the one-variable model from step one above. Equation 7 shows the best two-variable model, with 'best' defined as the model with the lowest AIC.

Equation 7: Best fitting two-variable models and their mean adj- R^2

$dni_a = f(dni_b, month)$	mean adj-R^2 = 0.83	(a)
$temp_a = f(temp_b, ghi_b)$	mean adj-R^2 = 0.95	(b)
$hum_a = f(hum_b, dni_b)$	mean adj-R^2 = 0.86	(c)
$pres_a = f(wet_b, month)$	mean adj-R^2 = 0.67	(d)

These two-variable models are used in step 3 to find the three variable models as discussed in Section 3.3.3. This process was continued for 11 steps with the results discussed in the next section.

Table 9: Step 2 – Selecting the best two-variable model ranked by mean AIC

(a) DNI

Rank		AIC	adj-R^2
	DNI	44634	0.81
1	mon	44346	0.83
2	ghi	44375	0.82
3	dhi	44411	0.82
4	maxMin	44413	0.82
5	zenith	44426	0.82
6	hour	44429	0.82
7	azimuth	44431	0.82
8	cloud	44435	0.82
9	altitude	44454	0.82
10	solar	44525	0.81
11	pressure	44550	0.81
12	hum	44570	0.82
13	temp	44590	0.82
14	evap	44597	0.81
15	wet	44606	0.81
16	rain	44617	0.81
17	speed	44620	0.81
18	weather	44636	0.81
19	dew	44637	0.81
20	direct	44657	0.81
21	vis	44704	0.81

(b) Temperature

	AIC	adj-R^2
TEMP	2072	0.93
ghi	400	0.95
hour	779	0.95
altitude	811	0.95
zenith	853	0.95
azimuth	1039	0.95
dni	1184	0.94
maxMin	1323	0.94
dhi	1536	0.94
weather	1547	0.94
pressure	1601	0.94
rain	1745	0.94
dew	1822	0.93
solar	1840	0.93
cloud	1846	0.93
wet	1851	0.93
hum	1886	0.93
direct	1912	0.93
mon	1921	0.93
speed	1975	0.93
vis	2005	0.93
evap	2102	0.93

(c) Humidity

	AIC	adj-R^2
HUM	17575	0.81
dni	16622	0.86
ghi	16632	0.86
hour	16727	0.86
azimuth	16813	0.85
zenith	16932	0.84
altitude	16980	0.84
cloud	17008	0.84
mon	17019	0.84
pressure	17200	0.83
dew	17239	0.83
dhi	17265	0.83
rain	17283	0.83
solar	17285	0.83
wet	17288	0.83
temp	17296	0.83
weather	17310	0.83
speed	17411	0.82
direct	17442	0.82
evap	17456	0.82
maxMin	17477	0.82
vis	17578	0.82

(d) Pressure

	AIC	adj-R^2
WET	7544	0.57
mon	7012	0.67
dew	7085	0.68
temp	7100	0.67
hum	7107	0.68
maxMin	7123	0.65
hour	7128	0.66
dni	7140	0.61
azimuth	7239	0.66
cloud	7269	0.64
ghi	7274	0.63
evap	7330	0.63
pressure	7351	0.63
rain	7363	0.61
weather	7392	0.61
speed	7439	0.61
solar	7445	0.64
zenith	7448	0.62
vis	7452	0.62
dhi	7477	0.63
altitude	7507	0.62
direct	7604	0.62

4.2.3 Step 11 – Selecting the eleven-variable models

Equation 8 shows the best eleven-variable model from Table 10 that is the model with the lowest AIC. The previous 10 steps provide similar tables but to aid clarity and save space these tables are excluded. Table 10 provides a composite of the previous 10 steps.

Equation 8: Best fitting eleven-variable models

dni_a = f(dni_b, month, ghi_b, dhi_b, $cloud_b$, pressure, vis_b, $maxMin_b$, $direct_b$, $solar_b$, $temp_b$) (a)
mean adj-R^2 = 0.86

$temp_a$ = f($temp_b$, ghi_b,$solar_b$, $maxMIn_b$, $weather_b$, dew_b, vis_b, hour, $evap_b$, altitude, month) (b)
mean adj-R^2 = 0.97

hum_a = f(hum_b, dni_b, hour, month, dew_b, $solar_b$, $weather_b$, $cloud_b$, zenith, vis_b, azimuth) (c)
mean adj-R^2 = 0.93

$pres_a$ = f(wet_b, month, dew_b, $solar_b$, hour, $direct_b$, hum_b, dni_b, $cloud_b$, ghi_b, altitude) (d)
mean adj-R^2 = 0.83

The first ten rows, shaded grey in Table 10, indicate the accumulation of the previous ten steps to find the ten-variable model. The first greyed row shows the one-variable model and its mean AIC and R^2 values. The second greyed row shows the second variable of the two-variable model and its mean AIC and R^2 values. The third greyed row shows the third variable of the three-variable model and its mean AIC and R^2 values and so forth until the tenth row.

The 12 unshaded rows in Table 10 show the eleventh variable of the eleven-variable models and their mean AIC and R^2 values. There are 12 eleven-variable models and they are ranked in ascending order of mean AIC. Equation 8 shows the best models. The next section discusses pruning these models.

Table 10: Step 11 – selecting the best eleven-variable model by mean AIC

(a) DNI

Rank		Mean AIC	Mean adj-R^2
1	dni	44654	0.81
2	mon	44414	0.82
3	ghi	44187	0.83
4	dhi	43892	0.85
5	cloud	43806	0.85
6	pressure	43706	0.85
7	vis	43716	0.85
8	maxMin	43828	0.85
9	direct	43764	0.85
10	solar	43707	0.85
1	temp	43587	0.86
2	speed	43618	0.86
3	hour	43640	0.86
4	dew	43652	0.86
5	rain	43718	0.85
6	wet	43760	0.85
7	altitude	43782	0.85
8	evap	43790	0.85
9	hum	43818	0.85
10	azimuth	43858	0.85
11	zenith	43908	0.85
12	weather	43933	0.85

(b) Temperature

	Mean AIC	Mean adj-R^2
temp	2144	0.93
ghi	433	0.95
solar	152	0.96
maxMin	-237	0.96
weather	-478	0.96
dew	-909	0.97
vis	-786	0.97
hour	-1363	0.97
evap	-1351	0.97
altitude	-1220	0.97
mon	-1766	0.97
dni	-1723	0.97
dhi	-1536	0.97
zenith	-1535	0.97
pressure	-1519	0.97
wet	-1502	0.97
speed	-1498	0.97
rain	-1484	0.97
azimuth	-1438	0.97
direct	-1418	0.97
hum	-1360	0.97
cloud	-1236	0.97

(c) Humidity

	Mean AIC	Mean adj-R^2
hum	17550	0.81
dni	16592	0.86
hour	15825	0.89
mon	14962	0.91
dew	14666	0.92
solar	14619	0.92
weather	14595	0.92
cloud	14324	0.92
zenith	14280	0.92
vis	13762	0.93
azimuth	13795	0.93
temp	13845	0.93
wet	13869	0.93
maxMin	13948	0.93
rain	13975	0.93
dhi	14028	0.93
ghi	14035	0.93
pressure	14050	0.93
altitude	14390	0.92
speed	14404	0.92
direct	14418	0.92
evap	14463	0.92

(d) Pressure

	Mean AIC	Mean adj-R^2
wet	7546	0.57
mon	7096	0.68
dew	6501	0.74
solar	6069	0.79
hour	5817	0.80
direct	5456	0.82
hum	5604	0.81
dhi	5575	0.81
cloud	5324	0.83
ghi	5399	0.82
altitude	5076	0.83
pressure	5085	0.83
evap	5145	0.84
dni	5242	0.82
rain	5255	0.83
maxMin	5313	0.82
vis	5345	0.82
weather	5375	0.83
speed	5378	0.82
azimuth	5438	0.82
temp	5448	0.82
zenith	5771	0.80

4.2.4 Pruning the models using signal to noise ratio

This section discusses pruning the models in Equation 8 using an information signal to noise ratio to arrive at the models in Equation 9.

Equation 9: Best fitting models after considering signal to noise

$$dni_a = f(dni_b, month, ghi_b, dhi_b, cloud_b, pressure) \quad (a)$$
$$\text{mean adj-}R^2 = 0.85$$
$$temp_a = f(temp_b, ghi_b, solar_b, maxMIn_b, weather_b, dew_b) \quad (b)$$
$$\text{mean adj-}R^2 = 0.97$$
$$hum_a = f(hum_b, dni_b, hour, month, dew_b, solar_b, weather_b, cloud_b, zenith, vis_b) \quad (c)$$
$$\text{mean adj-}R^2 = 0.93$$
$$pres_a = f(wet_b, month, dew_b, solar_b, hour, direct_b) \quad (d)$$
$$\text{mean adj-}R^2 = 0.83$$

Examining the mean adj-R^2 values in Table 10 shows that the information content available from adding an extra explanatory variable is nearly exhausted because the adj-R^2 is no longer increasing or increases very little. Additionally, for the shaded section of Table 10, the simulations of the models of various lengths have all been rerun and their mean AIC and R^2 values calculated. These AIC values no longer increase monotonically as was the case during their selection in the previous steps. This indicates that the noise is greater than the information being extracted in the current process. Section 3.2.3 discusses the source of noise or variability in the goodness-of-fit between simulations of neural network.

This lack of monotonicity in the mean AIC values could be addressed by averaging across more simulations. This would improve the stability of the mean AIC value and possible alter the order of the explanatory variables selected. This is an approach taken in (Woodd-Walker, Kingston & Gallienne 2001) who ran 100 simulations to address the variability in simulation results. However any increase in R^2 values is likely to be slight.

Alternatively, the instability of the mean AIC value also provides an indicator of the point at which adding the extra explanatory variables provides such a poor signal to noise ratio that the variable can be ignored. This poor signal to noise ratio can be seen in the greyed rows 6 and 7 in Table 10(a) where the mean AIC increases from pressure to visibility. The mean AIC also increases from visibility to max-min temperature in greyed rows 7 and 8. Furthermore, Table 8(a) shows that the mean adj-R^2 values for the visibility and max-min temperature are 0.06 and 0.07, respectively. Pruning the DNI model at pressure is appropriate.

In Table 10(b), pruning the temperature model at dew point is appropriate because the mean AIC value increases from dew point to visibility, and there is no increase in mean R^2 value.

In Table 10(c), pruning the humidity model at visibility is appropriate because the mean AIC values increase from visibility to azimuth, and there is no increase in mean R^2 value.

The mean adj-R^2 values in Table 8(d) for the explanatory variables for the driver pressure are the poorest of the four drivers. In Table 10(d), pruning the pressure model at *direct* that is wind direction is appropriate because the mean AIC value increases from *direct* to hum and the mean adj-R^2 values in Table 8(d) for the following explanatory variables in the

pressure model are small. Section 3.2.2.2 discusses the misleading aspect of interpolating wind direction and the requirement for a better algorithm.

Equation 9 shows the models from Equation 8 but the number of explanatory variables have been pruned after considering poor signal to noise ratio that is increasing AIC.

4.3 Conclusion

The models for the four drivers of yield have been selected, so the project can proceed to the next stage that is the development of the yield projections.

5 Discussion

5.1 Introduction

The discussion in this preliminary or draft report addresses the research questions but a more comprehensive discussion will be provided in the final report when all the results are available.

5.2 Can modelling the weather with limited datasets produce greater yield predictive power than using the historically more complete datasets from nearby sites?

The preliminary analysis in the literature review established that the climates between Collinsville and the coastal comparisons site differ considerably. This difference calls into question their use as proxies for the climate and yield in Collinsville. Given the climatic differences, Collinsville is expected to have slightly higher yield and the yield in the afternoon will be particularly higher. This comparatively intensive afternoon yield period requires the use of a daily yield profile in addition to average daily yield statistic when comparing yield between Collinsville and the other sites.

The Power and Energy Systems (Shah, Yan & Saha 2014b) group calculating the yield for the three comparison sites found that MacKay and Townsville had gaps in their datasets, which leaves only Rockhampton as a comparisons site.

5.3 Is BoM's DNI satellite dataset adequately adjusted for cloud cover at Collinsville?

Section 2.2.1 discusses discrepancies in Table 3 over cloud cover and the satellite derived daily solar exposure between Collinsville and the comparison sites. This research arose to address these discrepancies. Equation 10 helps address this research question. Equation 10 comprises of data from Table 8, Table 9 and Table 10.

Equation 10: Cloud cover and DNI modelling

(a) $Dni_a = f(dni_b)$ mean adj-$R^2 = 0.81$
(b) $Dni_a = f(cloud_b)$ mean adj-$R^2 = 0.33$
(c) $Dni_a = f(dni_b, cloud_b)$ mean adj-$R^2 = 0.82$
(d) $Dni_a = f(dni_b, mon, ghi_b, dhi_b, cloud_b, pressure)$ mean adj-$R^2 = 0.85$

Where

a = Allen's (2013) dataset
b = BoM's dataset
cloud = cloud cover

Equation 10(a) and Equation 10(b) show the mean adj-R^2 values for the single-variable models: dni_b and $cloud_b$ at 0.81and 0.33, respectively. Equation 10 (c) shows a mean adj-R^2 values of 0.82 for the two-variable model $(dni_b, cloud_b)$. The one percentage point increase in mean adj-R^2 from the one-variable model in Equation 10(a) to the two-variable model in Equation 10(c) indicates that the BoM's DNI data estimation from satellite images adequately incorporates cloud coverage. This comes with the caveat that $cloud_b$ is thrice daily dataset and dni_b is hourly. Equation 10(d) shows the final model selected for the driver DNI. The inclusion of $cloud_b$ after mon, ghi_b and dhi_b indicates that $cloud_b$ plays a very small

part in the model. ghi_b and dhi_b already incorporate cloud coverage in there derivation. This further reduces the role of $cloud_b$ in the final model.

5.4 Given dust and dew effects, is raw satellite data sufficient to model yield?

Section 2.3 discusses the dust and dew effect and introduces the concept of "effective" DNI to ameliorate concern over these effects for the use of terrestrial measurement of DNI for LFR yield calculations. However, satellite data uses image evaluation, so is unadjusted for dust and dew effects but may allow for some dust-in the atmosphere. Equation 10(a) shows the relation between "effective" DNI and satellite DNI represented by dni_a and dni_b, respectively. The mean adj-R^2 = 0.81 indicates that raw satellite data is good approximation to effective DNI but lacks a good fit.

5.5 Does elevation between Collinsville and nearby sites affect yield?

Section 2.2 discusses the lapse rates for temperature and pressure that is the change in temperature or pressure with change in elevation. The temperature and pressure decrease with increased elevation but this relation is far from simple with temperature having three lapse rates, which are used to study cloud dynamics. Using the simple pressure and environmental lapse rates (ELR) to perform a sensitivity analysis on the yield difference between Rockhampton and Collinsville provides an opportunity to recalibrate Rockhampton's yield data for the pressure and temperature difference. A simple application of the ELR would imply a small temperature decrease for Collinsville but examining Table 10 shows that the temperature range for Collinsville is wider relative to the three comparison sites. So, other effects such as the presence or absence of a sea breeze are overwhelming any temperature change induced by elevation and the ELR.

5.6 How does the ENSO affect yield?

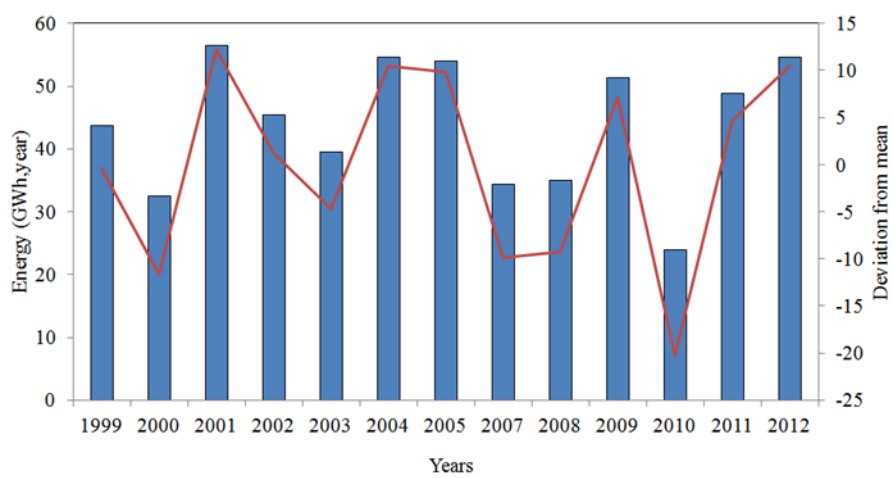
Section 2.4 discusses the ENSO where the *El Niño* phase relative to the *La Niña* phase increases DNI, temperature and pressure and reduces humidity. The overall *El Niño* effect is to increase yield and electricity demand.

5.7 Given the 2007-2012 constraint, will the TMY process provide a "Typical" year over the ENSO cycle?

Section 2.6 discusses the TMY process and the 2007-2012 constraint. The TMY process is expected to average out the ENSO cycle to provide a representative year. However, the 2007-2012 constraint may lack enough data to average out the ENSO cycle. Comparing the yield from the TMY for 2007-2012 with the TMY yield for previous years could ameliorate this concern.

However, Figure 3 show the percentage deviations in yield from the long term mean for 1999-2012 for the comparison site in Rockhampton. Collinsville could be expected to follow a similar pattern. This yield deviation pattern suggests a bias toward La Niña that is lower yield, during 2007-2012 compared to 1999-2006. This implies the report's period of study, 2007-2012, will under report the yield. The follow paragraph investigates the La Niña and El Niño phases to contrast their prevalence during the periods 2007-2012 and 1999-2006 to verify this interpretation of Figure 3.

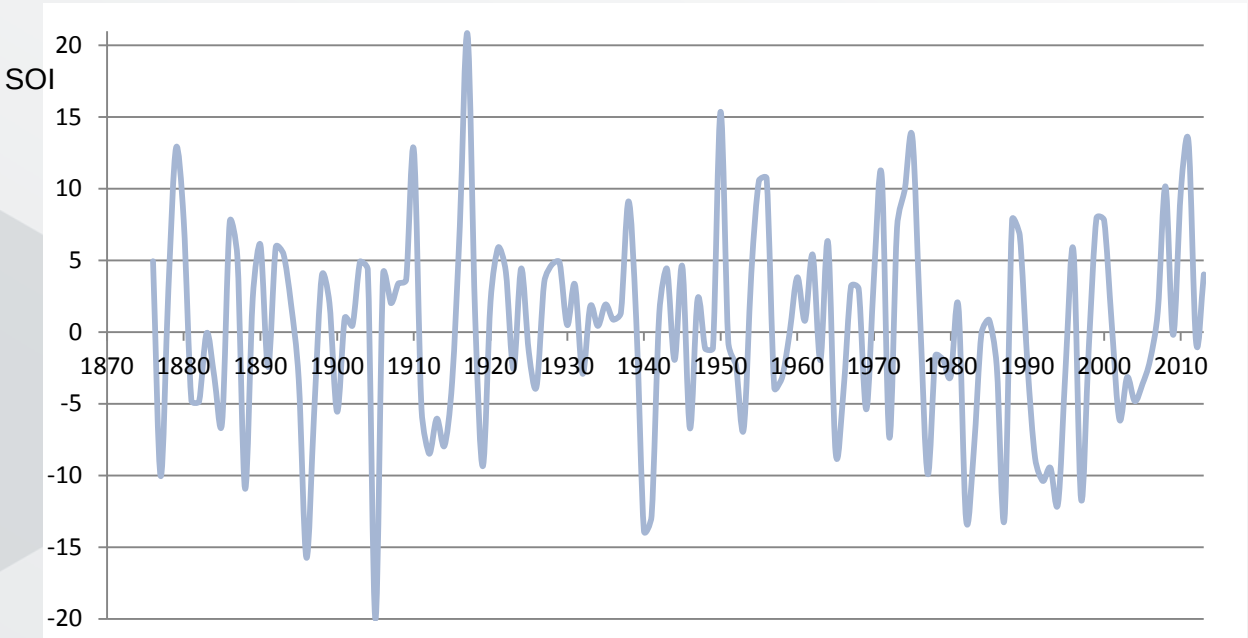
Figure 3 Yearly energy yield and deviations from long term mean values



(Source: Shah, Yan & Saha 2014b)

Figure 4 shows the mean annual SOI for 1875-2013 where positive SOI indicates a La Niña (BoM 2014c) bias and the negative SOI indicates an El Niño (BoM 2014b) bias. The mean annual SOI for 2007-12 is 5.63, which indicates a strong La Niña bias, and for 1999-2006 is -0.4167, which indicates El Niño bias. This supports the interpretation in of Figure 3. However, the mean annual SOI for 1875-1998 is -0.05, so both 1999-2006 and 2007-12 are bias in opposite directions compared to the longer-term average SOI. Hence comparing these periods overstates the La Niña bias of the period understudy. Nevertheless, there is a La Niña bias in the period understudy. This bias will cause an underreporting the yield from the plant.

Figure 4: Mean annual SOI 1875-2013



(Source: BoM 2014d)

5.8 How does climate change affect yield?

Section 2.5 discusses that the most likely effects of climate change are to reduce humidity and increase temperature and DNI but increase DNI only by a tiny amount. The data series of climate change lack projections for pressure. However, this scenario is similar to the *El Niño* phase described above, which indicates an increase in pressure. Therefore, the *El Niño* phase and climate change have similar implications for the LFR at Collinsville that are increasing yield and electricity demand.

Modelling the NEM becomes complex very quickly, so it is essential to focus on the core issue of the feasibility study that is to gain a PPA for the solar thermal project. EEMG's three reports strive to strike a balance by avoiding too many complexities but providing sufficient complexity to address the core issue of the feasibility study. Incorporating climate change into the modelling in the subsequent reports would impose a great deal of complexity over a small effect for the proposed plant. However, a sensitivity analysis on the data in Table 5 would quantify how small the climate change effect is likely to be.

Section 7, further research, suggests options for investigating the effect of climate change on the proposed plant.

5.9 What is the expected frequency of oversupply from the Linear Fresnel Novatec Solar Boiler?

Section 3.4 discusses yield calculation using SAM's default "Linear Fresnel Novatec Solar Boiler" that is modified to reflect the LFR at Collinsville. These modifications indicate the boiler is oversized and is likely to exceed the 30MW limit imposed by AMEO.

The Power Energy Systems Group's (Shah, Yan & Saha 2014b) yield calculation using SAM's modified "Linear Fresnel Novatec Solar Boiler" with Rockhampton's historical complete environmental datasets found that the yield exceeded 30MW. This issue of exceeding the AMEO limit requires consideration of both frequency and size of exceeding the limit to determine whether exceeding the limit is acceptable by AMEO or spillage is required.

.

6 Conclusion

This preliminary or draft report presents the complete literature review, arising research questions and methodology to address the research questions. This draft report also presents the first set of results that are the models for the four drivers of yield and yield estimates for Rockhampton. The discussion in Section 5 analyses the research questions based on the results were possible.

The methodology is in place to address the remaining research questions to provide a more comprehensive discussion in the final report when all results are available. The models for the four drivers of yield have been selected, so the project can proceed to the next stage that is the development of the yield projections.

7 Further research

This section compiles the further research discussed elsewhere in this report.

7.1 Inter-year variability rather than TMY

Section 2.6.1 compares two approaches to yield analysis. The TMY approach that allows sensitivity analysis and the individual year approach that allows inter-year variability analysis to calculate a P90. This report and the subsequent reports use the TMY approach. The inter-year variability approach is suggested for further research.

7.2 Using BoM's Rockhampton one-minute solar dataset

Appendix A provides yield analysis based on hourly data for Rockhampton from US Energy Plus. The BoM provides one-minute solar data for Rockhampton starting from 1996. This data provides the opportunity to assess the yield more accurately.

7.3 Climate change adjusted demand forecasts

Section 2.5 discusses the effect of climate change on yield, using three GCMs models to represent three cases: most like temperature rise, hottest and coolest for the NEM's geographic area. But because the main driver for yield is DNI, this analysis could be improved by selecting GCMs for the three cases: most like change in solar radiance, dullest and brightest for the NEM area. In addition, comparing the GCM selection for Collinsville with the NEM would indicate if any local climate change effect on Collinsville is running counter to the global effect on the NEM. The local effect determines yield but the global effect determines demand and prices.

7.4 The effects of weights in the neural networks on adj-R2 and AIC

Section 3.3.5 discusses the effect of the internal weights within the neural networks to cause an overestimation of adj-R^2 values. This effect was shown to be trivial in the report.

In addition to adj-R^2 values to indicate model fit, this report uses AIC to select between models. The effect of the weights on AIC in this report was shown to be irrelevance because 10 weights are used throughout this report. However, keeping the weights fixed at 10 reduces the possibility of fine-tuning the neural networks.

Acknowledgements

The Australian Government via the Australian Renewable Energy Agency (ARENA) provided financial support for this project. The views expressed herein are not necessarily the views of the Commonwealth, and the Commonwealth does not accept responsibility for any information or advice contained herein.

We thank Janet Gray, Michelle Hall and Kate Newall of the Global Change Institute for their administration and coordination of the project amongst UQ, RATCH-Australia Corporation (RAC) and ARENA.

We thank Dr Greg Allen of Allen Solar for his exceeding quick response to data requests and going beyond the scope originally specified.

We thank the US National Renewable Energy Laboratory (NREL 2012) for providing their Systems Advisor Model (SAM) free of charge and Novatec Solar for providing SAM definition files for their Fresnel Lens Solar Boiler. Additionally, we thank Novatec Solar for their permission to use the photograph on cover of this report.

Solar radiation data derived from satellite imagery processed by the Bureau of Meteorology from the Geostationary Meteorological Satellite and MTSAT series operated by Japan Meteorological Agency and from GOES-9 operated by the National Oceanographic & Atmospheric Administration (NOAA) for the Japan Meteorological Agency.

8 References

Akaike, H 1974, 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716-23.

Allen, G 2013, *Solar thermal data: Collection, data, analysis report for Collinsville, January to May 2013*, Allensolar, Brisbane, Australia.

Bell, WP, Wild, P & Foster, J 2013, 'The transformative effect of unscheduled generation by solar PV and wind generation on net electricity demand', paper presented to 2013 IAEE International Conference, Daegu, Korea, 16-20 June 2013.

—— 2014a, *Collinsville solar thermal project: Energy economics and dispatch forecasting*, University of Queensland, Brisbane, Australia.

—— 2014b, *Collinsville solar thermal project: Yield forecasting*, University of Queensland, Brisbane, Australia.

BoM 2005, *El Niño, La Niña and Australia's Climate*, Bureau of Meteorology, <http://www.bom.gov.au/info/leaflets/nino-nina.pdf>.

—— 2007, *Climate statistics for Australian locations: Definitions for other daily elements*, Bureau of Meteorology, viewed 22 Mar 2014, <www.bom.gov.au/climate/cdo/about/definitionsother.shtml>.

—— 2011, 'Climate Data Online', viewed 14 Nov 2011, <http://www.bom.gov.au/climate/data/>.

—— 2013, 'Australian Hourly Solar Irradiance Gridded Data', viewed 16 Oct 2013, <http://www.bom.gov.au/climate/how/newproducts/IDCJAD0111.shtml>.

—— 2014a, 'Climate statistics for Australian locations', vol. 2014, no. 22 Mar, <http://www.bom.gov.au/>.

—— 2014b, *El Niño - Detailed Australian Analysis*, Bureau of Meterology, viewed 23 Mar 20014, <http://www.bom.gov.au/climate/enso/enlist/>.

—— 2014c, *La Niña – Detailed Australian Analysis*, Bureau of Meterology, viewed 23 Mar 20014, <http://www.bom.gov.au/climate/enso/lnlist/>.

—— 2014d, *Monthly Southern Oscillation Index*, 23 Mar, Bureau of Meterology, <ftp://ftp.bom.gov.au/anon/home/ncc/www/sco/soi/soiplaintext.html>.

Burnham, KP & Anderson, DR 2002, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach (2nd ed.)*, Springer-Verlag.

Cebecauer, T, Skoczek, A, Betak, J & Suri, M 2011, *Site Assessment of Solar Resource: Upington Solar Park, Province Northern Cape, South Africa*, GeoModel Solar, Bratislava, Slovakia.

Clarke, JM & Webb, L 2011, *Meeting to discuss climate futures*, Tailored Project Services, CSIRO Division of Marine and Atmospheric Research, Aspendale, Victoria.

CSIRO 2011, 'OzClim: Exploring climate change scenarios for Australia', viewed 2 Nov 2011, <http://www.csiro.au/ozclim/home.do>.

Deoras, A 2010, 'Electrcity load and price forecasting with Matlab', *MathWorks*, <http://www.mathworks.com.au/webex/recordings/loadforecasting_090810/index.html>.

Foster, J, Bell, WP, Wild, P, Sharma, D, Sandu, S, Froome, C, Wagner, L, Bagia, R & Misra, S 2013, *Analysis of Institutional adaptability to redress electricity infrastructure vulnerability due to climate change*, National Climate Change and Adaptation Foundation, Brisbane, Australia.

Fovell, RG 2010, *Meteorology: An Introduction to the Wonders of the Weather*, The Great Courses, Virginia, USA.

Glaenzel, J 2013, *Advised changes from the SAM default setting for a 'Linear Fresnel Novatec Solar Boiler'*, Novatec Solar, <http://www.novatecsolar.com/>.

Guan, Z, Yu, S & Gurgenci, H 2014, *Collinsville solar thermal project: Solar mirror cleaning requirements*, University of Queensland, Brisbane, Australia.

Hippert, HS, Pedreira, CE & Souza, RC 2001, 'Neural Networks for Short-Term Load Forecasting: A Review and Evaluation', *IEEE Transactions on Power Systems*, vol. 16, pp. 44-55.

Lovegrove, K, Franklin, S & Elliston, B 2013, *Australian Companion Guide to SAM for Concentrating Solar Power*, IT Power (Australia) Pty Limited, ACT Australia.

Marion, W & Urban, K 1995, *User's Manual for TMY2s (Typical Meteorological Years) - Derived from the 1961-1990 National Solar Radiation Data Base*, National Renewable Energy Laboratory, Colorado, USA.

MathWorks 2014a, *Divide Data for Optimal Neural Network Training*, viewed 14 Mar 2014, <http://www.mathworks.com.au/help/nnet/ug/divide-data-for-optimal-neural-network-training.html>.

—— 2014b, *Improve Neural Network Generalization and Avoid Overfitting*, viewed 14 Mar 2014, <http://www.mathworks.com.au/help/nnet/ug/improve-neural-network-generalization-and-avoid-overfitting.html>.

NOAA 2012, *Air Pressure*, National Oceanic and Atmospheric Administration, viewed 24 Feb 2014, <http://www.srh.weather.gov/srh/jetstream/atmos/pressure.htm>.

—— 2014, *Environmental Temperature Lapse Rates*, National Oceanic and Atmospheric Administration, viewed 14 Mar 2014, <www.spc.noaa.gov/exper/soundings/help/lapse.html>.

NREL 2012, 'System Advisor Model (SAM)', <https://sam.nrel.gov>.

Page, CM & Jones, RN 2001, 'OzClim: the development of a climate scenario generator for Australia', in F Ghassemi, P Whetton, R Little & M Littleboy (eds), *Integrating models for natural resources management across disciplines, issues and scales (Part 2)*, MODSIM 2001, International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, Canberra, pp. 667-72.

Reda, I & Andreas, A 2008, *Solar Position Algorithm for Solar Radiation Applications*, National Renewable Energy Laboratory, Colorado, USA, <http://www.nrel.gov/docs/fy08osti/34302.pdf>.

Roy, V 2004, 'Sun position given observer time/location', *Matlab Central*, viewed 10 Mar 2004, <http://www.mathworks.com/matlabcentral/fileexchange/4605-sunposition-m>.

Shah, R, Yan, R & Saha, T 2014a, *Collinsville solar thermal project: Power system assessment*, University of Queensland, Brisbane, Australia.

—— 2014b, *Collinsville solar thermal project: Yield analysis of a linear Fresnel reflector based CSP by long-term historical data*, University of Queensland, Brisbane, Australia.

Singh, R & Gurgenci, H 2014a, *Collinsville solar thermal project: Fossil fuel boiler integration*, University of Queensland, Brisbane, Australia.

—— 2014b, *Collinsville solar thermal project: Optimisation of operational regime*, University of Queensland, Brisbane, Australia.

Stoffel, T, Renné, D, Myers, D, Wilcox, S, Sengupta, M, George, R & Turchi, C 2010, *Concentrated Solar Power: Best practices handbook for the collection and use of solar resources data*, National Renewable Energy Laboratory, Colorado, USA, <http://www.nrel.gov/docs/fy10osti/47465.pdf>.

Wagner, MJ 2012, 'Results and Comparison from the SAM Linear Fresnel Technology Performance Model', in *2012 World Renewable Energy Forum*, Denver, Colorado, USA, May 13–17, 2012.

Wagner, MJ & Zhu, G 2012, 'A Direct-steam Linear Fresnel Performance Model for NREL's System Advisor Model', in *ASME 2012 6th International Conference on Energy Sustainability & 10th Fuel Cell Science, Engineering and Technology Conference*, San Diego, CA, USA.

Wilcox, S & Marion, W 2008, *Users Manual for TMY3 Data Sets*, NREL/TP-581-43156, National Renewable Energy Laboratory.

Woodd-Walker, RS, Kingston, KS & Gallienne, CP 2001, 'Using neural networks to predict surface zooplankton biomass along a 50°N to 50°S transect of the Atlantic', *Journal of plankton research*, vol. 23, no. 8, pp. 875-88.

About the
Global Change Institute

The Global Change Institute at The University of Queensland, Australia, is an independent source of game-changing research, ideas and advice for addressing the challenges of global change. The Global Change Institute advances discovery, creates solutions and advocates responses that meet the challenges presented by climate change, technological innovation and population change.

This technical report is published by the Global Change Institute at The University of Queensland.
A summary paper is also available.
For copies of either publication visit
www.gci.uq.edu.au

T: (+61 7) 3443 3100 / E: gci@uq.edu.au
Global Change Institute Building (Bldg. 20)
Staff House Road
University of Queensland
St Lucia QLD 4072, Australia

www.gci.uq.edu.au