Collinsville solar thermal project: Yield forecasting – Final report

Bell, William Paul and Wild, Phillip and Foster, John

University of Queensland

4 November 2014

COLLINSVILLE SOLAR THERMAL PROJECT

Yield Forecasting
Final Report

THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA

gciQ
Global Change Institute

Prepared for

RATCH-Australia Corporation

RAC
RATCH-Australia Corporation

Chief Investigators

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

Authors

Dr William Paul Bell, p.bell2@uq.edu.au, Energy Economics and Management Group
Dr Phillip Wild, p.wild@uq.edu.au, Energy Economics and Management Group
Professor 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

Copyright

Preface

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) partially 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. We restructured the suggested content for the three reports in the matrix 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 our *'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.   Our 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 projections but its three nearby neighbours 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 Typical Metrological Year (TMY) technique to overcome technology specific TMYs.   We discuss the effect of climate change and the El Niño Southern Oscillation (ENSO) on yield and their implications for a TMY.

2.1   Research questions

Research questions 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:

- Does BoM adequately adjust its DNI satellite dataset 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 cycle affect yield?

- Given the 2007-12 electricity demand data constraint, will the 2007-13 based TMY provide a "Typical" year over the ENSO cycle?
- How does climate change affect yield?
- Is the method to use raw satellite DNI data to calculate yield and retrospectively adjusting the calculated yield with an effective to satellite DNI energy per area ratio suitable?
- How has climate change affected the ENSO cycle?

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. We also discuss the development of the technology specific TMY and sensitivity analysis to address the research questions on climate change and elevation.

4   Results and analysis

In the results section we present the selection process for the four driver models. We also present the effective to satellite DNI ratio, the annual variation in gross yield, the selection of TMMs for the TMY based on monthly yield, the sensitivity analysis results on climate change and elevation, and the frequency of gross yield exceeding 30 MW.

5   Discussion

We analyse the results within a wider context, in particular, we make a comparison with the yield calculations for Rockhampton to address the overarching research question. We find that the modelling of weather at Collinsville using incomplete weather data has higher predictive performance that using the complete weather data at Rockhampton but recommend using the BoM's one-minute solar data to improve the comparative test. Other findings include the requirement to increase the current TMM's selection period 2007-13 to incorporate more of the ENSO cycle. There is less than 0.3% change in gross yield from the plant in the most likely case of climate change but there is a requirement to determine the effect of climate change on electricity demand and the ensuing change in wholesale electricity prices.

6   Conclusion

In this report, we have addressed the key research questions, produced the yield projections for our subsequent report *'Energy economics and dispatch forecasting'* (Bell, Wild & Foster 2014a) and made recommendations for further research.

Contents

Tables

Figures

Equations

Abbreviations

A1FI	Fossil intensive rapid economic growth SRES
Adj-R^2	Adjust R-squared
AIC	Akaike Information Criteria
ANEM	Australian National Electricity Market Model (from EEMG)
ARENA	Australian Renewable Energy Agency
BoM	Australian Bureau of Meteorology
CSIRO	Commonwealth Scientific and Industrial Research Organisation
CSIRO-Mk3.5	GCM from the CSIRO used for the worst case (hottest) in this project
CSP	Concentrated Solar Power
DALR	Dry Adiabatic Temperature Lapse Rate
DHI	Diffused Horizontal Irradiance
DNI	Direct Normal Irradiance
EEMG	Energy Economics and Management Group (from UQ)
ELR	Environmental Lapse Rate
ENSO	El Niño Southern Oscillation
GCI	Global Change Institute
GCM	Global Climate Model
GHI	Global Horizontal Irradiance
ITEE	Information Technology and Electrical Engineering (from UQ)
JAMI	Japanese Advanced Meteorological Imager
kPa	kilopascals (pressure)
kW	kilowatt (power)
LFR	Linear Fresnel Reflector
MIROC3.2	GCM used for the best case (coolest) in this project
MRI	Meteorological Research Institute (Japan)
MRI-CGCM2.3.2	GCM from the MRI used for the most likely case in this project
MTSAT	Multi-Functional Transport Satellite

MW	Megawatt (power)
MWh	Megawatt-hour (energy)
NEM	Australian National Electricity Market
NREL	US National Renewable Energy Laboratory
NOAA	US National Oceanic and Atmospheric Administration
P50	50% of estimates exceed the P50 estimate
P90	90% of the estimates exceed the P90 estimate
PESG	Power and Energy Systems Group (from UQ ITEE)
PPA	Power Purchase Agreement
PV	Photovoltaic
RH	Relative Humidity
RSS	Residual Sum of the Squares
SALR	Saturated Adiabatic Lapse Rate
SAM	System Advisor Model (from NREL)
SMME	School of Mechanical and Mining Engineering (from UQ)
SOI	Southern Oscillation Index
SRES	Special Report on Emission Scenarios
TMM	Typical Metrological Month
TMY	Typical Metrological Year
UQ	University of Queensland

1   Introduction

The primary aim of this report is to produce hourly yield projections of electricity power for the proposed LFR plant at Collinsville, Queensland, Australia 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 our 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 is:

> *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 inform 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 beyond the infant industry stage 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 infant industry more problematic. Therefore, there is a requirement for a larger 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 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, 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 (ENSO) and climate change on the four drivers and yield to scope sensitivity analysis. Section 6 discusses the format and technique "Typical Meteorological Year" (TMY) and its implications for inter-year variation and sensitivity analysis and introduces modifications to the TMY 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 2012) Systems Advisor Model (SAM 2014) 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 (2014) 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 (2014) methodology in more detail. Other environment variables also affect the amount of electricity produced but the main drivers for yield are these four variables. 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 SAM (2014) 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-meteorological means for the period 1981-2010. The four meteorological drivers for yield for the four sites group these means.

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 hydrological 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 meteorological drivers for SAM (2014) while considering two aspects for each driver. First, do the alternative sites provide suitable weather proxies for Collinsville? Second, we inform the methodology in section 3 about the limited weather datasets available at Collinsville to model the four drivers.

2.2.1 Direct normal irradiance

DNI is the first of the four drivers for yield in SAM (2014) 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 December 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. First, 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. Second, 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 site 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, which should ameliorate this concern. A clearness index can measure atmospheric transmittance.

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. Section 2.3.2 discusses the dust effect further.

2.2.1.2 Collinsville DNI data

Table 3 shows the satellite derived solar daily exposure but SAM (2014) 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. We calculate the altitude 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/m^2 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

$$DHI = GHI - DNI \cos(zenith)$$

The BoM (2013) produces hourly grids based on satellite images starting in 1990; the grids consist 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, this report uses grids from images from two satellites: the Japanese Advanced Meteorological Imager (JAMI) and the Multi-Functional Transport Satellite (MTSAT) series operated by the Japan Meteorological Agency. Section 5.2.1 discusses how the frequency of missing hourly grids increases in years prior to 2007 when different satellites took the images.

Table 3 shows the coverage dates of the two satellites. A satellite produces a grid for each hour it is in range but the satellite take time to traverse Australia hence latitude relates to the minutes past the hour that the satellite made the image. 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 SAM (2014). In this report, 'temperature' means, 'dry bulb temperature'. In contrast, we refer to wet bulb and dew point temperatures 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, 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 coastal 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 2.3.1 discusses the dew effect further.

There is a relationship between elevation and temperature but this relationship 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 site 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 wholesale 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. The relatively higher electricity prices in the late afternoon compound this climate issue. Moderating this concern is the proposal to use a gas generator to top-up any short falls by the LFR yield to 30 MW from 8 am to 10 pm.

2.2.2.2 Collinsville temperature data

There are currently three daily temperature measurements taken at 6 am, 9 am, and 3 pm at the BoM's Collinsville Post Office weather station operational since 1939. This BoM coverage is far short of the hourly input required for SAM (2014) but at least BoM takes the three measurements during the daylight hours when the LFR produces yield. 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 in our 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. There lacks atmospheric pressure measurements at the BoM's Collinsville weather station. The following sections discuss alternative indicators for atmospheric pressure such as wind speed and direction.

In addition to the direct relationships just discussed there is the hydrological 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 which 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 SAM (2014). This paragraph provides a brief description of the relationship amongst RH and the three temperatures: dry bulb, wet bulb, and dew point to inform the discussion in this section. RH is the ratio between vapour supply and vapour capacity. The dew point temperature indicates vapour supply because it is the lowest air temperature before reaching saturation, that is, where the current vapour supply remains unchanged. 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 RH at MacKay is the least comparable to Collinsville. The 9 am RH 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 increasing, 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 both 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 single-daily measurements for evaporation, precipitation, and solar exposure. Another consideration is wind direction because 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 Atmospheric pressure

Atmospheric pressure is the last of the four drivers for yield in SAM (2014). As there is an absence of BoM atmospheric data for Collinsville, the use of the ideal gas law becomes invaluable to the following discussion. The ideal gas law in Equation 2 stipulates that pressure, temperature, and density are dependent on one another, meaning that a change in one causes 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 kPa/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. Section 3.6 in the methodology discusses operationalising the sensitivity analysis and Section 4.4.5 presents the results.

2.2.4.2 Collinsville pressure data

The absence of BoM atmospheric data requires consideration of other variables to model pressure. Candidates include those variables in the direct relationships of irradiance causing temperature change, 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. Mostly solar irradiance and to a lesser extent the lunar cycle drive these atmospheric tides. Therefore, both a daily and annual cycle in the pressure is expected. These tides have small oscillations at low elevations becoming larger at higher elevations. There is an extensive literature on atmospheric tides. Nevertheless, 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 (2014) fails to include wind in its calculation of yield to allow for a chill factor. However, building a linear Fresnel technology plant in a site with high winds is unlikely because the plant would be subject to more damage than at low wind speeds. Therefore, the chill factor is ignorable. Consistently, Table 3 shows that there is considerably lower wind speed in Collinsville than in the three comparison sites because the three comparison sites are subject to the sea breeze cycle and Collinsville is sheltered inland at higher elevation and within a valley.

The higher wind speed at Collinsville's three coastal neighbours makes both SAM (2014) 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 (2014) calculation of yield does make the yield calculated from these comparison sites more comparable with the yield from the Collinsville site.

We disregard wind speed observed 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 thrice-daily 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 SAM (2014) 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 compared to satellite DNI and ameliorates concerns about the discrepancy between terrestrial 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 produce electricity, whether by the LFR or the instrument measuring DNI. The dew effect is considerable in places with clear nights and 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, we can ignore the dew-effect because the dew–effect affects both the terrestrial DNI measuring instrument and the LFR plant. We assume the dew-effect is the same for both. Therefore, measured DNI is the "effective" DNI.

The automatic adjustment for the dew effect on terrestrial measuring instruments to read effective DNI is absent in satellite data. Therefore, the dew effect makes the unmodified use of satellite DNI data questionable, particularly at Collinsville.

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, factors affecting this assumption about constant dust levels include a change in wind patterns or coal-mining intensity or coal-mining methodology. The El Niño Southern Oscillation (ENSO) or climate change can affect both wind speed and direction. However, the implication of ENSO and climate change for the dust-effect is too complex to analyse and probably slight. Therefore, we ignore ENSO and climate change implications for the dust-effect. Nevertheless, Section 2.4 discusses ENSO and Section 2.5 discusses climate change in relation to the four drivers.

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

Similarly, "dust-on" Allen's (2013) terrestrial 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.

The same self-compensating reasoning between measuring instrument and the LFR for the dew-effect applies to the dust-effect because both the measuring instrument for DNI and the LFR are subject to dust-effects. However, we acknowledge the potential for unequal

cleaning regimes between the terrestrial measuring instruments and the LRF could invalidate this assumption. Section 7.6 in further research recommends collaboration between Allen (2013) and Guan, Yu and Gurgenci (2014) to ensure equivalent cleaning regimes.

The question arises about whether it is possible to use raw satellite DNI data to calculate yield and then retrospectively adjust the yield by an effective to satellite DNI ratio.

In summary, we assume that the terrestrially measured DNI is the "effective" DNI for the LFR and both the dust-effect and dew-effect make the unmodified use of satellite DNI data questionable. We also seek to test the arising research question: "Is the method to use raw satellite DNI data to calculate yield and retrospectively adjusting the calculated yield with an effective to satellite DNI energy per area ratio suitable?"

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 cycle is irregular and can span more than a year. Therefore, there is an expectation of many ENSO cycles during the lifetime of the proposed plant at Collinsville.

This paragraph presents a brief description of the ENSO to inform the remainder of the report. BoM (2005) discusses ENSO within a worldwide context. In contrast, this section discusses ENSO cycle implications for Collinsville, Queensland. The ENSO spans the Pacific and consists of two main phases: the La Niña and El Niño phases. We consider La Niña the normal weather phase and El Niño the abnormal weather phase within the ENSO cycle.

In La Niña, the warmer waters off Queensland and cooler waters of Central America create an atmospheric convection current between Central America and Queensland causing the trade winds to blow from South America toward Queensland called the Walker Circulation. 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

Southern Oscillation Index (SOI) 1994 - 2007
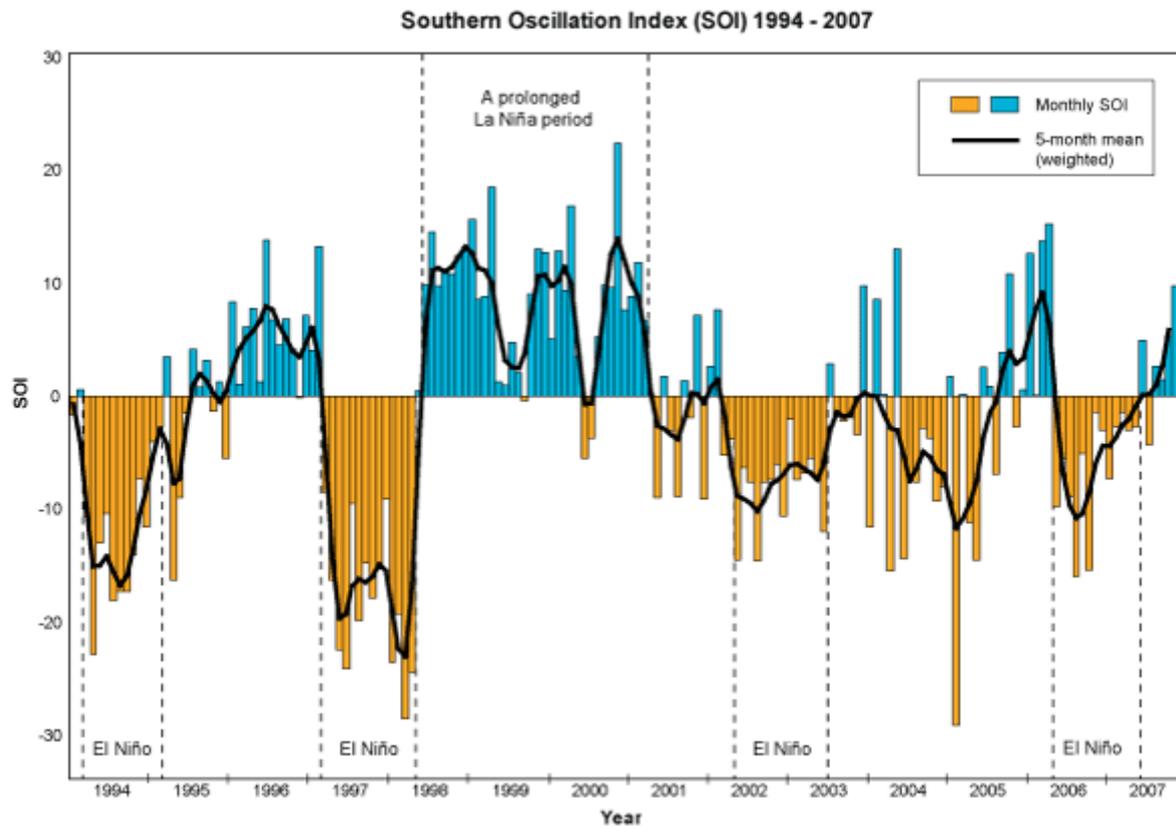
(Source: BoM 2005)

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 the 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 daytime temperatures and lower night-time temperatures because the reduction in moisture ameliorates its moderating effects.

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 more 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. Section 2.6.3 discusses ENSO implications for selecting a TMY.

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,

climate change produces a gradual change in the long-term means of the four drivers to affect the plant's yield permanently over its expected lifetime.

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 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 magnitude of the projected mean temperature changes 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 because 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 the coolest case, the changes may slightly decrease yield. Section 3.5 and Section 4.4.4 discuss the methodology and results, respectively, for the sensitivity analysis to provide estimates that are more exact.

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 meteorological conditions and calculating the associated wholesale spot price and dispatch. Our ensuing 'Energy economics and dispatch forecasting report' (Bell, Wild & Foster 2014a) uses the TMY yield projections from this report to help forecast wholesale prices and dispatch.

2.6.1　TMY as 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 versions 2 and 3. TMY is both a format and a technique. SAM (2014) can use both TMY3 and TMY2 formats. This report uses the TMY3 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 calculated 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 renewable energy generation sector now uses the TMY technique, which required extension of the technique for use in the sector. The TMY technique involves finding the 12 most typical meteorological months (TMMs) from a range of years. The existing TMY technique requires weighting the meteorological variables of interest according to their importance to yield or heating requirements. This weighted average helps 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 meteorological 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) to ensure technologically appropriate implicit weights for the four drivers.

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 our subsequent report (Bell, Wild & Foster 2014a).

2.6.2 TMY implications for demand and supply in the NEM

The dependence of electricity demand on meteorological conditions is a long established relationship but the production of electricity in a predominantly fossil fuel generation fleet is relatively independent of meteorological conditions. 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, meteorological conditions now have an even more dramatic effect on wholesale market 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 introduction of solar PV largely explains the increasing midday depression in net demand. Further solar PV installations without battery storage will exacerbate this reduction in net midday demand. In contrast, the introduction of solar water heaters replacing electrical water heaters reduces demand in the early hours of the morning because electrical water heaters generally use off-peak power that is available in the early hours of the morning. This transformation of the net demand curve requires a consistent application of the TMMs calculated for Collinsville's LFR across the NEM to determine generation mix and net demand to calculate realistic wholesale spot prices. We discuss in more detail the implications for net demand, wholesale prices and dispatch in our ensuing 'Energy economics and dispatch forecasting' report Bell, Wild and Foster (2014a).

Figure 2: NEM's net average demand for 2007 to 2011
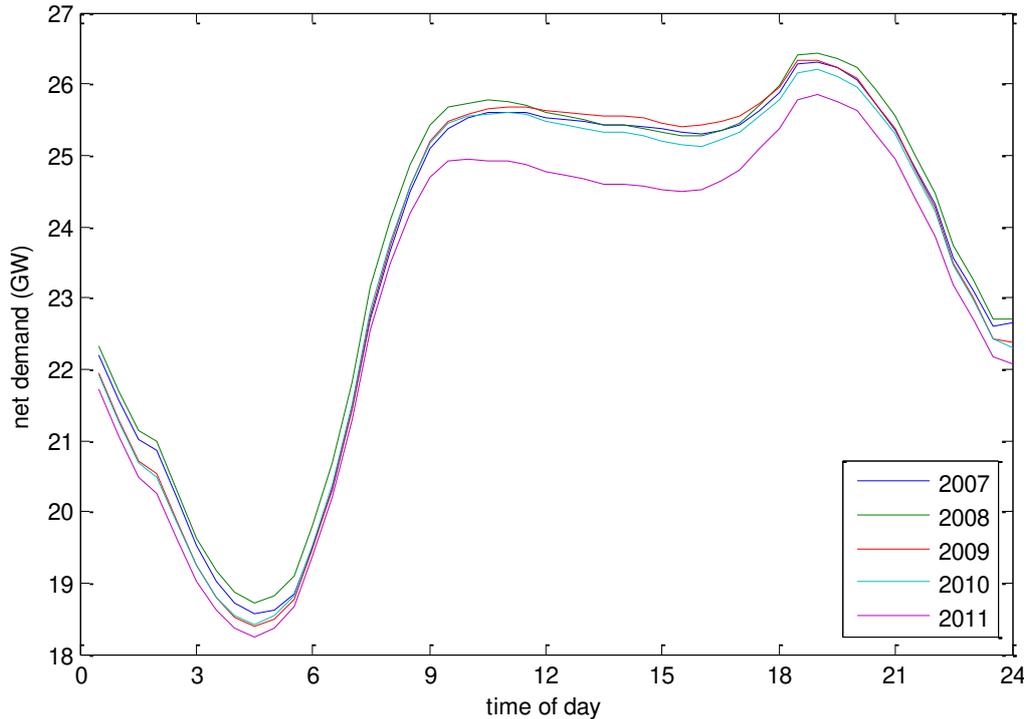
(Source: Bell, Wild & Foster 2013)

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.

How has climate change affected the ENSO cycle? If climate change produces a La Niña or El Niño bias, then this bias restricts the use of previous years to average out the ENSO.

Sections 5.6 and 5.7 discuss a comparative analysis of yield from years 2007-2013 with earlier years to investigate this concern. Section 7.7 in further research recommends extending the TMY process to include earlier years.

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 yield proxies for 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 coastal 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 sites in the mid-afternoon. In Queensland, the wholesale price of electricity is generally higher in the late afternoon. Therefore, Collinsville's climate relatively engenders a shift in LFR yield production to become 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 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 environmental variables. The methodology further develops these two items.

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 (2014) 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.

2.7.1 Research questions

The research questions arising from the literature review.

The overarching research question is:

> *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:

- Does BoM's DNI satellite dataset adequately adjust 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?
- Is the method to use raw satellite DNI data to calculate yield and retrospectively adjusting the calculated yield with an effective to satellite DNI energy per area ratio suitable?

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 2014) 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 (2014) and the range of years, 2007-2012, for our subsequent report (Bell, Wild & Foster 2014a). However, as discussed in the literature review, the dew and dust effects and ambiguity over cloud cover makes the use of this DNI dataset questionable without modification for the aforementioned effects.

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 (2014). In addition, the weather station lacks any datasets for the fourth driver, pressure.

We develop models of the four drivers to overcome the inadequacies in the BoM datasets and satisfy the hourly requirement for SAM (2014) 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 (2014) 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.

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, are selected to model the four drivers, in a procedure known as principle component analysis. For instance, three variables temperature, wind direction, and month could be sufficient to model pressure.

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. Therefore, 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, a simpler two-variable model (temperature and wind direction) or a more complex three-variable model (temperature, wind direction, and month) could model pressure. 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 independent, normally distributed errors with a mean of zero. 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, it is suitable to ignore the constant *c.*

Equation 3: Akaike Information Criteria

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

Where

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

We use the selected models to produce projections of the four drivers for the years 2007-13, which SAM (2014) uses to calculate yield for 2007-13. 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). Section 4.4.3 presents the results from selecting the 12 typical meteorological months (TMM) from the years 2007-12. Section 4.4.1 helps validate the four drivers models for the year 2013 by comparing the yields calculated from driver models with the yields calculated from Allen (2013).

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 comparison site at Rockhampton.

3.2 Preparing the data

This section discusses the preparation of the datasets for use in this report. The outline below shows a hierarchy of the datasets and of 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 the Collinsville site for the four drivers of yield. Allen (2013) converts the one-minute data into an hourly form specifically to meet the requirements of SAM (2014). SAM's (2014) 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 variables determines their preparation, so the following discussion groups the datasets or variables by frequency. We interpolated missing values using the average of the previous and following day. Similarly, we interpolated missing values in the thrice or six times daily measurements using the average of 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.

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

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

We assume that minimum temperatures occur in the early hours of the morning and maximum temperatures occur in the remaining hours of the day. Therefore, we assign the same minimum dry bulb temperature that day for the hours 00:00 to 09:00 and the same maximum dry bulb temperature for the hours 10:00 to 23:00 for the previous day.

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

We simply assigned solar exposure to every hour of the day. We calibrated the model using data from only the daylight hours. Therefore, assigning solar exposure during the night is a non-issue.

3.2.2.2 Thrice daily data

BoM takes thrice-daily readings at 6 am, 9 am, and 3 pm for the following variables.

- o Temperature
    - ▪ Dry bulb
    - ▪ Wet bulb
    - ▪ Dew point
- o Relative humidity
- o Wind
    - ▪ Speed
    - ▪ Direction
- o Cloud cover
- o Visibility
- o 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 environmental 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. We simply interpolated the thrice-daily readings with the following exceptions.

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

We recommend improving the interpolation for wind direction for two reasons. First, 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. Second, 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 past and present weather recordings taken thrice daily in effect give six daily readings. We simply assigned the codes for the present times to the respective hour of the day and the codes for the past times to the previous intervening hours. For example, we assigned the code for the past weather reading taken at 9 am 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. We recommend developing a more sophisticated algorithm to handle the missing readings.

Nevertheless, two factors ameliorate any concern over the averaging of missing codes. First, the weather phenomena are a composite of existing variables such as visibility, relative humidity, precipitation, temperature, cloud cover, and evaporation. Therefore, we have already modelled many of the components of the codes in Table 6. Second, the same code exists on either side of the missing code in many instances. Therefore, the frequency of interpolation between differing codes is low.

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. We adjust the BoM's DNI and GHI readings for the minutes past the hour then interpolate 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

We calculate the astrological angles for every hour of the year for Collinsville, using the algorithm described in section 2.2.1.2.

We use Equation 1 to calculate DHI using the original DNI and GHI datasets discussed in Section 3.2.3. We then adjust DHI for the satellite latitude / minute past the hour deviation as described in Section 3.2.3.

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) to capture the atmospheric tide as discussed in section 2.2.4.

Lastly, we assume that the hour of the day and month of the year capture any other cyclical variation.

3.3 Selecting the best model for the four drivers

As discussed in section 3.1, we use AIC 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 the 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. Therefore, an R^2 value closer to one denotes a better fit as the SS_{res} approaches 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, we use the best fitting explanatory variable for each of the four drivers 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

In the third step, we select the two-variable model with the lowest AIC and the remaining 20 explanatory variable 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, we have exhausted the information value in the remaining explanatory variables 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

In this report, all the neural networks optimised to provide model fit have 10 internal weights. These weights in effect add extra variables to the models by increasing the value of *k* in Equation 3. We can ignore the effect of these weights on AIC because the weights add the same constant to each AIC and constants can be ignored as discussed in Section 3.1.

However, we cannot ignore the weight effect in Equation 5 for adjusted-R^2 (adj-R^2). 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 (n = 4380), gives the following results of *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 one. Ameliorating this effect even further is that the degrees of freedom of the 10 weights require only adding 9 weights to the term *k*. Furthermore, only a fraction of each weight may require reflecting in the term *k*. Section 7.5 in further research discusses this weight issue but for this report, we can safely ignore this issue.

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

In the above steps, we use the mean adj-R^2 or AIC of 10 simulations because the goodness-of-fit of each simulation of a neural network 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 testing (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 providing a scientific foundation for the modelling. We ran the neural network over a number of simulations and averaged the adj-R^2 or AIC values to help improve the veracity of the results. The veracity derives from randomly assigned data into the training, validation and testing sets for each simulation.

3.4 *Calculating yield with the Systems Advisor Model from four drivers*

SAM (2014) calculates the hourly yield for LFR given hourly values for the four drivers in TMY format (Wagner 2012; Wagner & Zhu 2012). We use SAM version 2014.1.14 for this report. The company Novatec Solar will provide the LFR technology for Collinsville. SAM (2014) 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	Field aperture:	174,624 m^2	862,848
	Number of modules in boiler section:	11	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:	120 bar	90
Plant Design	Design turbine gross output:	30.07 MWe	49.998
	Rated cycle efficiency	0.407	0.3941

(Source: Glaenzel 2013)

The field aperture size results from Novatec's techno-economical optimization process for the plant where various dimensions are tested until reaching a minimum levelised cost of electricity. This process considers the physical limitations of the propose site. For instance, reducing the number of modules in the boiler section from 12 to 11 allows the Novatec boiler to fit on the land space available. Additionally, the collector azimuth angle *-10°,* meaning a 10° inclination to the west, allows the LFR plant to maximise the collection field within the land space available.

The increase in the default 'design turbine gross output' is consistent with the increase in 'design turbine gross output'. This implies the plant will exceed the 30 MW AEMO imposed dispatch limit under ideal climatic conditions, which may involve some spillage of excess supply but the amount is uncertain. Therefore, we add the supplementary research question:

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

Section 4.4.6 presents the results from analysing the frequency of oversupply.

3.5 What is the effect of climate change on plant yield?

A sensitivity analysis provides the methodology to determine the effect of climate change on yield. Section 2.5 in the literature review discusses the selection of three Global Climate Models (GCM) for the coolest, most likely and hottest cases. Table 5 provides the expected change in solar radiation, mean temperature, and relative humidity from 1990 to 2040 for the three GCM's. The GCMs lack atmospheric pressure information. Therefore, we assume no change in atmospheric pressure from 1990 to 2040.

In Table 8, we derive the representative values for 1990 for temperature and relative humidity from BoM (2014b) climate statistics for the Collinsville Post Office for the periods 1971-2000 and 1981-2010 by averaging the 9 am and 3 pm values. The averaging between the 9 am and 3 pm values provides representative dry temperature and relative humidity for when the plant is operating. The averaging between the periods 1971-2000 and 1981-2010

weights the averages in the period 1981-2000 to more heavily reflect conditions a decade either side of 1990 the baseline year for the GCMs.

Table 8: Collinsville average temperature and humidly for 1971-2000 and 1981-2010

Driver	1971-2000		1981-2010		mean
	9 am	3 pm	9 am	3 pm	
Dry temperature (°C)	23.1	28.8	23.3	29.3	26.1
Relative humidity (%)	67	44	66	43	55

(Source: BoM 2014b)

BoM (2014b) also lacks data on atmospheric pressure as do the three comparison sites at Rockhampton, Townsville and MacKay. Therefore, we use the average atmospheric pressure supplied by Allen (2013) for the period 12 December 2012 to 11 February 2014 that is 989.5 hPa or 98.95 kPa. Collinsville's average atmospheric pressure is slightly less than the pressure defined in both the 'standard temperature pressure' and 'standard ambient temperature and pressure' that use 100 kPa (1 bar). We expect a lower pressure given Collinsville's 197 m altitude and both the 'standard temperature pressure' and 'standard ambient temperature and pressure' definitions assume the atmospheric pressure at sea level.

The BoM (2013) states that typical values for DNI are up to around 1000 W/m^2. We calculate an effective to satellite DNI ratio of 0.767 in Section 4.3.1. Therefore, we use a DNI value of 767 W/m^2 for 1990 for the sensitivity analysis.

Section 4.4.4 presents the yield sensitivity analysis results using the values derived in this section.

3.6 Does elevation between Collinsville and nearby sites affect yield?

A sensitivity analysis provides the methodology to determine the effect of elevation on yield. Section 2.2.4.1 in the literature review discusses the requirement for a sensitivity analysis on the effect of altitude between Collinsville and the nearby sites of which only Rockhampton presents sufficiently complete data for analysis.

In Table 9, we derive the representative values for 1990 for temperature and relative humidity from BoM (2014b) climate statistics for Rockhampton Aero for the periods 1971-2000 and 1981-2010 by averaging the 9 am and 3 pm values. This averaging technique parallels that for Collinsville in Table 8. Therefore, the Collinsville values in Table 8 can also act as the baseline in this research question.

Table 9: Rockhampton average temperature and humidly for 1971-2000 and 1981-2010

Driver	1971-2000		1981-2010		mean
	9 am	3 pm	9 am	3 pm	
Dry temperature (°C)	22.3	27.2	22.7	27.4	24.9
Relative humidity (%)	69	48	67	46	57.5

(Source: BoM 2014b)

We also assume the same DNI value of 767 W/m^2 derived in Section 3.5 for Collinsville. For atmospheric pressure, we assume a value of 1 bar or 1000 hPa because Rockhampton is much closer to sea level than Collinsville. Section 3.5 also discusses the selection of a

value for atmospheric pressure.   Section 4.4.5 presents the results of the elevation sensitivity analysis.

3.7 Is the method to use raw satellite DNI data to calculate yield and retrospectively adjusting the calculated yield with an effective to satellite DNI energy per area ratio suitable?

A sensitivity analysis also provides the methodology to determine the suitability of the method to use raw satellite data to calculate yield and retrospectively adjusting the yield with an effective to satellite DNI energy per area ratio.   Section 2.3 in the literature review discusses the requirement for a sensitivity analysis to test the suitability of the suggested method.

In the sensitivity analysis, we assume constant temperature, humidity, pressure and effective to satellite DNI energy per area ratio as 26.1°C, 55%, 989.5 hPa and 0.767 as given in Section 3.5. Table 10 shows the change in DNI values for use in the sensitivity analysis where the raw satellite values represent high, medium, and low DNI values and the adjusted satellite values are the raw satellite values factored by the effective to satellite DNI energy per area ratio.

Table 10: DNI values to test suitability of retrospectively adjust yield based on raw satellite DNI data

DNI (W/m2)	high	medium	low
raw satellite	1,000	667	333
adjusted satellite	767	511	256

Section 4.4.7 presents the results of conducting the sensitivity analysis.

3.8 How has climate change affected the ENSO cycle?

Section 2.6.3 discusses this research question. We simply compare the mean of the SOI for the period 1876-1944 with 1945-2013 to test for an increase in a La Niña or El Niño bias.

3.9 Conclusion

This section, building on the literature review, has discussed the methodologies that are 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?*

Now 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 2014). We calibrate the model using the dataset from Allen (2013) for the year 2013. Section 3 calculates annual DNI energy per area to provide an effective to satellite DNI ratio and a benchmark for the LFR yield against the reference year 2013. Section 4 calculates the yield, compares the model yield for 2013 against Allen's (2013) data, and benchmarks the yield for earlier years against the annual DNI energy per area. Section 4 also presents the results from two sensitivity analyses on climate change and elevation.

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 (2014) 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 11 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. We use the mean adj-R^2 of 10 simulations because the results from each simulation of a neural network can differ slightly. Section 3.3.6 discusses the importance of using more than one simulation.

The selection of the first explanatory variable in the first row of Table 11 for DNI, temperature, and relative humidity is unsurprising. However, 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 11 (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 11: Step 1 - selecting the one-variable models for the four drivers using R^2

Rank	(a) DNI Explanatory Variables	mean adj-R^2	(b) Temperature	mean adj-R^2	(c) Humidity	mean adj-R^2	(d) Pressure	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 11 for step 1.

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

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

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)

We use the results from the one-variable model selection in step 2.

4.2.2 Step 2 – Selecting the two-variable models

The unshaded rows in Table 12 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 12 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

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

We use these two-variable models in step 3 to find the three variable models as discussed in Section 3.3.3. We continued this process for 11 steps with the results discussed in the next section.

Table 12: Step 2 – Selecting the best two-variable model ranked by mean AIC

Rank	(a) DNI	AIC	adj-R^2	(b) Temperature	AIC	adj-R^2	(c) Humidity	AIC	adj-R^2	(d) Pressure	AIC	adj-R^2
	DNI	44634	0.81	TEMP	2072	0.93	HUM	17575	0.81	WET	7544	0.57
1	mon	44346	0.83	ghi	400	0.95	dni	16622	0.86	mon	7012	0.67
2	ghi	44375	0.82	hour	779	0.95	ghi	16632	0.86	dew	7085	0.68
3	dhi	44411	0.82	altitude	811	0.95	hour	16727	0.86	temp	7100	0.67
4	maxMin	44413	0.82	zenith	853	0.95	azimuth	16813	0.85	hum	7107	0.68
5	zenith	44426	0.82	azimuth	1039	0.95	zenith	16932	0.84	maxMin	7123	0.65
6	hour	44429	0.82	dni	1184	0.94	altitude	16980	0.84	hour	7128	0.66
7	azimuth	44431	0.82	maxMin	1323	0.94	cloud	17008	0.84	dni	7140	0.61
8	cloud	44435	0.82	dhi	1536	0.94	mon	17019	0.84	azimuth	7239	0.66
9	altitude	44454	0.82	weather	1547	0.94	pressure	17200	0.83	cloud	7269	0.64
10	solar	44525	0.81	pressure	1601	0.94	dew	17239	0.83	ghi	7274	0.63
11	pressure	44550	0.81	rain	1745	0.94	dhi	17265	0.83	evap	7330	0.63
12	hum	44570	0.82	dew	1822	0.93	rain	17283	0.83	pressure	7351	0.63
13	temp	44590	0.82	solar	1840	0.93	solar	17285	0.83	rain	7363	0.61
14	evap	44597	0.81	cloud	1846	0.93	wet	17288	0.83	weather	7392	0.61
15	wet	44606	0.81	wet	1851	0.93	temp	17296	0.83	speed	7439	0.61
16	rain	44617	0.81	hum	1886	0.93	weather	17310	0.83	solar	7445	0.64
17	speed	44620	0.81	direct	1912	0.93	speed	17411	0.82	zenith	7448	0.62
18	weather	44636	0.81	mon	1921	0.93	direct	17442	0.82	vis	7452	0.62
19	dew	44637	0.81	speed	1975	0.93	evap	17456	0.82	dhi	7477	0.63
20	direct	44657	0.81	vis	2005	0.93	maxMin	17477	0.82	altitude	7507	0.62
21	vis	44704	0.81	evap	2102	0.93	vis	17578	0.82	direct	7604	0.62

4.2.3 Step 11 – Selecting the eleven-variable models

Equation 8 shows the best eleven-variable model from Table 13, that is, the model with the lowest AIC. The previous 10 steps provide similar tables but we exclude these tables to aid clarity and to save space. Table 13 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)

$$\text{mean adj-}R^2 = 0.86$$

$temp_a$ = f($temp_b$, ghi_b, $solar_b$, $maxMIn_b$, $weather_b$, dew_b, vis_b, hour, $evap_b$, altitude, month) (b)

$$\text{mean adj-}R^2 = 0.97$$

hum_a = f(hum_b, dni_b, hour, month, dew_b, $solar_b$, $weather_b$, $cloud_b$, zenith, vis_b, azimuth)   (c)

$$\text{mean adj-}R^2 = 0.93$$

$pres_a$ = f(wet_b, month, dew_b, $solar_b$, hour, $direct_b$, hum_b, dni_b, $cloud_b$, ghi_b, altitude)   (d)

$$\text{mean adj-}R^2 = 0.83$$

The first ten rows, shaded in grey in Table 13, indicate the accumulation of the previous ten steps to find the ten-variable models. 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 adj-R^2 values. The third greyed row shows the third variable of the three-variable model and its mean AIC and adj-R^2 values and so forth until the tenth row.

The 12 unshaded rows in Table 13 show the eleventh variable of the eleven-variable models and their mean AIC and adj-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 13: Step 11 – selecting the best eleven-variable model by mean AIC

Rank	(a) DNI	Mean AIC	Mean adj-R^2	(b) Temperature	Mean AIC	Mean adj-R^2	(c) Humidity	Mean AIC	Mean adj-R^2	(d) Pressure	Mean AIC	Mean adj-R^2
1	dni	44654	0.81	temp	2144	0.93	hum	17550	0.81	wet	7546	0.57
2	mon	44414	0.82	ghi	433	0.95	dni	16592	0.86	mon	7096	0.68
3	ghi	44187	0.83	solar	152	0.96	hour	15825	0.89	dew	6501	0.74
4	dhi	43892	0.85	maxMin	-237	0.96	mon	14962	0.91	solar	6069	0.79
5	cloud	43806	0.85	weather	-478	0.96	dew	14666	0.92	hour	5817	0.80
6	pressure	43706	0.85	dew	-909	0.97	solar	14619	0.92	direct	5456	0.82
7	vis	43716	0.85	vis	-786	0.97	weather	14595	0.92	hum	5604	0.81
8	maxMin	43828	0.85	hour	-1363	0.97	cloud	14324	0.92	dhi	5575	0.81
9	direct	43764	0.85	evap	-1351	0.97	zenith	14280	0.92	cloud	5324	0.83
10	solar	43707	0.85	altitude	-1220	0.97	vis	13762	0.93	ghi	5399	0.82
1	temp	43587	0.86	mon	-1766	0.97	azimuth	13795	0.93	altitude	5076	0.83
2	speed	43618	0.86	dni	-1723	0.97	temp	13845	0.93	pressure	5085	0.83
3	hour	43640	0.86	dhi	-1536	0.97	wet	13869	0.93	evap	5145	0.84
4	dew	43652	0.86	zenith	-1535	0.97	maxMin	13948	0.93	dni	5242	0.82
5	rain	43718	0.85	pressure	-1519	0.97	rain	13975	0.93	rain	5255	0.83
6	wet	43760	0.85	wet	-1502	0.97	dhi	14028	0.93	maxMin	5313	0.82
7	altitude	43782	0.85	speed	-1498	0.97	ghi	14035	0.93	vis	5345	0.82
8	evap	43790	0.85	rain	-1484	0.97	pressure	14050	0.93	weather	5375	0.83
9	hum	43818	0.85	azimuth	-1438	0.97	altitude	14390	0.92	speed	5378	0.82
10	azimuth	43858	0.85	direct	-1418	0.97	speed	14404	0.92	azimuth	5438	0.82
11	zenith	43908	0.85	hum	-1360	0.97	direct	14418	0.92	temp	5448	0.82
12	weather	43933	0.85	cloud	-1236	0.97	evap	14463	0.92	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

$$dni_a = f(dni_b, month, ghi_b, dhi_b, cloud_b, pressure) \qquad \text{(a)}$$
$$\text{mean adj-}R^2 = 0.85$$
$$temp_a = f(temp_b, ghi_b, solar_b, maxMIn_b, weather_b, dew_b) \qquad \text{(b)}$$
$$\text{mean adj-}R^2 = 0.97$$
$$hum_a = f(hum_b, dni_b, hour, month, dew_b, solar_b, weather_b, cloud_b, zenith, vis_b) \qquad \text{(c)}$$
$$\text{mean adj-}R^2 = 0.93$$
$$pres_a = f(wet_b, month, dew_b, solar_b, hour, direct_b) \qquad \text{(d)}$$
$$\text{mean adj-}R^2 = 0.82$$

Examining the mean adj-R^2 values in Table 13 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, the simulations of the models of various lengths in the shaded section of Table 13 have all been re-run and their mean AIC and adj-R^2 values recalculated. 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 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.

We could address this lack of monotonicity in the mean AIC values by averaging across more simulations. This would improve the stability of the mean AIC value and possibly alter the order of the explanatory variables selected. This is an approach taken in Woodd-Walker, Kingston and Gallienne (2001) who ran 100 simulations to address the variability in simulation results. However, any increase in adj-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 13(a) where the mean AIC increases from pressure to visibility. The mean AIC also increases from visibility to max-min temperature in greyed rows seven and eight. Furthermore, Table 11(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 13(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 13(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 11(d) for the explanatory variables for the driver pressure are the poorest of the four drivers. In Table 13(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 11(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 we have pruned the number of explanatory variables after considering poor signal to noise ratio that is increasing AIC.

4.3   Approximating expected yield using DNI data

DNI is the main driver for yield; therefore, it is informative to evaluate the DNI before examining the results of the yield modelling. This evaluation looks at two aspects: effective DNI versus satellite DNI and the annual variation in satellite DNI.

4.3.1   Effective versus Satellite DNI energy per area

Section 2.3 discusses effective DNI versus Satellite DNI providing reasons for effective DNI to be less then satellite DNI. Therefore, people basing yield calculations on satellite DNI will expect a higher yield than those using effective DNI. Table 14 compares the effective DNI from Allen (2013) with satellite DNI from BoM (2013) to calculate an effective DNI to satellite DNI ratio. Allen's data provides an incomplete coverage for year 2013 and SAM (2014) requires complete data for every driver for every hour of the year to calculate yield. Therefore, the comparison in Table 16 only totals energy per area (MWh/m^2) for 309 days in the first column. However, Allen (2013) has 343 complete days of DNI data. Hence, the second column in Table 14 shows 343 days of data. This extra number of days in the dataset of 343 days helps confirm the ratio 0.767 for the dataset of 309 days.

Table 14: Effective versus satellite DNI and ratio for 2013 for 309 and 343 days

	309 days	343 days
Effective DNI (Allen) – MWh/m^2	1.717	1.916
Satellite DNI (BoM) – MWh/m^2	2.239	2.500
Effective-satellite DNI ratio	0.767	0.767

A sensitivity analysis using SAM (2014) helps verify whether simply using the effective to satellite DNI energy per area ratio of 0.767 to adjust yield calculated using raw satellite DNI data is appropriate. Section 4.4.7 presents the results from performing such a sensitivity analysis. The conclusion is that it is unsuitable to adjust the yield using the effective to satellite DNI energy per area ratio. This finding has implications for the comparison site at Rockhampton that uses raw satellite DNI data to calculate yield. We are unable to simple adjust the Rockhampton yield with the effective to satellite DNI energy per area ratio. Therefore, in much of the subsequent analyses, we normalise the yield on year 2012 or 2013 before making comparisons between Rockhampton's yield calculated from raw satellite DNI data and Collinsville's yield calculated from terrestrially derived DNI data from Allen (2013). The normalisation process eliminates the need for the ratio. Section 5.2 discusses the effective-satellite DNI ratio further.

4.3.2   Variation in annual satellite DNI energy per area for 2007-13

The major driver for yield is DNI. Table 15 shows the variation in annual satellite DNI energy per area (MWh/m^2) for the years 2007-13 in the first row and the second row is the energy normalised to the year 2013. The model calibration year of 2013 has a markedly higher annual DNI energy per area than the projections years that are 2007-12. The annual variation in the plant's electricity power will reflect this annual variation in DNI energy.

Table 15: Variation in annual satellite DNI energy per area for 2007-13 for Collinsville

	2007	2008	2009	2010	2011	2012	2013
MWh/m^2	1.917	1.943	1.968	1.467	1.882	2.032	2.662
Normalised on 2013	72%	73%	74%	55%	71%	76%	100%

(Source: BoM 2013)

4.4 Calculating the yield from the Systems Advisor Model from four drivers

SAM (2014) produces both a gross yield and net yield projection from TMY files containing the four drivers. Gross yield less net yield gives the parasitic load. The gross yield is electricity generated by the LFR boiler and exported to the grid. The parasitic load is electricity used by the plant to operate and is imported from the grid. The price paid for the exported and imported electricity differs; therefore, this report uses gross yield analysis rather than net yield analysis. Additionally, the report uses 96% of the gross yield from SAM (2014) because we assume a 4% loss in yield to repair, maintenance and other down time.

Subsection 1 validates the driver models by comparing yield calculated using these models against yield calculated using Allen's (2013) data. Subsection 2 presents the 2007-2013 yield projections using the validated model and benchmarks the projections using the annual variation in DNI energy per area from the previous section. Subsection 3 uses the 2007-13 yield projections to calculate a TMY for the next report. Subsections 4 and 5 perform sensitivity analysis for the effect of climate change and elevation on yield, respectively. Section 6 uses the 2007-13 yield projections to calculate the exceedance rate of 30 MW.

4.4.1 Validating the weather modelling by comparing the solar electricity output

This section compares the electricity output or yield calculated using the modelled weather variables with the yield calculated using Allen's (2013) weather station data. This comparison helps validate the modelling of the four weather variables. The next section uses the validated weather models to calculate yield projections for years without Allen's weather station data. SAM (2014) calculates the yield from the four weather variables. Table 16 compares the average day's yield by month for 2013 calculated from the four drivers using the neural networks described in the previous section. Allen (2013) has data missing for some hours. These missing hours present a modelling problem because SAM (2014) is a dynamic model where the previous hour's values affect the next hour's yield. Additionally, the SAM's (2014) input files require a complete set of values for each of the 8,760 hours in a year for each of the four drivers. Therefore, the comparative analysis between Allen's (2013) data and the reports model uses 309 days because Allen's data has fifty-six days with one or more hour's data missing.

Table 16: Comparing average day's MWh gross yield by month for 2013 between Allen and Model

Yield (MWh)	Allen (309 days)	Model (309 days)	Model (365 days)
Jan	148.5	145.3	143.6
Feb	113.6	112.7	110.0
Mar	106.4	112.1	115.9
Apr	101.8	96.4	96.4
May	69.9	69.6	67.3
Jun	85.6	84.3	84.3
Jul	59.9	55.7	60.8
Aug	183.1	173.1	171.9
Sep	188.1	188.5	188.5
Oct	208.8	202.6	202.6
Nov	141.7	129.7	129.7
Dec	179.8	193.4	193.4
Annual	130.9	129.6	130.6

For the 309 days, Table 16 shows that there is discrepancy between Allen and the Model's average daily yield per month but the discrepancy in average daily yield per annum is less than 1%. This validates the model for yield projections for 2007-12 because the gas component of the hybrid gas-solar plant tops-up any variation in solar yield to 30 MW. Therefore, the plant would use the same quantity of gas over the year within 1% whether Allen's (2013) or the model's environment variables are prevalent. Additionally, the model projection over the entire 365 days of the year agrees with Allen to within 0.2%. Section 7.8 in further research discusses reasons for monthly distribution of yield not following the expected high yield in summer and low yield in winter.

4.4.2 Estimating solar electricity output for 2007-13 using weather model projections

Table 17 shows the solar plant's gross electrical output calculated using the modelled weather inputted into SAM (2014) for the years 2007-13. The second to last row in Table 17 shows the annual average daily electricity output normalised on the year 2013. This normalised electricity output is comparable to the normalised DNI energy per area for 2007-13 from Table 15. The last row of Table 17 shows this normalised DNI information for ease of comparison.

Table 17: Comparing average day's MWh yield by month for years 2007-13

Yield (MWh)	2007	2008	2009	2010	2011	2012	2013
Jan	54.9	57.5	13.9	32.6	102.9	106.6	143.6
Feb	68.2	47.0	38.4	36.0	64.3	86.8	110.0
Mar	92.2	69.9	123.0	41.5	19.2	43.7	115.9
Apr	97.9	110.2	98.1	41.3	68.8	96.7	96.4
May	44.8	70.2	37.8	81.5	93.0	66.6	67.3
Jun	37.8	50.0	80.1	64.6	67.4	52.7	84.3
Jul	107.9	68.2	99.1	41.2	94.0	53.4	60.8
Aug	70.7	109.0	134.3	75.5	81.9	101.6	171.9
Sep	152.8	125.6	158.9	55.4	153.3	121.7	188.5
Oct	165.3	136.5	177.0	104.8	113.9	170.1	202.6
Nov	81.7	134.8	94.1	21.3	117.9	172.3	129.7
Dec	85.5	116.2	123.6	38.4	75.1	155.1	193.4
Annual Yield	88.4	91.5	98.6	53.1	87.7	102.3	130.6
Annual Yield normalised on 2013	68%	70%	75%	41%	67%	78%	100%
Annual DNI energy per area normalised on 2013	72%	73%	74%	55%	71%	76%	100%

4.4.3 Selecting typical meteorological months from 2007-13 using the solar electricity output

Table 18 shows the selection of the twelve typical meteorological months (TMMs) from the years 2007-13 using the plant's solar electricity output to form the typical meteorological year (TMY) for this report and our subsequent *Energy economics and dispatch forecasting report* (Bell, Wild & Foster 2014a). Section 2.6 discusses the reason for using the plant's output to determine the TMMs.

Table 18 shows the plant's average daily energy output in MWh by month and year in the left panel. In contrast, the right panel shows the difference from the average month. The yellow highlighted differences are the smallest absolute differences from the average month. The months of the years that have the months' with the smallest difference from the average month form twelve TMMs to produce the TMY. For instance, the average daily energy output in January 2008 is 57.5 MWh, highlighted in blue. The January monthly average daily energy output is 73.1 MWh. Therefore, the difference is -15.7 MWh. This is the smallest absolute difference in January for years 2007 to 2013. Therefore, the TMM for January is from year 2008. The year 2009 has no TMMs. We have no demand data for 2013. Therefore, we expediently use year 2012 rather than year 2013 as to select the TMM for April. Section 5.7 discusses the importance of the year 2013 having no TMMs. Section 7.8 in further research compares the TMY selection process for yield with a TMY selection process for DNI to evaluate why the TMY monthly yield distribution fails to follow an expected summer-winter cycle.

Table 18: Selecting typical meteorological months from the years 2007-13 using the plant's solar electricity output

(MWh)	Plant's average daily energy output							Month's Ave.	Difference from month's average							TMM		
	2007	2008	2009	2010	2011	2012	2013		2007	2008	2009	2010	2011	2012	2013	year	value	diff.
Jan	54.9	57.5	13.9	32.6	102.9	106.6	143.6	73.1	-18.3	-15.7	-59.2	-40.5	29.8	33.4	70.4	2008	57.5	-15.7
Feb	68.2	47.0	38.4	36.0	64.3	86.8	110.0	64.4	3.8	-17.4	-26.0	-28.4	-0.1	22.4	45.6	2011	64.3	-0.1
Mar	92.2	69.9	123.0	41.5	19.2	43.7	115.9	72.2	20.0	-2.3	50.8	-30.7	-53.0	-28.5	43.7	2008	69.9	-2.3
Apr	97.9	110.2	98.1	41.3	68.8	96.7	96.4	87.0	10.8	23.1	11.1	-45.8	-18.3	9.6	9.4	2012	96.7	9.6
May	44.8	70.2	37.8	81.5	93.0	66.6	67.3	65.9	-21.1	4.3	-28.1	15.6	27.1	0.7	1.4	2012	66.6	0.7
Jun	37.8	50.0	80.1	64.6	67.4	52.7	84.3	62.4	-24.6	-12.4	17.7	2.1	5.0	-9.7	21.9	2010	64.6	2.1
Jul	107.9	68.2	99.1	41.2	94.0	53.4	60.8	74.9	33.0	-6.7	24.1	-33.7	19.1	-21.6	-14.2	2008	68.2	-6.7
Aug	70.7	109.0	134.3	75.5	81.9	101.6	171.9	106.4	-35.7	2.6	27.9	-30.9	-24.5	-4.8	65.4	2008	109.0	2.6
Sep	152.8	125.6	158.9	55.4	153.3	121.7	188.5	136.6	16.2	-11.0	22.3	-81.2	16.7	-14.9	51.9	2008	125.6	-11.0
Oct	165.3	136.5	177.0	104.8	113.9	170.1	202.6	152.9	12.4	-16.4	24.1	-48.1	-39.0	17.2	49.7	2007	165.3	12.4
Nov	81.7	134.8	94.1	21.3	117.9	172.3	129.7	107.4	-25.7	27.4	-13.3	-86.1	10.5	64.9	22.3	2011	117.9	10.5
Dec	85.5	116.2	123.6	38.4	75.1	155.1	193.4	112.5	-27.0	3.7	11.1	-74.1	-37.3	42.7	80.9	2008	116.2	3.7
Annual	88.4	91.5	98.6	53.1	87.7	102.3	130.6	93.0	-4.6	-1.5	5.6	-39.9	-5.3	9.3	37.6	2008	93.5	0.5

4.4.4 What is the effect of climate change on yield?

Section 2.5 in the literature review and Section 3.5 in the methodology discuss the effect of climate change on yield. Table 5 in Section 2.5 shows the expected changes in the drivers for three Global Climate Models for the coolest, most likely and hottest cases. Table 19 compiles from Section 3.5 the values for the four drivers for yield for the 1990 baseline year and from Table 5 the expected changes from the baseline year.

Table 19: Values representing the four drivers for the base year 1990 and sensitivities for Collinsville

Driver	1990 baseline	Coolest case MIROC3.2-Medres	Most likely case MRI-CGCM2.3.2	Hottest case CSIRO-Mk3.5
DNI (W/m^2)	767	759	768	773
Dry temperature (°C)	26.1	27.3	27.2	27.5
Relative humidity (%)	55	55.4	54.6	54.5
Pressure (hPa)	989.5	989.5	989.5	989.5

Table 20 shows SAM's (2014) annual gross yield calculations using the values for the four drivers in Table 19. Table 20 also shows the percentage change in annual yield induced by climate change from 1990 to 2040 in the three GCMs. The lifetime of the plant is less than the period 1990 to 2040. Therefore, the magnitude of the percentage change in gross yield will be less.

Table 20: Climate change induced percentage change in yield

	1990 baseline	Coolest case MIROC3.2-Medres	Most likely case MRI-CGCM2.3.2	Hottest case CSIRO-Mk3.5
Gross Yield (GWh)	82.80	81.81	83.06	83.82
Change in yield (%)	0.0%	-1.2%	0.3%	1.2%

4.4.5 Does elevation between Collinsville and nearby sites affect yield?

Section 2.2.4.1 in the literature review and Section 3.6 in the methodology discuss the effect of altitude on yield. Table 21 compiles from Sections 3.5 and 3.6 the values for the four drivers for yield for the Collinsville to Rockhampton altitude sensitivity analysis.

Table 21: Values for the four drivers for Collinsville-Rockhampton altitudes sensitivity analysis

Driver	Collinsville 1990 baseline	Rockhampton 1990
Altitude (m)	197	13
DNI (W/m^2)	767	767
Dry temperature (°C)	26.1	24.9
Relative humidity (%)	55	57.5
Pressure (hPa)	989.5	1000

Table 22 shows SAM's (2014) gross yield calculations using the values for the four drivers in Table 21. Table 22 also shows the percentage change in yield induced by altitude change from Collinsville to Rockhampton given constant DNI but allowing dry temperature and relative humidity to change for altitude. Ignoring any altitude effects on DNI, the effect of

altitude on yield is slight. Therefore, Rockhampton remains a potential candidate to proxy yield for Collinsville provided we adjust DNI.

Table 22: Altitude induced percentage change in yield

	Collinsville 1990 baseline	Rockhampton 1990
Gross Yield (GWh)	82.80	82.67
Change in yield (%)	0.0%	-0.1%

4.4.6 Analysis of the plant's LFR gross electricity output exceeding 30 MW

The frequency of exceedance of 30 MW will determine whether to pay for higher transmission capacity for the plant. This decision, in turn, will determine whether the modelling in our subsequent report (Bell, Wild & Foster 2014a) will dispatch the entire yield of the LFR or spill power exceeding 30 MW.

Table 23 : Analysis of the plant's gross electricity output exceeding 30 MW

	2007	2008	2009	2010	2011	2012	2013
Tot energy (GWh)	198	158	148	3	203	249	1,065
Max power (GW)	5.9	5.5	5.9	1.4	6.0	6.0	6.1
Mean power (GW)	2.9	2.6	3.2	0.5	3.3	2.4	3.7
3rd Quartile (GW)	4.3	4.2	5.1	0.7	5.1	3.6	5.9
Median (GW)	3.0	2.7	3.1	0.3	3.5	1.9	4.0
1st Quartile (GW)	0.2	0.1	0.0	0.1	0.1	0.0	0.1
Number of hours	68	60	46	6	62	102	288
Operational (%)	1%	1%	1%	0%	1%	1%	3%

4.4.7 Is the method to use raw satellite DNI data to calculate yield and retrospectively adjust the calculated yield with an effective to satellite DNI energy per area ratio suitable?

Section 3.7 discusses the methodology for the sensitivity analysis in this section. Table 24 presents the results of the sensitivity analysis. We use the same effective to satellite DNI ratio of 0.767 on three raw satellite DNI values of high, medium, and low to produce the adjusted satellite DNI data. We use SAM (2014) to calculate the annual energy yield from the raw and adjusted satellite DNI data. The raw to adjusted satellite yield ratio shows considerable variation without a simple correlation with the effective to satellite DNI ratio. We conclude that it is unsuitable to adjust retrospectively the yield calculated using raw satellite data with an effective to satellite DNI energy per area ratio.

Table 24: Results of the suitability of retrospectively adjusting yield based on raw satellite DNI data

	Satellite data	high	medium	low
DNI (W/m2)	raw	1,000	667	333
	adjusted	767	511	256
Yield (MWh)	raw	108,192	68,574	20,171
	adjusted	82,795	46,247	8,102
Raw to adjusted Satellite yield ratio		0.77	0.67	0.40

4.4.8 How has climate change affected the ENSO cycle?

Section 3.8 discusses the methodology for this research question. We simply compare the mean of the SOI for the period 1876-1944 with 1945-2013 to test for an increase in a La Niña or El Niño bias.

We find the period 1876-1944 has an SOI of 0.26 and the period 1945-2013 and SOI of 0.14. Therefore, climate change is inducing an El Niño bias. This result calls into question the suitability of including earlier years in a TMY to average out ENSO effects. However, we recommend further research to test the statistical significance of the result. Additionally, would the bias outweigh the benefit of including earlier years in a TMY?

5   Discussion

5.1   Introduction

This section discusses the research questions and results within a wider context.

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, we expect Collinsville to have a slightly higher yield and the yield in the afternoon will be particularly higher. The results of the elevation sensitivity analysis in Section 4.4.5 confirm this initial evaluation. However, the sensitivity analysis finds Rockhampton's yield is only 0.1% less than Collinsville's yield allowing pressure, temperature, and humidity to vary with elevation but assuming that DNI is constant.

Section 5.2.1 compares the BoM DNI satellite data for Collinsville with three nearby sites MacKay Aero, Rockhampton Aero, and Townville Aero. Section 5.2.2 compares this report's yield calculations for Collinsville with the yield calculated by Shah, Yan and Saha (2014b) for a comparable plant sited at Rockhampton Aero.

5.2.1   Comparing DNI satellite data for Collinsville and Rockhampton Aero

Figure 3 compares the total annual BoM (2013) satellite hourly gridded DNI energy per area (MWh/m2) for MacKay Aero, Rockhampton Aero and Townsville Aero normalised against Collinsville's annual DNI energy per area for the year 1990 to 2013. Figure 3 shows that the comparison sites have higher satellite DNI energy per area than Collinsville for most years and this difference reduces from about 2007.

Figure 3 shows that Rockhampton for the years 2007-13 has a slightly higher DNI per area than Collinsville, which ameliorates concerns over the results from the elevation sensitivity analysis in Section 4.4.5, which finds that Rockhampton's yield is 0.1% less than Collinsville's yield. Therefore, Rockhampton remains a potential yield proxy for Collinsville for 2007-13. However, earlier years show a considerable deviation in DNI between Rockhampton and Collinsville, which would make Rockhampton a less than ideal proxy for yield at Collinsville.

Figure 3: Percentage deviation of DNI energy from Collinsville for comparison sites
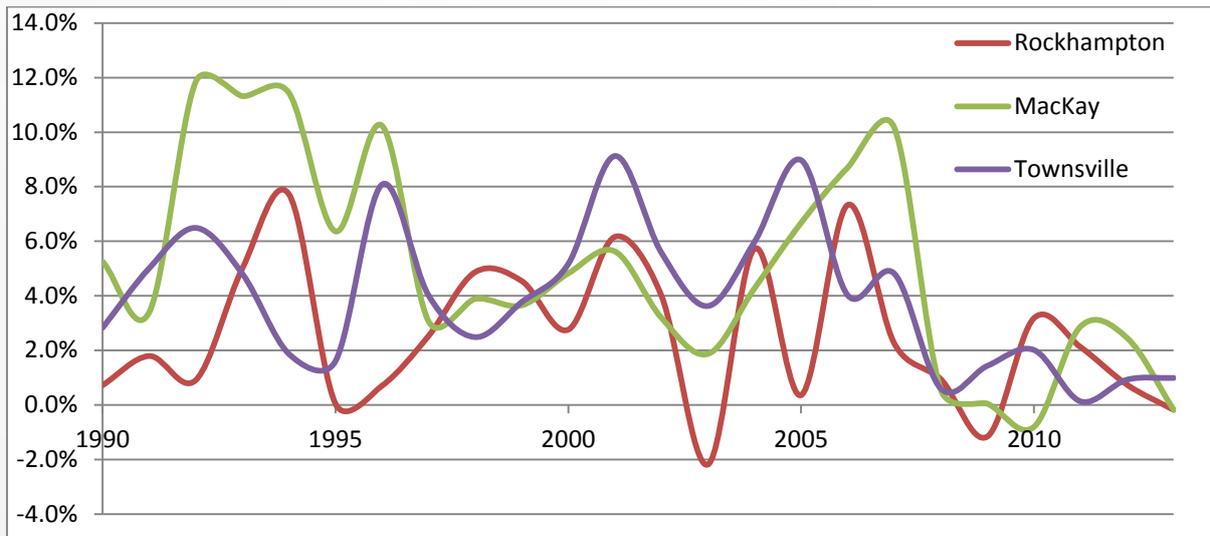
(Source: BoM 2013)

This normalisation process in Figure 3 serves two further functions. First, the normalisation more clearly identifies any deviation from Collinsville's annual DNI energy per area. The next section uses this percentage deviation to adjust the yield calculation by Shah, Yan and Saha (2014b) for the Rockhampton comparison site. Second, the BoM (2013) hourly gridded data has missing days and hours particularly in the earlier years, see Figure 4. This missing data makes inter-year comparisons inaccurate but the same hourly gridded data is missing for all sites, which allows inter-site comparison. This inter-site comparison is the focus of this section.

Figure 4: Annual number of hours of DNI satellite data for years 1990-2013
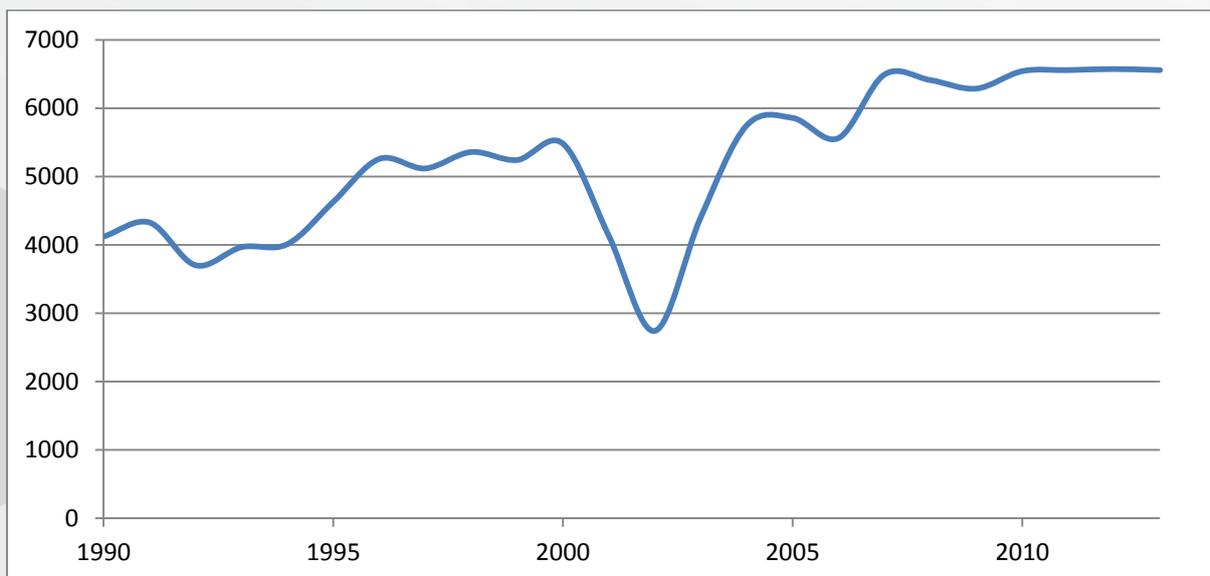
(Source: BoM 2013)

The BoM (2013) provides three general reasons for the missing satellite DNI data in Figure 4: the satellite image was unavailable, the image was unprocessed, or the image failed quality control. Reasons that are more specific include:

- The lack of values for the first two hours and last two hours of the day for the period up until 30 Jun 1994, due to the absence of satellite images at these times during the initial period of operation of satellite GMS4.
- The period July 2001 to June 2003 has fewer values because there was reduced imaging frequency at the end of the life of satellite GMS-5, and the initial few weeks of operation of satellite GOES-9 in the Australian region.
- BoM calculates the DNI from GHI. Therefore, if any GHI is missing, there is no DNI.

This report develops its TMY from the period 2007-13 using the satellite DNI data to model effective DNI and then yield. Figure 4 shows that the report's TMY period has more annual DNI satellite images than the previous years, which provides for less interpolation and more accurate interpolation. Additionally, Figure 3 shows a convergence of the percentage deviation of the comparison sites in the period 2007-13. Both factors provide more confidence in the use of the percentage deviation during the report's TMY period to compare yield for Rockhampton Aero calculated by Shah, Yan and Saha (2014b) with the yield for Collinsville calculated in this report. Conversely, the period prior to 2007 carries less confidence. Sections 5.6 and 5.7 make comparisons between the report's TMY period and periods prior to 2007 to evaluate whether the reports TMY adequately represents any El Niño Southern Oscillation effects.

Section 7.3 in further research discusses an alternative way to calculate the effective DNI at Collinsville using the inter-site variation in Figure 3 and the one-minute solar data from BoM (2012).

5.2.2 Comparing Collinsville and Rockhampton yield calculations

In order to address the research question, Shah, Yan and Saha (2014b) calculate the yield for a comparable Collinsville LFR plant situated at three comparison sites using datasets from Exemplary Energy (2014). However, Shah, Yan and Saha (2014b) found that MacKay and Townsville had gaps in their datasets, which leaves only Rockhampton Aero as a comparison site. Furthermore, the data for the year 2013 was unavailable from Exemplary Energy (2014) when Shah, Yan and Saha (2014b) wrote their report.

Therefore, to calculate yield for 2013, Shah, Yan and Saha (2014b) use a TMY, as proxy for the year 2013. The Commonwealth of Australia, Department of the Environment and Water Resources, Australia Greenhouse Office developed this TMY for use in Building Code compliance. The Greenhouse Office TMY includes twelve TMMs from the years 1974 to 2004 and is available for download from EnergyPlus (2014).

5.2.2.1 Comparing Collinsville and Rockhampton 2013 yield using the Greenhouse Office TMY and Allen's data

Shah, Yan and Saha (2014b) use the Greenhouse Office TMY as a proxy for 2013 weather. However, there are limitations to comparing the yield calculated using the Greenhouse Office TMY as a proxy for weather in 2013 with the yield calculated using the 2013 weather station data from Allen (2013). Specifically, the Greenhouse Office TMY consists of twelve TMMs from the years 1974 to 2004 and Allen's 2013 data lies outside these years.

Moreover, the effect of El Niño Southern Oscillation (ENSO) on yield makes comparison between these periods problematic because the average SOI for 1974 to 2004 is -1.98 and for 2013 is 4.033, indicating both periods are in different phases of the ENSO cycle.

Additionally, there are limitations to comparing the yield calculated from the Greenhouse Office TMY with the yield calculated from this report's TMY for two reasons. First, this report's TMY provides a TMY constructed from the yield of the LFR plant whereas the Greenhouse Office's TMY provides a TMY constructed for building code compliance. Second, this report's TMY uses the years 2007 to 2013 whereas the Greenhouse Office's TMY uses the years 1974 to 2004. The effect of ENSO on yield makes comparison between these periods problematic because the average SOI for 1974 to 2004 is -1.98 and for 2007 to 2013 is 5.4, indicating both periods are in different phases of the ENSO. Sections 5.6 and 5.7 discuss further the ENSO with respect to this report's TMY period and earlier periods.

Table 25 allows ease of comparison between the yield results for Collinsville and Rockhampton for the year 2013. The Collinsville panel in Table 25 replicates Table 16 and the Rockhampton panel shows the yield calculations from Shah, Yan and Saha (2014b). The average daily annual yield calculated using the Greenhouse Office TMY data for Rockhampton is about 8% higher than that calculated using Allen's ground based data. Two factors can explain this difference:

- the effective-satellite DNI energy per area ratio of 0.767 shown in Table 14; and
- the Rockhampton-Collinsville satellite DNI ratio in Figure 3

The Rockhampton-Collinsville satellite DNI ratio for 2013 is 0%. Therefore, we exclude any adjustment for the ratio in Table 25.

Table 25: Comparing average day's MW yield by month for 2013 among Allen, Model and Rockhampton

Month	Collinsville			Rockhampton	
	Allen (2013) (309 days)	Model (309 days)	Model (365 days)	Shah, Yan and Saha (2014b) (365 days)	(x 0.767)
Jan	148.5	145.3	143.6	163	125
Feb	113.6	112.7	110.0	143	110
Mar	106.4	112.1	115.9	151	116
Apr	101.8	96.4	96.4	130	100
May	69.9	69.6	67.3	100	77
Jun	85.6	84.3	84.3	106	81
Jul	59.9	55.7	60.8	122	94
Aug	183.1	173.1	171.9	130	100
Sep	188.1	188.5	188.5	150	115
Oct	208.8	202.6	202.6	167	128
Nov	141.7	129.7	129.7	171	131
Dec	179.8	193.4	193.4	163	125
Annual	130.9	129.6	130.6	141	108

The annual yield calculated in this report's model for Collinsville and that yield based on the Greenhouse Office TMY calculated by Shah, Yan and Saha (2014b) are about 17% different from Allen's weather station data after adjusting the Rockhampton yield calculation by the effective-satellite energy per area ratio of 0.767. Adjusting the yield calculated using the

Greenhouse Office TMY by the effective-satellite DNI energy per area ratio over compensates but comparing yield from the TMY and 2013 is a poor test for the usefulness of the ratio.

Nevertheless, factoring the yield results from Shah, Yan and Saha (2014b) by 0.767 does allow easy comparison between the patterns of monthly yield amongst the yield based on Allen's data and this reports' model. The report's model more closely reflects the monthly variation in the yield based on Allen's data than does the Rockhampton yield. This Rockhampton's monthly yield disparity result is unsurprising given Rockhampton's yield is based on a TMY. This report's yield model has more predictive power of the yield based on ground-based data from Allen's instrument than the yield based on the Rockhampton data. However, the result is ambiguous because it fails to address the research question adequately. Rockhampton yield based on 2013 weather is required for a fair comparison between this report's model of Collinsville's yield and a Rockhampton yield model. Sections 7.2 and 7.3 in further research discuss ways to improve the Rockhampton comparison.

5.2.2.2 Comparing Rockhampton 2007-12 yield using Exemplary Datasets against BoM DNI

The major driver for yield is DNI. Table 26 shows the variation in annual satellite DNI energy per area (MWh/m^2) for the years 2007-12 for Rockhampton in the first row and the second row is the energy normalised to the year 2012. The annual variation in the Rockhampton plant's electricity yield should reflect this annual variation in DNI energy.

Table 26: Variation in annual satellite DNI energy per area for 2007-12 for Rockhampton

	2007	2008	2009	2010	2011	2012
MWh/m^2	1.955	1.955	1.939	1.510	1.918	2.044
Normalised on 2012	96%	96%	95%	74%	94%	100%

(Source: BoM 2013)

Table 27 shows the average day's MWh yield by month for the years 2007-12 for Rockhampton calculated by Shah, Yan and Saha (2014b) using the data from Exemplary Energy (2014). The second to last row in Table 27 shows the annual average daily electricity output normalised on the year 2012. This normalised electricity output should be comparable to the normalised DNI energy per area for 2007-12 in Table 26. The last row of Table 27 shows this normalised DNI information for ease of comparison. The years 2007, 2008 and 2010 show considerable divergence between the Normalised DNI and yield.

Exemplary Energy (2014) use the satellite data from BoM (2013) satellite in developing their weather input files but modify the data (Exemplary Energy 2013). This modification could account for some of the discrepancy between the normalised DNI energy per area and yield in Table 27.

Table 27: Comparing average day's MWh yield by month for years 2007-12 for Rockhampton

Yield (MWh)	2007	2008	2009	2010	2011	2012
Jan	112	72	68	109	204	167
Feb	79	64	126	61	162	196
Mar	140	127	176	81	87	128
Apr	132	128	145	89	119	163
May	87	93	96	98	136	113
Jun	40	65	123	64	98	59
Jul	108	58	139	36	130	78
Aug	64	113	172	80	104	124
Sep	139	110	187	63	196	181
Oct	157	144	217	124	151	199
Nov	102	139	215	30	192	218
Dec	99	157	172	67	107	234
Annual Yield	105	106	153	75	141	155
Annual Yield Normalised on 2012	68%	68%	99%	48%	91%	100%
DNI energy per area normalised on 2012	96%	96%	95%	74%	94%	100%

(Source: Shah, Yan & Saha 2014b)

5.2.2.3 Comparing Collinsville and Rockhampton yield by referencing BoM's satellite DNI energy per area for 2007-13

We adjust Rockhampton's yield by the effective to satellite DNI ratio and Rockhampton to Collinsville DNI ratio before making a comparison.

Table 28 shows Rockhampton's annual average daily yields adjusted for the satellite to effective DNI ratio of 0.767. Rockhampton's annual average daily yields are from Table 25 and Table 27. Section 4.3.1 discusses the satellite to effective DNI ratio.

Table 28: Adjusting Rockhampton's yield for satellite-effective ratio and Rockhampton-Collinsville ratio

	2007	2008	2009	2010	2011	2012	2013
Rockhampton Yield (MWh)	105	106	153	75	141	155	141
Adjusted for Satellite-effective DNI ratio 0.767	80	81	117	58	108	119	108
Adjusted for Rockhampton-Collinsville DNI ratio	79	80	119	56	106	118	108

Table 29 calculates the Rockhampton to Collinsville DNI energy per area ratio for the years 2007-13 using data from BoM (2012). Figure 3 shows this ratio graphically as a percentage deviation.

Table 29: Collinsville and Rockhampton's annual satellite DNI energy per area and ratio

(MWh/m^2)	2007	2008	2009	2010	2011	2012	2013
Collinsville	1.917	1.943	1.968	1.467	1.882	2.032	2.662
Rockhampton	1.955	1.955	1.939	1.510	1.918	2.044	2.652
Ratio	1.02	1.01	0.99	1.03	1.02	1.01	1.00

Table 30 makes a comparison between the annual average daily yield from Collinsville and Rockhampton normalised on the year 2013. Table 30 repeats from Table 17 for ease of

comparison of both the Collinsville annual Satellite DNI and annual average daily yields normalised on 2013. We calculate the normalised Rockhampton yield data in Table 30 from the yield data in Table 25 and Table 27, which we adjust by the Collinsville to Rockhampton DNI ratio in Table 29.

Table 30 shows that the normalised modelled yield for Collinsville more closely follows the normalised DNI energy per area for Collinsville than the normalised yield for Rockhampton even after adjusting Rockhampton's yield for both Rockhampton to Collinsville DNI ratio and satellite to effective DNI ratio.

Table 30: Comparing normalised annual Rockhampton and Collinsville daily average yield and BoM's satellite annual DNI energy per area for the years 2007-13

	2007	2008	2009	2010	2011	2012	2013
Collinsville Satellite DNI	72%	73%	74%	55%	71%	76%	100%
Collinsville yield	68%	70%	75%	41%	67%	78%	100%
Rockhampton yield	73%	74%	110%	52%	97%	109%	100%

In summary, we find that modelling the weather with limited datasets from Collinsville produces greater yield predictive power than using the historically more complete datasets from nearby Rockhampton. However, we recommend using one-minute solar data from BoM (2012) for Rockhampton to both improve the effectiveness of the comparison and potentially provide a way to improve the predictive power of the model for Collinsville. Section 7.2 and 7.3 in further research discuss the one-minute data from the BoM in more detail.

5.3 Does BoM adequately adjust its DNI satellite dataset 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 11, Table 12 and Table 13.

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 value 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 their derivation. This further reduces the role of *cloud$_b$* in the final model.

In summary, we find that the BoM adequately adjusts for cloud coverage in the DNI satellite dataset.

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. Additionally, Section 4.3.1 discusses the DNI effective to satellite ratio of 0.767 for the Allen's data against the BoM's satellite data. The effective-satellite DNI ratio of 0.767 and the mean adj-R^2 = 0.81 indicate that the raw satellite data overestimates DNI energy.

However, cleaning regimes can alter the effective-satellite DNI ratio. For instance Allen (2013) in his January to May 2013 report states that construction at the Powerlink Substation 50 m from the DNI measuring instrument is causing dust and mud contamination and bird fouling on the instruments has necessitated a daily cleaning regime. Additionally, consecutive wet days caused a number of problems including short-circuit and battery failure. Allen (2013) in his September to November 2013 report states that Transfield are cleaning the equipment daily but recently they changed to every second day. The reductions in cleaning regime, consecutive wet day problems and surrounding earth works would reduce the effective-satellite DNI ratio.

Furthermore, differences in cleaning regimes for the measuring instruments (Allen 2013) and LFR boiler (Guan, Yu & Gurgenci 2014) would induce disparity between the effective-satellite DNI ratios for each. Reducing this disparity requires some coordination between the people overseeing the DNI measuring instrument and the people overseeing the cleaning regime for the LFR, so the cleaning regime for the measuring instrument mimics the cleaning regime for the LFR. Currently, cleaning for both the LFR and measuring instrument are undergoing change induced by ongoing research. Therefore, coordination for the current project is difficult but future projects will have the findings of the current research to help establish the required coordination.

Determining whether this effective-Satellite DNI ratio is more widely applicable to other sites warrants further research. Section 7.2 discusses using the one-minute solar data collected from ground-based weather stations available from BoM (2012) with the hourly gridded satellite data from BoM (2013) to determine the wider applicability of the effective-satellite DNI ratio.

In summary, we find that raw satellite data overestimates effective DNI. Therefore, we recommend adjusting raw DNI satellite data either by modelling effective DNI or by using an effective-satellite DNI energy per area ratio before calculating yield to prevent overestimation. We also recommend coordination between overseers of the DNI measuring instrument and LFR to ensure the effective-satellite DNI ratio is comparable.

5.5   Does elevation between Collinsville and nearby sites affect yield?

Section 2.2.4 discusses the lapse rates for temperature and pressure, that is, the change in temperature or pressure with change in elevation. Temperature and pressure decrease with increased elevation but this relationship is far from simple with temperature having three lapse rates interrelated with 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 3 shows that the temperature range for Collinsville is wider relative to the three comparison sites.

Section 4.4.5 finds a 0.1 decrease in yield by decreasing elevation from Collinsville to Rockhampton. This sensitivity analysis kept DNI constant and adjusted for changes in temperature, humidity, and pressure. However, Figure 3 shows the inter-year variability in DNI between Rockhampton and Collinsville often exceeds this elevation effect.

In summary, other effects such as variability in DNI, the presence or absence of a sea breeze play a more important role in determining yield.

5.6   How does the ENSO cycle affect yield?

Section 2.4 discusses the ENSO cycle 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.

To address this research question quantifiably, Shah, Yan and Saha (2014b) calculate the yield from a comparable LFR plant located at Rockhampton using datasets of the four drivers from Exemplary Energy (2014). Table 31 and Table 32 show the average daily yield of the Rockhampton plant for the period 2007-12 and 2000-2005, respectively. In comparison, Figure 5 shows the mean annual SOI for 1875-2013 where positive SOI indicates a La Niña (BoM 2014d) bias and the negative SOI indicates an El Niño (BoM 2014c) bias. The mean annual SOI for the period 2007-12 is 5.63, which indicates a strong La Niña bias, and for the period 1999-2006 is -0.4167, which indicates El Niño bias. Therefore, we would expect a higher average yield for the period 1999-2006 than 2007-12. In agreement, the average daily yield for 2000-05 in Table 21 is 0.140 GWh and for 2007-12 in Table 20 is 0.122 GWh. Hence, the El Niño weather pattern provides a 14% increases in yield over La Niña weather pattern.

Table 31: Average daily yield of Rockhampton 2007-2012

(GWh)	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec
2007	0.112	0.079	0.14	0.132	0.087	0.040	0.108	0.064	0.139	0.157	0.102	0.099
2008	0.072	0.064	0.127	0.128	0.093	0.065	0.058	0.113	0.11	0.144	0.139	0.157
2009	0.068	0.126	0.176	0.145	0.096	0.123	0.139	0.172	0.187	0.217	0.215	0.172
2010	0.109	0.061	0.081	0.089	0.098	0.064	0.036	0.08	0.063	0.124	0.03	0.067
2011	0.204	0.162	0.087	0.119	0.136	0.098	0.13	0.104	0.196	0.151	0.192	0.107
2012	0.167	0.196	0.128	0.163	0.113	0.059	0.078	0.124	0.181	0.199	0.218	0.234
Monthly	0.122	0.115	0.123	0.129	0.104	0.075	0.092	0.11	0.146	0.166	0.149	0.138

(Source: Shah, Yan & Saha 2014b)

Table 32: Average daily yield of Rockhampton 2000-2005

(GWh)	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec
2005	0.163	0.213	0.148	0.118	0.126	0.052	0.099	0.089	0.17	0.14	0.209	0.247
2004	0.166	0.17	0.168	0.135	0.118	0.096	0.09	0.167	0.19	0.21	0.187	0.185
2003	0.179	0.09	0.117	0.115	0.077	0.059	0.078	0.098	0.174	0.147	0.073	0.126
2002	0.156	0.112	0.153	0.112	0.087	0.07	0.093	0.091	0.168	0.193	0.209	0.189
2001	0.219	0.172	0.151	0.106	0.16	0.072	0.128	0.198	0.179	0.234	0.167	0.217
2000	0.147	0.102	0.116	0.112	0.109	0.095	0.09	0.115	0.134	0.162	0.104	0.145
Monthly	0.172	0.144	0.142	0.116	0.113	0.074	0.096	0.128	0.169	0.181	0.158	0.185

(Source: Shah, Yan & Saha 2014b)

Figure 5: Mean annual Southern Oscillation Index 1875-2013
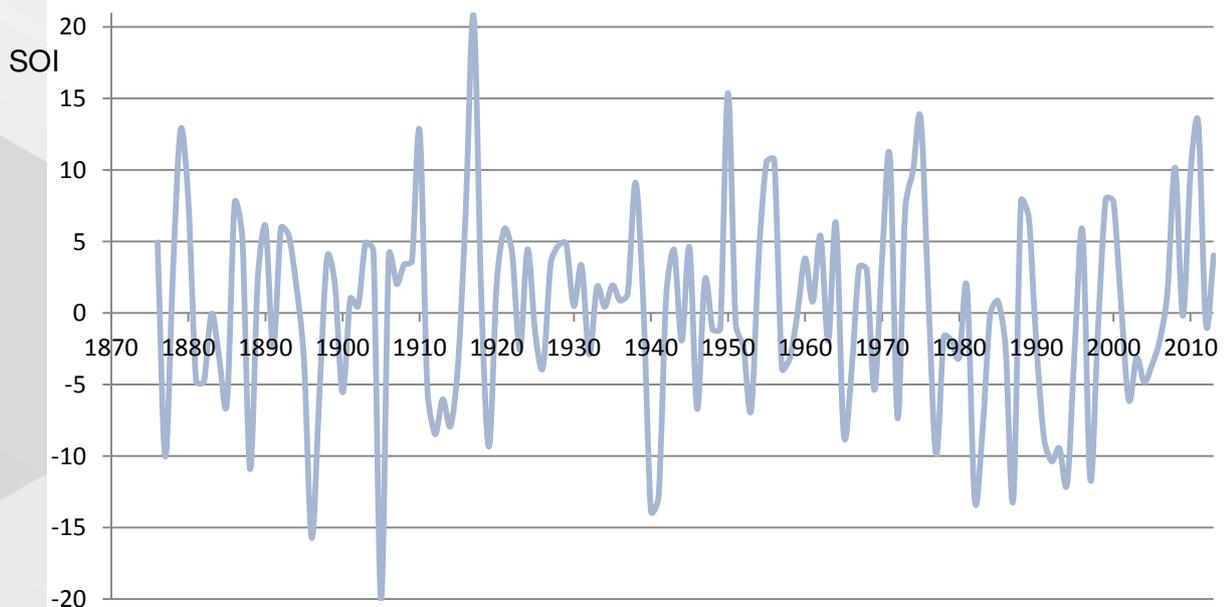
(Source: BoM 2014e)

However, this analysis comes with three caveats.

- Shah, Yan and Saha (2014b) use DNI data from Exemplary Energy (2014) in their yield calculations. Exemplary Energy (2014) use DNI satellite data from BoM (2013) to develop their DNI projects. Figure 4 shows that the number of hours of satellite imagery for BoM (2013) differs considerably between the periods 2000-05 and 2007-12. For instance, there are 2,741 hours of images in 2002 and 6,572 hours in 2012. Therefore, the above analysis compares a period using a high proportion of interpolated data with a period containing mostly original data.

- Section 5.2.2.2 discusses the discrepancies between the yield calculated in the period 2007-12 using Exemplary Energy (2014) data and the BoM (2013) satellite DNI data highlighted in Table 27.

- The TMY developed in this report uses the period 2007-13 and the comparison period in Shah, Yan and Saha (2014b) uses the period 2007-12. Section 5.7 discusses how comparing the average yield from 2000-05 with that yield from 2007-12 rather than 2007-13 will overstate the ENSO induced difference in yield.

In summary, we recommend readdressing this research question with one-minute solar data (BoM 2012) to ensure measurement consistency between the periods being analysed. Sections 7.2, 7.3 and 7.7 in further research discuss this recommendation in more detail.

5.7 Given the 2007-12 electricity demand data constraint, will the 2007-13 based TMY provide a "Typical" year over the ENSO cycle?

Section 2.6 discusses the TMY process and the period 2007-2012 constraint due to available electricity demand data in our subsequent report (Bell, Wild & Foster 2014a). However, in Section 4.4.3 during the development of the report's TMY, we found it possible to both meet the electricity data constraint and include the year 2013 within the TMY processes to select the 12 TMMs because the year 2013 had relatively high yields for all months. Therefore, the report's TMY process includes the years 2007-13 but selects no TMMs from 2013. Consequently, this 2007-13 based TMY both meets the 2007-12 constraint of the electricity demand data and the inclusion of the year 2013 acts to increase the yield of each TMM to improve the TMY's ability to represent the ENSO as Section 5.6 discusses.

The analysis in section 5.6 discusses the average daily yield for Rockhampton for the El Niño period 2000-05 at 0.140 GWh and for the La Niña period 2007-12 at 0.122 GWh. The yield calculated in the El Niño period is 14% higher than the La Niña period. We would expect Collinsville to follow a similar pattern. However, three factors ameliorate concerns over this 14% difference in yield between the two periods.

- The 40-year lifetime of the proposed plant requires considering the long-term mean annual SOI rather than the short comparison periods. The mean annual SOI for 1875-2013 is 0.02 and the mean annual SOI for the comparison periods 1999-2006 and 2007-12 is -0.4167 and 5.63, respectively. Both these comparison periods are bias in opposite directions to longer-term mean annual SOI. Hence, the 14% overstates the difference from the long-term 1875-2013 mean annual SOI.

- The report's TMY incorporates the extra year 2013 to cover the period 2007-2013 whose mean annual SOI is 5.40. This SOI is less than the mean annual SOI 5.63 for the comparison period 2007-2012. Hence, the 14% overstates the difference in yield since this report's TMY incorporates the year 2013.

- Figure 4 shows that the number of hours of satellite imagery per year prior to 2007 is sparser than after 2007. This sparseness reduces the confidence in the yield calculation for the comparison period 1999-2006.

Extending the 2007-13 period of the TMY would encompass more ENSO cycles to allow a more representative TMY. However, the current TMY period 2007-13 uses actual electricity demand and extending this period would entail modelling demand, which would introduce modelling error into any calculations. Exacerbating the demand modelling error is the sparseness of satellite DNI data prior to 2007. An alternative to using satellite data is the one-minute data from BoM (2012) which starts in 1996 for Rockhampton and data collection continues to date. Additionally, the period 1996-2013 has a mean average SOI of 1.52 that is much closer to the long-term SOI of 0.02.

In summary, there are good grounds to assume that yield based on the report's TMY will underreport yield due to ENSO but this underreporting will be less than 14%. We recommend readdressing this research question with one-minute solar data (BoM 2012) to ensure measurement consistency between the periods being analysed and to accurately quantify the underreporting of yield. Sections 7.2 and 7.3 in further research discuss the use of the one-minute data in more detail. Section 7.7 discusses increasing the number of years in the TMY to average out the ENSO cycle.

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 climate change data-series lack projections for atmospheric 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.

Section 4.4.4 presents the results of a sensitivity analyses on the climate change induced change in gross yield from 1990 to 2040 in Table 20. In the most likely and hottest cases, the increase in yield is less than 0.3% and 1.2% respectively and in the coolest cases, the decrease is less than 1.2%. 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 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, Section 4.4.4 presents only the change in gross yield and for a complete analysis the change in wholesale electricity prices induced by climate change needs consideration. Bell and Wild (2013) calculate the effect of climate change on the NEM's demand from 2009 to 2030. They find a 1.57% increase in total net demand and a 2.23% increase in peak demand for the Collinsville node, which would have an upward pressure on wholesale spot prices. Our subsequent report (Bell, Wild & Foster 2014a) specifically addresses wholesale spot prices and will make any recommendations for further research.

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 (2014) default "Linear Fresnel Novatec Solar Boiler" that is modified to reflect the LFR at Collinsville. These modifications indicate the boiler is likely to exceed the 30MW limit imposed by AEMO. This issue of exceeding the AEMO limit requires consideration of both frequency and size of exceeding the limit to determine whether exceeding the limit is acceptable by AEMO or spillage is required.

In summary, section 4.4.6 and Shah, Yan and Saha (2014b) present yield exceedance analysis for Collinsville and Rockhampton respectively. Consequently, RATCH Australia decided to export all yield from the LFR to the grid based on the analysis.

6   Conclusion

In this report, we have addressed the research questions and produced a TMY of yield for the proposed plant at Collinsville for use in our subsequent report *Energy economics and dispatch forecasting* (Bell, Wild & Foster 2014a). We have calibrated our yield model for the proposed plant against weather observations provided by Allen (2013) to within 1% of gross annual yield, see Section 4.4.1. We have also introduced a new technique to develop TMYs for renewables energy generation, see Sections 2.6 and 4.4.3. Additionally, we have introduced the term effective DNI to help supplement our discussion of raw and adjusted satellite DNI data and terrestrially measured DNI, see Sections 2.3 and 4.3.1.

This report uses the 2013 terrestrially measured DNI provided by Allen (2013) to model DNI for the years 2007 to 2013. In contrast, other reports use satellite data provided by BoM (2013) either directly or indirectly. We found a ratio of 0.767 between the terrestrial and satellite DNI energy per square meter for 2013. Consequently, the yield calculated in this report based on terrestrially measured DNI is lower than reports that use satellite DNI data. Hence, we recommend further research using the Rockhampton terrestrially measured one-minute solar data (BoM 2012) to investigate this discrepancy.

We express some concerns about yield calculated from satellite DNI data (BoM 2013) prior to 2007 and suggest alternative methods, such as, using the terrestrial based one-minute solar data (BoM 2012). We acknowledge that the report may underreport the yield from the proposed plant because the report's TMY development period of 2007-13 has a La Niña bias. Sections 7.2, 7.3 and 7.7 discuss extending the study using one-minute data that can both overcome the La Niña bias and sparseness of satellite DNI data prior to 2007.

The overarching research question is:

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

The results show that modelling the weather for the four drivers of yield at Collinsville from limited data provides higher predictive performance in yield modelling then using the more complete data from Rockhampton.

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 requires further research.

7.2   Using BoM's one-minute solar dataset for Rockhampton site comparison

Sections 5.6 and 5.7 discuss the comparative analysis of the yield calculations between Collinsville and Rockhampton. For the comparative analysis, Shah, Yan and Saha (2014b) calculate yield for the Rockhampton comparison site using datasets from two sources: for the period 1999 to 2012 they use the datasets from Exemplary Energy (2014) and for 2013 they use a TMY from the Green House Office downloadable from EnergyPlus (2014).

Using the one-minute solar data from the BoM (2012) provides a way to improve the Rockhampton yield calculations for comparative analysis. This one-minute solar data overcomes two shortcomings in the existing Rockhampton yield calculations for use in comparative analysis:

- Fragmentation of the Rockhampton yield analysis across two different sources; and
- Remove the requirement for any satellite to effective DNI adjustment.

The one-minute solar data for Rockhampton Aero data from the BoM (2012) starts in 1996 and data collection continues, which removes fragmentation of the Rockhampton yield calculations. BoM (2012) uses ground based weather station observation, which removes the need to perform satellite to effective DNI adjustment. In comparison, Exemplary Energy (2014) derives its DNI data using satellite data from BoM (2013).

7.3   Adjusting BoM's one-minute solar data using BoM's satellite data to model Collinsville's DNI

Section 5.2.1 discusses how the BoM (2013) satellite data has far fewer hourly satellite images prior to 2007, which reduces the accuracy of inter year comparisons prior to the year 2007 but still allows inter-site comparison because the same hours are missing for all sites.

Adjusting BoM (2012) one-minute solar data for Rockhampton Aero with an hourly inter-site satellite DNI ratio provides an alternative way to model effective DNI for the Collinsville site starting in 1999. Figure 3 illustrates the inter-site satellite DNI variability. Nevertheless, modelling a small adjustment to reflect differing atmospheric opacity would still be required between Allen's (2013) effective DNI data and the adjusted one-minute solar data (BoM 2012).

7.4   Climate change adjusted yield and demand forecasts

Section 2.5 discusses the effect of climate change on yield, using three GCMs to represent three cases: most likely temperature rise, hottest and coolest for the NEM's geographic area.

Using climate change induced change in DNI rather than change in temperature in selecting GCMs provides a way to improve this analysis because the main driver for yield is DNI rather than temperature. However, the local effect in Collinsville determines yield but the global effect in the NEM region determines demand and wholesale electricity prices. Therefore, this would require selecting GCMs for the three cases: most likely change in solar radiance, dullest and brightest for both the NEM region and for Collinsville. Comparing the GCM selection for Collinsville with the NEM region would indicate if any local climate change effect in Collinsville is running counter to the global effect in the NEM region.

7.5   The effects of weights in the neural networks on adj-R^2 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 and shows the effect is trivial for this 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 irrelevant because 10 weights are used throughout this report.

However, keeping the weights fixed at 10 reduces the possibility of fine-tuning the neural networks. Using alternative weighting regimes to improve predictive performance would entail some research into the consequences for the comparisons of AIC and adj-R^2 values from differing weight regimes.

7.6   Ensuring consistent cleaning regimes between LFR and DNI terrestrial measurement instrument

Section 2.3.2 discusses the dust-effect and the requirement to maintain consistent cleaning regimes between the LFR and DNI terrestrial measurement instrument to ensure the effective DNI is the same for both. This would require some collaboration between Allen (2013), who manages the terrestrial DNI measuring instrument, and Guan, Yu and Gurgenci (2014) who are researching the LFR mirror cleaning.

7.7   Increasing the number of years in the TMY selection process to average out the effects of the ENSO cycle on both yield and demand

Sections 2.4, 2.6.3, 5.6 and 5.7 discuss the ENSO cycle and the TMY's selection period and the implications for electricity demand, wholesale spot prices and yield. We find a La Niña bias in the reports' TMY section period 2007-13. This bias results in underestimating demand, wholesale spot prices, and the LFR yield. Together, these will underestimate the revenue of the plant. We recommend extending the TMY selection period to reduce ENSO effects causing an over or underestimation of yield for the lifetime of the plant.

7.8   The DNI's model's month variable as a latent variable for changes in cleaning regimes or the idiosyncrasies of a particular year

Sections 4.4.1 and 4.4.3 present respectively the yield for the calibration year 2013 and the yield for the TMY selected from the years 2007-13. The monthly distributions of these yields fail to follow the expected summer-winter cycle. This section investigates why the yield fails to follow the expected cycle using a comparative analysis between yield and DNI energy per area, the main driver for yield. We find three factors contributing to the unexpected non-cyclic yield:

- January and February are the months with the lowest DNI for the years 2007-13
- the nonlinearity of the LFR amplifies this effect
- the TMY selection process also amplifies the effect.

While these three factors can explain the non-cyclical behaviour, they fail to preclude the possibility that the variable *month* in the DNI model is acting as a latent variable for other factors.

Table 33 and Table 34 show the daily average satellite DNI from BoM (2013) and the proposed plant's yield for each month for the years 2007-13 in the left panel. BoM (2013) calculates these DNI readings from satellite imagery and we calculate the yield using SAM and this report's weather modelling of the four drivers for yield. Our weather modelling is calibrated using data from Allen (2013). The normalised percent daily average yield and DNI allows easy comparison between them. Both January and February are the lowest normalised percentage DNI (83% and 79%), which supports the lower yield forecast in the summer months.

Additionally, Section 4.4.7 shows considerable nonlinearity between DNI input and yield output in SAM (2014). This nonlinearity further attenuates the low DNI input signal into SAM (2014), which is consistent with the relatively lower normalised percent daily average yield of 79% and 69% in January and February compared to DNI of 83% and 79%.

The right panel in Table 33 and Table 34 shows the TMY selection process for both DNI and yield from the years 2007-13. Section 4.4.3 described the TMY selection process in detail. In the last column, Table 33 and Table 34 show the normalised daily average DNI and yield for the TMMs of the TMY. Comparing this TMY normalised percentage with the normalised percent daily average of each month shows the TMY selection process has amplified the percentage variation. For instance, the DNI percentage daily average for January DNI is 83% and the TMM is 76%.

While these three factors contribute to the unexpected non-cyclical annual yield, they fail to rule out latent variables within the variable *month* for the DNI model. We suspect that the variable *month* in the DNI model may contain latent variables for the idiosyncrasy of the DNI measuring instrument or the year 2013. Allen's (2013) data could contain latent variables that affect DNI measurement emanating from changes in cleaning regimes, nearby earthworks and months when instrument faults are more prevalent.

Cures for the latent variables in Allen's (2013) data include dropping the variable *month* in the DNI model or including data from the year 2014 to make the model fitting more robust. In addition, replacing the variable *month* with the more theoretically derived variable *altitude* could further ameliorate any concerns.

Table 33: Daily average BoM satellite DNI each month and TMY selection for Collinsville (Wh/m²)

	DNI daily average each month									Absolute difference from the monthly average								TMY		
	2007	2008	2009	2010	2011	2012	2013	Ave	percent	2007	2008	2009	2010	2011	2012	2013	Min	year	value	percent
Jan	4,060	3,756	1,853	2,972	5,821	5,693	7,531	4,527	83%	466	770	2,673	1,554	1,295	1,166	3,004	466	2007	4,060	76%
Feb	4,277	3,353	2,417	3,535	4,362	5,440	6,709	4,299	79%	22	945	1,882	764	63	1,141	2,410	22	2007	4,277	80%
Mar	5,692	4,527	6,537	4,102	2,423	3,261	7,143	4,812	89%	880	286	1,724	710	2,389	1,551	2,331	286	2008	4,527	84%
Apr	6,224	6,019	5,534	3,839	4,594	6,090	6,307	5,515	102%	709	504	19	1,676	922	575	792	19	2009	5,534	103%
May	4,688	5,663	4,187	5,677	5,741	4,709	5,831	5,214	96%	526	450	1,027	463	527	504	617	450	2008	5,663	106%
Jun	3,097	5,076	6,168	5,436	4,978	4,574	6,915	5,178	95%	2,081	102	991	258	200	604	1,737	102	2008	5,076	95%
Jul	6,786	4,939	6,528	4,221	6,245	4,255	5,950	5,561	102%	1,225	621	967	1,340	684	1,305	389	389	2008	4,939	92%
Aug	4,742	5,953	6,789	4,449	5,019	5,768	8,465	5,884	108%	1,142	70	906	1,434	865	115	2,581	70	2008	5,953	111%
Sep	6,783	6,006	6,917	3,709	6,748	5,838	8,636	6,377	118%	406	371	540	2,668	371	538	2,259	371	2009	6,917	129%
Oct	6,900	6,367	7,212	5,064	5,339	6,728	8,388	6,571	121%	329	204	641	1,507	1,232	157	1,817	157	2012	6,728	125%
Nov	4,637	5,942	3,871	1,992	5,953	7,183	6,356	5,134	95%	496	809	1,263	3,142	820	2,050	1,223	496	2007	4,637	86%
Dec	5,046	6,062	6,421	3,148	4,640	7,085	9,205	5,944	110%	898	118	477	2,796	1,304	1,141	3,262	118	2008	6,062	113%
Ave	5,253	5,316	5,391	4,019	5,157	5,549	7,294	5,426	100%	173	109	35	1,407	268	123	1,868	35	2009	5,364	100%

(Source: BoM 2013)

Table 34: Daily average yield each month and TMY selection for the LFR at Collinsville (MWh)

	Yield daily average each month									Absolute difference from the monthly average								TMY		
	2007	2008	2009	2010	2011	2012	2013	Ave	percent	2007	2008	2009	2010	2011	2012	2013	Min	year	value	percent
Jan	55	57	14	33	103	107	144	73	79%	18	16	59	40	30	33	70	16	2008	57	61%
Feb	68	47	38	36	64	87	110	64	69%	4	17	26	28	0	22	46	0	2011	64	69%
Mar	92	70	123	41	19	44	116	72	78%	20	2	51	31	53	28	44	2	2008	70	75%
Apr	98	110	98	41	69	97	96	87	94%	11	23	11	46	18	10	9	9	2012	97	103%
May	45	70	38	82	93	67	67	66	71%	21	4	28	16	27	1	1	1	2012	67	71%
Jun	38	50	80	65	67	53	84	62	67%	25	12	18	2	5	10	22	2	2010	65	69%
Jul	108	68	99	41	94	53	61	75	81%	33	7	24	34	19	22	14	7	2008	68	73%
Aug	71	109	134	76	82	102	172	106	114%	36	3	28	31	25	5	65	3	2008	109	117%
Sep	153	126	159	55	153	122	188	137	147%	16	11	22	81	17	15	52	11	2008	126	134%
Oct	165	136	177	105	114	170	203	153	164%	12	16	24	48	39	17	50	12	2007	165	177%
Nov	82	135	94	21	118	172	130	107	115%	26	27	13	86	10	65	22	10	2011	118	126%
Dec	85	116	124	38	75	155	193	112	121%	27	4	11	74	37	43	81	4	2008	116	124%
Ave	88	91	99	53	88	102	131	93	100%	5	2	6	40	5	9	38	2	2008	93	100%

Acknowledgements

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 2013, 'The impact of climate change on electricity demand: Research', in WP Bell (ed.), *Analysis of institutional adaptability to redress electricity infrastructure vulnerability due to climate change*, National Climate Change Adaptation Research Facility, Gold Coast, Queensland, Australia, pp. 51-84.

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/>.

—— 2012, 'About One Minute Solar Data', viewed 31 Jul 2014, <http://www.bom.gov.au/climate/data/oneminsolar/about-IDCJAC0022.shtml>.

—— 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, 'Climate statistics for Australian locations: Period 1981-2010: COLLINSVILLE POST OFFICE', viewed 5 Jun 20014, <http://www.bom.gov.au/jsp/ncc/cdio/cvg/av?p_stn_num=033013&p_prim_element_index=0

&p_comp_element_index=0&redraw=null&p_display_type=full_statistics_table&normals_yea rs=1981-2010&tablesizebutt=normal>.

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

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

—— 2014e, *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>.

EnergyPlus 2014, *Weather Data: All Regions : Southwest Pacific WMO Region 5 : Australia* <http://apps1.eere.energy.gov/buildings/energyplus/cfm/weather_data3.cfm/region=5_south west_pacific_wmo_region_5/country=AUS/cname=Australia#instructions>.

Exemplary Energy 2013, *Weather, Climate and Solar Data for Australian Locations*, Exemplary Energy, Fyshwick, ACT, Australia.

—— 2014, *Hourly Climate Data*, <http://www.exemplary.com.au/download/Flyer%20EXEMPLARY%20CLIMATE_DATA%202 013.pdf>.

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>.

SAM 2014, *System Advisor Model Version 2014.1.14*, National Renewable Energy Laboratory, Golden, CO, USA, <https://sam.nrel.gov/content/downloads>.

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 (Draft)*, 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