Norwegian University
of Life Sciences

Master's Thesis 2020 30 ECTS
Realtek

Monitoring PV Systems: A Recommendation for Appropriate Measurement to Improve System Longevity

Lars Jetmund Svartis Engesæth
Industrial Economics and Technology Management

Abstract

Solar photovoltaic (PV) systems demand improved operation and maintenance (O&M) measures to increase system longevity and overall performance. Installation of solar PV systems are rapidly increasing, and while technology is improving, O&M measures are neglected or poorly understood. Appropriate O&M will improve solar PV market competitiveness. This study aims to design requirements for operators to provide necessary information to perform an analysis of the state of the system. Building on existing literature for monitoring PV system, it asks: Which parameters and indicators are necessary to evaluate the state of the system, which quality assessments must be made to produce reliable results, and what must operators do the first year to aid this process?

Based on literature regarding monitoring and maintenance of solar PV, an analysis was performed on the operational data set from the PV system at Glava Energy Center in Glava, Sweden. The methodology was based on CRISP-DM and was exploratory. Data were analysed, enhanced, and analysed further in the context of Overall Equipment Effectiveness. The approach was adjusted as understanding grew. The results provide insights into the parameters, indicators and the general quality of data. Temperature and plane-of-array irradiance is suggested as the most influential parameters to evaluate system projections, although the parameters should be accompanied by other measurements. There are also presented an analysis of the quality of the data set. The quality is periodically low, which leads to recommendations regarding maintaining quality of monitoring of the PV system.

Sammendrag

Solcelle fotovoltaiske systemer krever forbedrede tiltak for drift- og vedlikehold (O&M) for å øke systemets levetid og generelle ytelse. Installasjonen av solcelleanlegg øker raskt, og mens teknologien forbedrer seg, er O&M-tiltak neglisjert eller dårlig forstått. Passende O&M vil forbedre konkurranseevnen til PV. Denne studien tar sikte på å utforme krav til operatører til å gi nødvendig informasjon for å utføre en analyse av systemets tilstand. På bakgrunn av eksisterende litteratur for overvåking av PV-system spør den: Hvilke parametere og indikatorer er nødvendige for å evaluere tilstanden til systemet, hvilke kvalitetsvurderinger som må gjøres for å gi pålitelige resultater, og hva må operatørene gjøre det første året for å hjelpe denne prosessen?

Basert på litteratur om overvåking og vedlikehold av solcelleanlegg, ble det utført en analyse av driftsdatasettet fra PV-systemet ved Glava Energy Center i Glava, Sverige. Metodikken var basert på CRISP-DM og var utforskende. Data ble analysert, forbedret og analysert videre. Tilnærmingen ble justert etter hvert som forståelsen vokste. Resultatene gir innsikt i parameterne, indikatorene og den generelle kvaliteten på data. Temperatur og plan-av-array-irradians foreslås som de mest innflytelsesrike parameterne for å evaluere systemframskrivninger, selv om parameterne bør ledsages av andre målinger. Det blir også presentert en analyse av kvaliteten på datasettet. Kvaliteten er periodisk lav, noe som fører til anbefalinger om opprettholdelse av kvaliteten på overvåkingen av PV-systemet.

Foreword

This thesis is the final part of a five-year study in industrial economics at NMBU (Norwegian University of Life Sciences). The thesis equates 30 ECTS and have been written in the spring of 2020. The thesis proved to be incredibly challenging and rewarding, and I am proud to present it. The research was done with close cooperation with PhD candidate Jesper Frausig, and I will trust our results to him for further research into the topic.

The thesis would never have been possible without the incredible help and counselling from my advisors, Associate Professor Tor Kristian Stevik and PhD candidate Jesper Frausig. Tor Kristian proved invaluable as the strategic mastermind behind the thesis, while Jesper and I shared countless interesting discussions regarding our results and solar technology. I could not express the extent of my gratitude to them in this paragraph alone.

I am also grateful for the help and support of my friends and family, especially the support I needed when it all seemed impossible and difficult. Additionally, I wish to thank my fellow students and professors at NMBU for the great experiences throughout these five years.

Finally, this thesis was also made possible by Glava Energy Center. They provided the necessary material to perform a meaningful analysis and to make a conclusion in this paper. Without their help, this paper would not be as meaningful as it is. They have my gratitude, and I wish them the best of luck with their own research.

I was lucky enough to live in good health and safety during the COVID-19 outbreak, but unlucky enough to write my thesis during it. For that reason, I would like to dedicate the end of this foreword to a quote from an important student song for NMBU students.

«Kanskje verden er litt stri,
men når det gråner skal du si:
at du har hatt en bra studentertid!»
- *"Studentenes Kall" 1933*

Ås, June 2020

Lars Jetmund Svartis Engesæth

Contents

1. Introduction

1.1. Background

A transition to renewable energies is necessary to reach the global 2-degree target, as proposed by the Paris Agreement. Globally, 65% of electricity is produced by burning coal, oil and fossil gas (International Energy Agency 2019). A large portion of the carbon emissions come from burning fossil fuels for electricity. Solar power could prove to be an important technology for energy production with low carbon emissions per MWh.

Yearly solar flux exceeds yearly energy demand by four orders of magnitude (Hofstad 2019). The worlds combined energy consumption in 2018 was 161 248 TWh (BP plc 2019), while the sun provides 89 000 TW annually across the globe (Tsao and Crabtree 2006). This means that the sun can power a year of energy consumption in less than 2 hours. All this solar flux cannot, of course, be converted to useful electricity. After considering the suitable surface area for solar PV and assuming a high 40% efficient energy conversion (Tsao and Crabtree 2006) (although the highest recorded efficiency is 46% (Geisz, Steiner et al. 2017)) the technical potential equates to roughly 7500 TW. Harvesting all technically possible solar flux would then meet the world's yearly energy demand in 22 hours. There are therefore clear indications that the sun could meet a large part of the global energy demand.

Over the past 10 years, solar photovoltaic (PV) capacity has been increased by a factor of 28.55. In 2018 alone, there were installed 136 additional TWh, increasing the total capacity by 31% (Heymi Bahar, D'Ambrosio et al. 2019). This surge of new installed capacity comes with a greater need for operation and maintenance (O&M). Proper O&M mitigates potential risks, improves levelized cost of electricity (LCOE) and Power Price Agreements (PPA) prices and positively impacts the return on investment (ROI). The O&M phase of the PV project is also by far the longest phase, typically lasting for more than 30 years (SolarPower Europe 2019). It is therefore crucial that O&M is studied and improved. Longer lifetime positively impacts the LCOE, since a system with a longer lifetime will produce more total power than a shorter-lived system.

Prices for solar PV capacity has decreased by 85% to 57 USD/MWh, during a ten year period from 2009 to 2019 (Jäger-Waldau 2019). With steadily decreasing prices per MWh, the market competitiveness of solar PV is growing, and solar energy represented 42.5% of all new renewable energy investments in 2018 (Jäger-Waldau 2019). 2018 was also the 9th year in which solar power represented the largest share of new investments (Jäger-Waldau 2019).

Improved knowledge regarding O&M for solar modules will lead to a higher ROI. One study from Spain reveals that a reduction of O&M tasks by 76% resulted in a 26% reduction in energy production (Muñoz-Cerón, Lomas et al. 2018). The reason for the deviation is not yet understood, but there is clear indications that effective O&M results in decreased LCOE (Muñoz-Cerón, Lomas et al. 2018). There must be knowledge regarding the state of the system to perform suitable O&M.

1.1.1. Scope and research questions

The complexity of O&M for Solar PV system is challenging, as failure appears due to different causes in different climates. The failures also have different consequences in different climates, as well as climate-dependent cascade-effects. For that reason, this paper will limit the scope to focus on these OEE parameters:

- *Performance – the usefulness of the system when it is operating.*
- *Uptime – The availability of the system, when required.*
- *Quality – The ability to provide useful function.*

The goal for this thesis is that it should design instructions for operators so that the operators produce a basis for evaluating the state of the solar PV system.

The research questions related to that goal are as follows:

1. Which indicators and parameters are necessary to perform said evaluation of the state of the system?
2. Which demands are there to quality of data, monitoring and analysis?
3. What sort of O&M must be performed the first year to make this evaluation?

There should be attention to understanding O&M for solar PV systems, but there must be insights into the failures and deviations that degrade PV systems. To understand failures and deviations within the system an analysis of the state of the solar PV system must be performed, to determine a baseline.

The motivation for the approach is largely determined by the context. The thesis is preliminary work for the project that PhD candidate Jesper Frausig is performing. In that context, it is relevant to analyse the failure modes for PV systems. Therefore, there will be attention to degradation nodes for different climates as presented below.

1.1.1.1. Performance indicators

The primary OEE parameter influencing lifetime is performance. One example of a performance issue for PV systems is ribbon degradation. Ribbon degradation is uniquely manifesting as failure clusters in polar and tropical environments. It manifests differently in the two climates, and the review ought to establish when the climate-specific degradation pathway indicates a performance issue.

1.1.1.2. Uptime indicators

The secondary parameter affecting the PV lifetime is uptime or the availability of the system, when required. An example of an uptime indicator is physical damage of the cell. Physical damage is uniquely manifesting as a failure cluster in snow climates and moderate climates. It manifests differently in each climate, and the review ought to establish when the climate-specific degradation pathway indicates an uptime issue.

1.1.1.3. Quality indicators

The third parameter supporting performance and uptime is quality. Quality is defined as the ability to provide useful function. One example of a quality issue is backsheet damage. Backsheet damage is uniquely manifesting as a failure cluster in polar, arid and moderate environments. It manifests differently in each climate, and the review ought to establish when the climate-specific degradation pathway indicates a quality issue.

1.1.2. Outline

The thesis begins with the summary of the main findings.

A theoretical basis is presented to give the reader the necessary understanding and common agreement of core aspects that will be discussed in the thesis.

Then the methodology is presented. It is assumed that the reader knows why certain methodology is used in different circumstances. This chapter will aim to explain and defend why the chosen methodology will give credibility to the results.

After the methodology follows the results from the analysis. They will give insight into what was discovered and provide examples for the reader. There are numerous other examples that have been excluded, but it was concluded that including more examples would not provide any additional and meaningful information. They will be used further by PhD candidate Jesper Frausig in his research.

Subsequently follows a discussion of the results and their implication. The discussion aims to answer the research questions and will highlight several relevant aspects. This chapter will also discuss several topics on the periphery of the scope of this paper that are important to the topic and the results.

The thesis will end with some concluding words to summarize the findings and the most important parts from the discussion. This section provides answers to the research questions posed in the introduction. Finally, this chapter ends with some suggestions for future work that have been considered relevant after the process of writing this thesis.

After the conclusion to the thesis follows the addendum with attachments and the reference list used for researching the topic.

2. Theoretical basis

2.1. Solar PV basics

The module and cells compromising a solar module are commonly referred to as solar a module and solar cells. For the purpose of disambiguation from other types of solar technology, the term photovoltaics will be used. Photovoltaics (PV) cells are made using semiconductors. Semiconductors are insulators in their pure form but will conduct electricity under certain temperatures or when they are mixed with other materials. PV cells are generally built using p-type and n-type doped silicon wafers. When the sides of a silicon wafer are oppositely doped a layer between them called the depletion region will be produced. As a few electrons move from the n-side to the p-side, a bandgap of negative charge preventing more electrons from moving that way is formed. PV cells utilize solar radiation to excite electrons to pass the bandgap, creating an asymmetric abundance of electrons. When these two sides are connected through a cable and a resistance, electricity will flow.

2.1.1. Operational Equipment Effectiveness as applied to PV

Overall equipment effectiveness (OEE) is a measure of how well a system manufactures its products during the period it is supposed to operate (Vorne Company s.a.). By analysing quality, performance, and uptime, insights into how the process can be optimized are possible. For PV this means analysing quality of equipment, while performance refers to the output of the system, and uptime refers to the availability of the system and what percent of expected uptime where the system produces. Due to the nature of solar power, there will be periods where weather will reduce the output of the PV system. Without proper monitoring, it can be difficult or impossible to determine if any reduction in output is caused by system degradation, monitoring errors, or simply weather.

Some studies have shown that using risk analysis and the Risk Priority Number (RPN) can help to identify which degradation and failures are most critical (Colli 2015), although more research into the method is still necessary. One article showed that degradation, static shading and variable shading can be detected successfully from analysing voltage and current of the PV system (Pei and Hao 2019).

2.1.2. Overview of select maintenance issues

One challenge for O&M for PV modules is that PV modules degrade only 0.8-1% annually (Jordan and Kurtz 2012) (National Rewewable Energy Laboratory 2016) (Azizi, Logerais et al. 2018). This means that due to uncertainty in measurements, lack of attention, or low quality of data, several years may pass without detection of failures or degradation (Pearsall 2017).

When degradation is confirmed within an uncertainty interval, it can be challenging to identify which degradation cause(s) are responsible for the error. Damaged and otherwise impaired parts can have both synergetic and antagonistic effects on other parts of the PV module, which increases the challenge of correlating observations with only one cause. The microclimate in which the PV system is located will, however, indicate which causes are more likely to occur given the conditions.

Additionally, it appears results are skewed for degradation rates. PV modules with higher degradation rates have a tendency to be left shorter in operation, and thus gives the impression of sinking degradation rates as systems age (Jordan and Kurtz 2012).

It can be useful to view the degradation of PV modules as aging. As the PV module ages, the OEE will decrease accordingly. Factors that contributes to aging that can be hindered or reduced are highly relevant for O&M operators to register and evaluate. One study showed that the resistance in a module increased by 12.8% in 20 years, which reduced the power output by 30% (Azizi, Logerais et al. 2018).

The following subchapters will explore three specific cases of PV module degradation: ribbon degradation, cell physical damage and backsheet damage.

2.1.2.1. Ribbon degradation

Cell interconnect ribbon degradation reduces the power transmitted from the solar cell to the bus bar. When moisture leaks through the backsheet of the solar module, the ribbon will corrode, which increases the series resistance (Chattopadhyay, Dubey et al. 2014). Increased resistance results in lower power output as more power is converted into heat inside the solar module. Overheated PV modules can cause cascading effects of degradation. Since the initial quality of a part in a solar cell or module will vary, degradation can lead to varying consequences in each PV module, especially accompanied by electrical mismatching.

The ribbon located on top of the solar cells carry current from each string to the PV bus bar. When the series resistance increase, the power transmitted from solar cells will decrease, and consequently, leads to a reduction of OEE.

Corrosion of the cell's interconnect ribbon (and in their entirety), is sped up by corrosive products from the encapsulation. The ethylene vinyl acetate (EVA) film protects the solar cell, but will decompose into acetic acid that increases corrosion and degrade the metal of which the interconnect ribbon is made of (van Dyk, Chamel et al. 2004) (Kempe, Jorgensen et al. 2006). Van Dyk et al. also showed that during the drier summer months, there was less moisture inside the solar module, which resulted in improved performance (van Dyk, Chamel et al. 2004). Additionally, Kempe et al. discovered that solar modules in lower than *-15 ℃* will be more vulnerable to mechanical damage from snow and wind than similar solar modules in other climates (Kempe, Jorgensen et al. 2006).

2.1.2.2. Cell physical damage

Cell physical damage occurs in snow climates due to large amounts of snow on the PV module. When snow and ice loads are unevenly distributed across the module, the stress can bend the frame, break the glass and loosen screws (International Energy Agency 2014). The damage can be characterised as the following four effects:

1. Vertical loads from snow and ice on an inclined surface can be broken down into two component forces: The normal force and the downhill force.
2. Sliding snow on the surface of the module is distributed inhomogeneously.
3. Inhomogeneous loads cause moments and torques in the lower part of the module along the axial direction of the test specimen.
4. Temperatures below *0 ℃* may cause embrittlement of the adhesives and further reduce stability.

The extent to which cell cracks will influence power output directly is poorly understood. (Köntges, Kajari-Schröder et al. 2011) and (Köntges, Oreski et al. 2017) suggest that the orientation of a cell crack is a critical factor in determining the reduction in power output. The two papers also present illustrations to show how cell cracks were unevenly distributed across the module and the resulting uneven distributed power loss.

Additionally, isolated cell cracks will decrease the maximum power point. The combination of these sources of uncertainty means that the power loss of the module will be different from the sum of all losses from cracked cells (Köntges, Oreski et al. 2017). Nevertheless, cracks may lead to degradation from humidity, regardless of the total power loss.

Furthermore, it is suggested that PV modules located in cold and polar climates (D and E in Köppen-Geiger classification (Peel, Finlayson et al. 2007)) are more vulnerable to cell damage. The degradation rate of 7% annually is higher for those climates as opposed to the average degradation rate from cell physical damage in other climates (Köntges, Oreski et al. 2017). On the other hand, the same paper established that Köppen-Gieger classification works poorly to predict the failure modes in different climates, possibly since Köppen-Geiger classification is meant to classify plant growth under different temperature and precipitation conditions, and not weather affecting PV modules (Köntges, Oreski et al. 2017). Köppen-Gieger classification will defined below.

2.1.2.3. Backsheet damage

PV module backsheets normally consists of three-layer laminates:

- A weather resistant outer layer
- An electrical resistant middle layer
- An adhesive inner layer.

Degradation can occur as a result of exposure to light, heat, moisture, and other environmental stressors.

This damage can manifest as physical cracks in the backsheet, which allows moisture and oxygen a pathway into the solar cells. Cracks leads to degradation of the packaging and solar cells. Lyu et al. showed that the inner layer is the layer most affected by degradation (Lyu, Fairbrother et al. 2020). There is also indications from Lyu et al. that light, heat, and moisture facilitated cracking of the backsheet (Lyu, Fairbrother et al. 2020). These effects are highly dependent on the microclimate in which the module is placed, which could make it hard to generalize backsheet damage from one PV system to another, based solely on their macroclimate.

Several studies show the material that the backsheet consists of can be a significant contributor to the rate of degradation of the backsheet (Wiesmeier, Huelsmann et al. 2012) (de Oliveira, Cardoso et al. 2017) (Gebhardt, Bauermann et al. 2018). This effect further complicates the task of generalizing our understanding of degradation mechanisms in the backsheet of PV modules.

2.2. Relevant climates

The climates in which the systems are located will impact which degradation modes are relevant to consider. Similar degradation in different climates may have different causes, development, and impact, so the results given for one degradation mode in one climate cannot be transferred to another climate without further research. In this chapter, there will be attention to the polar and snow climates as classified by Wladimir Köppen (1846-1940) and Rudolf Geiger (1894-1981). The Köppen-Gieger classification breaks down climates into five groups (Kottek, Grieser et al. 2006):

A	B	C	D	E
• Tropical	• Arid	• Warm	• Cold	• Polar

Initially, these categories were created by Wladimir Köppen based on his experience from botany and thus were used to analyse ecosystem conditions and vegetation. Later, the system was improved by Rudolf Gieger. The system has since its creation been used in a wide range of climate, climate change, physical geography, hydrology, agriculture, and biology research (Kottek, Grieser et al. 2006).

While these classifications can clearly determine attributes of the macroclimate, the microclimate in which the solar modules are located is necessary to understand. Usually, microclimates only differ slightly from the surrounding area. These small differences can lead to significant changes in solar modules, especially since degradation modes are dependent on different causes than minimum and maximum temperatures and precipitation.

2.2.1. Polar climate

The polar climate (E) is commonly found near the poles.

Climate	Condition (Kottek, Grieser et al. 2006):
All polar climates	$T_{max} < 10$ ℃ in the warmest month
Tundra climate	0 ℃ $\leq T_{max} < 10$ ℃
Frost climate	$T_{max} < 0$ ℃.

Due to this temperature criterion, these climates will be largely determined by permafrost and/or snow (if precipitation allows for snow cover). This can lead to unique effects on the solar module, since these subclassification are the climates with the lowest temperatures. Considering the climate-specific factors with respect to the degradation types listed above, it is worth noting that:

- Some metals are brittle in lower temperatures. This might damage or weaken the solar modules.
- Low temperatures allow the module to work more efficiently. Solar modules in polar climates might have higher efficiency.

Examples of Polar climate include, but are not limited to, Svalbard, Norway; Mount Everest, China/Nepal; and Nord, Greenland.

2.2.2. Cold climate

Cold climates, also referred to as continental climates, snow climates, or microthermal climates, are defined by the mean temperature of the coldest month being below -3 ℃. Cold climates have more variations within the climates, deemed by the timing of the precipitation:

Climate	Condition (Kottek, Grieser et al. 2006):
All cold climates	$T_{mean} < -3$ ℃ in the coldest month
Snow climate with dry summer	$P_{smin} < P_{wmin}$, $P_{wmax} > 3P_{smin}$ and $P_{smin} < 40$ *mm*
Snow climate with dry winter	$P_{wmin} < P_{smin}$ and $P_{smax} > 10\ P_{wmin}$
Snow climate, fully humid	Neither of the two options above

P_{smin}, P_{wmin}, P_{wmax}, P_{smax} are the highest and lowest monthly precipitation values in the summer(s) months and winter(w) months (Kottek, Grieser et al. 2006). Examples of Cold climate include Oslo, Norway; Moscow, Russia; Pyongyang, North Korea; and Toronto, Canada. Interesting weather phenomena include, but are not limited to:

- Snow can remain for periods of time, causing low availability, while also partly melting. This can lead to an uneven mechanical stress on the PV modules, eventually lowering the performance, and even quality of the system.
- Seasonal humidity can cause differences in stressors for modules. It could prove to be different degradation between dry summers and wet summers for PV modules.

2.3. Solar PV system indicators and parameters

This section will introduce the indicators and parameters that have been relevant in this thesis to lay the foundation for shared definitions and understanding.

2.3.1. Irradiance parameters

In this thesis there will be results and discussions surrounding irradiance. The terms DHI, DNI, GHI, POA30, POA40 will be used to describe different types of ways of measuring irradiance.

DHI (diffuse horizontal irradiance) refers to the irradiance received per unit area by the modules or measuring equipment, that has not travelled a direct path from the sun after being scattered by molecules in the atmosphere. The irradiance arrives in equal amount from all directions (Vashishtha 2012).

DNI (Direct normal irradiance) is the irradiance received by the module or measuring equipment at a normal orientation to the irradiance that travels directly from the sun to the module. Since the orientation of modules are static, and the sun moves, this irradiance is multiplied by the cosines angle (Vashishtha 2012).

GHI (Global horizontal irradiance) is the total irradiance received by a surface that is horizontal to the ground. It is the sum of DNI (cosine adjusted) and DHI (Vashishtha 2012).

POA30 and **POA40** (plane-of-array) refers to the irradiance coming from directly from the sun multiplied by the angle of incidence (30 or 40 in this paper), diffuse irradiance from surroundings and reflected irradiance, resembling the amount irradiance incident on the plane of array of the PV system. The reflected irradiance is determined by several factors, like the module tilt angle and soil reflectance factor (EcoSmart s.a.).

Clearsky conditions is the conditions during a day where there are no clouds covering sky and casting shadows on the solar module. The irradiance resembles a bell curve and can help analyse measured irradiance against expected irradiance (Silva, Balanzategui et al. 2019) (Hatti 2014).

Figure 1: The image displays the sensor station at GEC. Photo: Jesper Frausig.

2.3.2. Module temperature

The module operating temperature is the second most influential parameter aside from irradiance when considering module performance (Pearsall 2017). The module temperature determines the rate of reaction of degradation, and by controlling temperature, one can improve the lifetime of the module. The parameter is a function of the following variables (Pearsall 2017):

- The irradiance received
- The ambient temperature
- The module design, regarding how it rejects heat
- The module efficiency, since a lower efficiency means that more irradiance is converted to heat
- The mounting system, regarding how ventilation affects the module
- Other ambient conditions, such as wind speed and direction

Generally, module temperature is proportional with irradiance received. Higher module temperature leads higher cell temperature, which results in lower efficiency (Pearsall 2017). This means that with higher irradiance, there will be lower performance. Additionally, there will be slight variations between cell temperature, module temperature and ambient temperature.

2.3.3. Power output

Power output will be possible to register as alternating current (AC) or direct current (DC). AC can be divided into real, reactive, and apparent power. Real power is measured by Watts, Reactive power is measured in Volt-Amps-Reactive (VAR), and apparent power is measured in Volt-Amps (VA). Apparent power is a complex value and is the vector sum of the real power and the reactive power (Engineering ToolBox 2005). Power measurements in this thesis will be based on apparent power.

2.4. Cloud focusing

Cloud focusing (or cloud enhancement) is a phenomenon where days with partly cloud cover can experience higher measurements of irradiance higher than clearsky estimations. The phenomenon is generally explained by reflection through cloud edges to the solar modules, but Järvelä et al. argue that this explanation is insufficient (Järvelä, Lappalainen et al. 2020). Their results indicate that cloud focusing can result in irradiance 1.5 times higher than clearsky irradiance (Järvelä, Lappalainen et al. 2020). Zehner et al. found that irradiance may reach up to 30% higher than clearsky (Zehner, Weigl et al. 2011). Generally, cloud focusing is a phenomenon that is not yet completely understood, but it may affect some measurements in this paper.

2.5. Definition of deviations

For the purpose of this text, and general definitions within the PV discipline, the following terms will be defined accordingly. In terms of solar PV, several definitions are used to describe the concept of problems regarding the PV system. Degradation, fault, malfunction, error, defect and failure are all definitions that have somewhat overlapping meaning in terms of deviation of a system's performance. This paper will focus on degradation and failure of PV systems.

2.5.1. Differentiation between deviation types: failure and degradation

Degradation of a system is a result of a malfunction and refers to the detected reduction of the system's ability to perform normally (Parhami 1997). A degraded system is only partially faulty. When a system has a failure, it cannot meet expectations and/or is unable to perform properly. A failure can happen independent of degradation, but a degradation in general will eventually lead to a failure. Neglecting to attend to the issues may later result in a catastrophic failure. From a warranty perspective, a module is considered failed when its power output is less than 80% of its original power output (Hatti 2014). However, even if PV modules fulfil their technical specifications, the expected performance may not satisfy economic requirements of the business case.

Since PV systems have no moving parts, the main cause of reduced reliability is instability and corrosion of the individual parts (Pearsall 2017). Degradation will slowly reduce the output of individual parts, until the degradation is noticeable. To prevent cascading degradation effects, discovering degradation as early as possible and rectify critical problems immediately is important.

2.6. Diagnostics, detection, identification and monitoring

2.6.1. Monitoring

The ways of monitoring solar power plants are several, but analysis of the data and measurements must be conducted by someone with expertise within the field (Hatti 2014).

Furthermore, Hatti raises the following perspectives regarding monitoring of systems (Hatti 2014):

- Inverters can be used to monitor the performance of the system
- Measurement of radiation is essential to ensure precise control of the performance of a facility. Both on-ground measurements and satellite measurements are possible. This thesis used a combination of these methods.

Solar PV systems have no moving parts, unless there is solar tracking installed. This means that it may not be easy to distinguish which part has broken down when the performance declines or the system fails (Pearsall 2017). Pearsall argues that the most relevant input parameters for a solar PV system is the irradiance it receives and the temperature of the module. Furthermore, Pearsall argues that measurement of electrical output together with other parameters will be essential to perform meaningful analysis of the performance of the system (Pearsall 2017).

Pearsall provided a table with common faults that could be identified in a system that has reduced output. The faults mentioned in the table is as follows (Pearsall 2017):

1. Inverter threshold
2. System outages
3. Shading
4. Poor MPPT (maximum power point tracking) behaviour
5. Grid voltage fluctuations
6. Poor inverter efficiency at low light levels
7. Inverter output plateau
8. Temperature effects

An example of these could be inverter power plateau. The table describes that the AC output will be the parameter that indicates this issue, and that the data should be plotted for a day with clearsky conditions and high sunlight levels. If inverter power plateau is indeed the issue, the AC should follow the irradiance most of the day, but will be restricted to a level that is lower than the expectation (Pearsall 2017). The other faults in the list are similarly explained. A comprehensive understanding of all such tests would simplify O&M and make monitoring of performance easier.

2.6.2. Health and safety

Health and safety of the workers are a fundamental part of performing O&M of any power plant. Solar power is no exception, as workers can be exposed to great heights, large voltages and hot surfaces. Hatti goes into detail of proper precautions regarding O&M and suggest that O&M should only be performed by qualified and trained personnel (Hatti 2014). This paper will focus more on O&M when applied to diagnostics, detection, and monitoring, and will not focus on the specific and necessary measures that should be taken when investigating and working near a solar PV plant.

2.7. CRISP-DM

In this thesis there has been used an approach from data science called *Cross-industry standard process for data mining* (CRISP-DM). The methodology is divided these six phases (Vorhies 2016):

1. **Business Understanding:** determine business objectives; assess situation; determine data mining goals; produce project plan
2. **Data Understanding:** collect initial data; describe data; explore data; verify data quality
3. **Data Preparation** (generally, the most time-consuming phase): select data; clean data; construct data; integrate data; format data
4. **Modelling:** select modelling technique; generate test design; build model; assess model
5. **Evaluation:** evaluate results; review process; determine next steps
6. **Deployment:** plan deployment; plan monitoring and maintenance; produce final report; review project

CRISP-DM is an intuitive process for understanding a data set by having a circular approach to the problem, and is the most widely-used analytics process standard (Brown 2015). The list above fails to explain that the process is also circular and does not necessarily only involve six steps. The steps are all parts of an cycle of activities, that might demand that the researchers work back and forth between the phases (Brown 2015).

2.8. O&M for solar PV

Several measures can be taken in order to perform O&M on solar power plants. Mustapha Hatti published a detailed book describing, among several issues, the details for annual inspection, detailed visual inspection and manufacturer-specific inverter inspection (Hatti 2014).

The solar power plant must be regularly maintained. Without proper care, one or more parts is likely to break, which leads to reduced power output or a catastrophic failure. The amount of O&M necessary for each individual plant will be different, both due to random incidents occurring and the micro- and macroclimate specific stressors affecting the plant. Operators can independently choose how much maintenance should be planned based on the simplified list below:

1: NO O&M. The system is left alone

2: Complete O&M. Close monioring and inspection of parts to rectify any issue

3: A solution between option 1 and 2

Option 1 should appear undesirable for most operators. The costs of investment in PV systems are large, contingent on third-party loans, and the possible reduction of power output is documented from several plants. Therefore, it can be assumed that option 1 will lead to a higher LCOE in the long term, even if saving costs on O&M in the short-term.

Option 2 should keep the plant in the best shape the operators are capable of. This will not reduce the risk completely, as the list of issues that can arise is incomplete even given a full understanding of the underlying issues that might affect the plant. This risk is likely to decrease over time, as additional research provides better understanding of these issues. Since option 2 involves the most time spent on O&M, it is easy to assume that this will lead to the highest operating costs.

Option 3 will be any solution between option 1 and 2. It will be chosen under the assumption that there will be a too great cost of closely investigating the plant daily. There is not consensus regarding how much O&M is economically optimal. Tentative figures are 0.5-1% per year of investment costs; however, the actual costs are dependent on a range of factors. NREL published a study indicating that O&M costs per kW has been reduced in the 2010-2018 timespan (Fu, Feldman et al. 2018). O&M costs for residential, commercial and utility-scale PV declined by 60%, 47%, and 49%, respectively (Fu, Feldman et al. 2018). Reduced costs for O&M will result in reduced LCOE for solar power.

2.9. Quality of solar PV measurements

Systems of high quality monitoring will exceed general system monitoring requirements for PV systems (Silva, Balanzategui et al. 2019). The extent of requirements for quality measurement systems might be different for each system, depending on the intended use of the data.

Silva et al. displays an overview for quality control, where it is the result of a triangle consisting of quality assurance, quality assessment and quality enhancement (Silva, Balanzategui et al. 2019). The four concepts can be described as follows:

1. **Quality control**: The complete process through the other three concepts whereas the quality of data is maintained.
2. **Quality Assessment**: The process of flagging data points that are incorrect. The process also involves comparing the data to itself.
3. **Quality Assurance**: This concept is about maintaining a high level of quality in the measurement process to prevent corrupt data from appearing.
4. **Quality Enhancement**: This concept refers to the act of improving the data quality after considering quality assurance and quality assessment. This is the only process that changes the data set. Since it changes the data set of which the analysis is built upon, it can change the result from analysis. Therefore, it must be carried out by an expert, and it is advised that the original data set is saved in case the enhancement was incorrect.

Logical charts with branches can sometimes help visualize which leads to which measures. The following charts are visual representations inspired by literature regarding the subject.

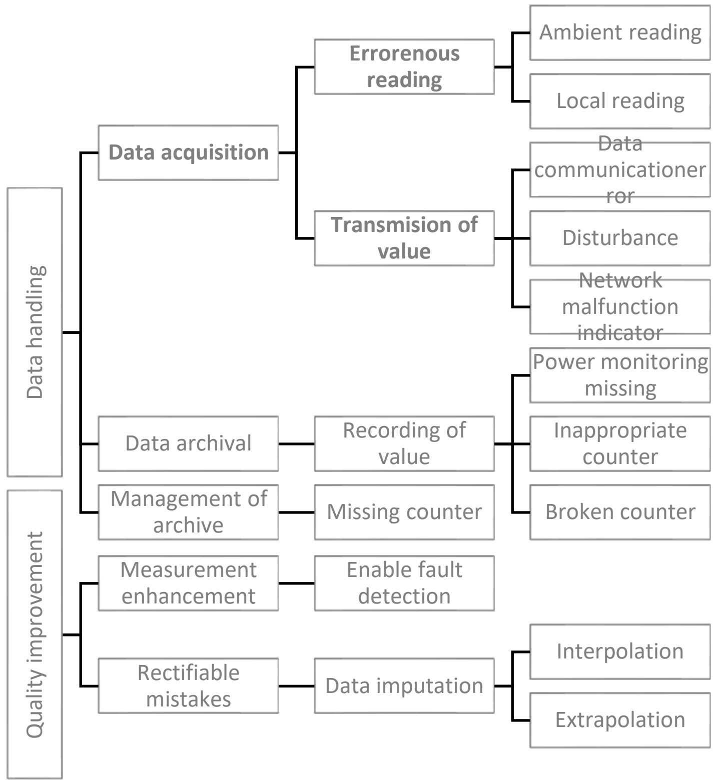


Figure 2: The chart indicates some logical connections for quality improvement and data handling. The figure is inspired by: (Silva, Balanzategui et al. 2019), NREL Quality Management Handbook and Solar Bankability Project.

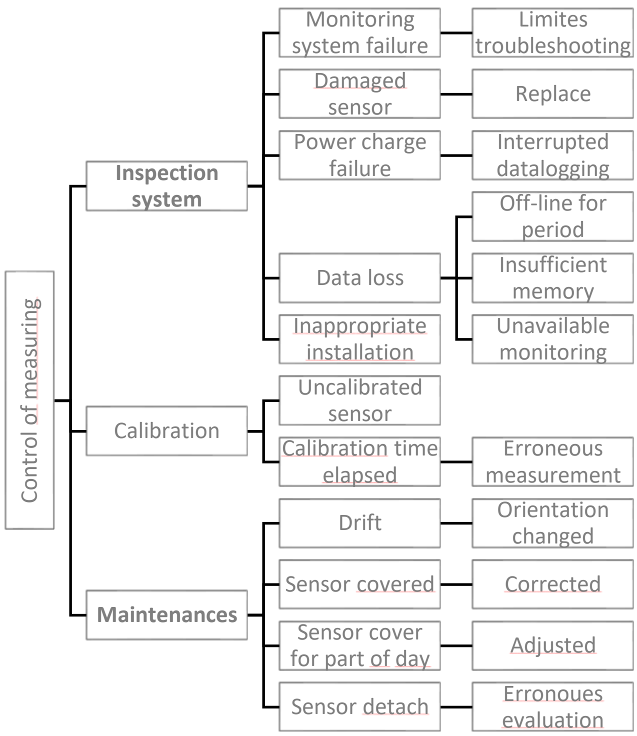


Figure 3: The chart indicates some logical connections for control of measurements. The figure is inspired by: (Silva, Balanzategui et al. 2019), NREL Quality Management Handbook and Solar Bankability Project.

2.9.1. Quality Assurance

Quality assurance include calibration and measurement procedures and suitable instrument selection for the intended use of data. Some sensors could have fitting accuracy, but the sensors could still be a suboptimal choice for the microclimate in which the module is located due to degradation of the sensor itself. Furthermore, it is necessary to acquire data correctly. Some important points to remember when assuring the quality of PV data sets are as follows (Silva, Balanzategui et al. 2019):

1. Sampling rates affect the data set and data analysis. A great sampling rate might generate a more accurate data set, but it might become excessively large and cumbersome to analyse afterwards. Similarly, a slow sampling rate will result in a data set that will be easier to handle, but it could hide certain trends or faults.
2. Measurement ranges could limit the extent to which faulty measurements deviate from the expected.
3. Weather conditions should be acquired to check the validity of the measurements.
4. Raw data files should never be modified, and there should exist extra copies in case of mistakes that prevent any future use of the data.

The software used should also be evaluated and monitored. Logging errors could result in broken data sets, as well as processing software might produce corrupt sets.

2.9.2. Quality Assessment

Quality assessment assure the validity of the data sets by checking if the data acquired are reasonable. Testing validation criteria and visual inspection will identify errors, which leads to reliable scientific studies and more accurate estimations for energy production (Silva, Balanzategui et al. 2019). Silva et al. suggest the following methodology for quality control of irradiance for PV data sets (Silva, Balanzategui et al. 2019):

1. Control the data recording time.
2. Visual inspection of solar radiation components
3. Confirming physically impossible values
4. Confirming physically possible, but extremely rare values
5. Testing for inconsistencies across values. Some values are known to have a correlation and when the data points do not, the data points could be wrong.

Further quality assessment should be done by experts. Graphing the data sets can be an additional tool, especially useful for investigating periods with many flagged data points. Data sets for days may consist of large data files, depending on sampling rate, and analysis may be easier with visual aids.

The aforementioned list for irradiance could be translated to use for meteorological data values. The microclimate in which the module is located will determine some upper and lower bounds for temperature, wind speed, pressure, and relative humidity, which can flag other data points (Silva, Balanzategui et al. 2019).

Finally, Silva et al. suggests that monitoring the performance of the measurement equipment should be performed to make sure that faults or breakdown of the equipment do not ruin an entire research project (Silva, Balanzategui et al. 2019).

2.9.3. Quality Enhancement

Wrong or missing data points identified from the quality assessment should be removed. Missing irradiance data points could be replaced by accurate values, if other irradiance parameters are in place to do so. The datapoints could also be interpolated from the larger data set, if enough of the data set is not corrupted (Silva, Balanzategui et al. 2019). Removed data points should be marked as "NULL", "NaN" or similar, depending on the programming language used.

Another option is to calculate mean values for larger time intervals if some data points are missing. Hourly mean values can be calculated if more than 50% of the data points from that hour is present. Daily mean values can be calculated if more than 75% of the data points are present. The remaining points can be interpolated from the hourly values.

3. Methodology

3.1. Introduction

To answer the research questions posed in the introduction chapters there has been gathered information from various scientific sources and analysis of data gathered from a PV system owned by Glava Energy Center (GEC) in Glava, Sweden. GEC owns several different systems, but the systems that have been analysed in this thesis is their "Ongrid System" (Glava Energy Center s.a.). The Ongrid System is grid-connected system intended to be analysed for research with two parks: Solar module park 1 and solar module park 2. Solar Module park 1 has been analysed in this paper. The four different technologies are presented in table 1 below:

Table 1: The specifications of each plant in Solar Module Park 1 (Glava Energy Center s.a.).

Plant name	**Size**	**Module number and producer**	**Inverter type**
Plant no 1	4,6 kW	20 from REC solar	SMA
Plant no 2	17,6 kW	80 from Innotech Solar	Eltek Valere
Plant no 3	86 kW	400 from REC Modules	ABB
Plant no 4	28 kW	120 from Innotech Solar and SweModule	Microinverter, Optistring, SolarMagic and SMA

Figure 4: The four inverters at for the different plants at GEC. Photo: Jesper Frausig.

The PV system is in Glava, Sweden. Map coordinates are 12.62 longitude and 59.53 latitude, and is marked on figure 5 on the map below from Google Maps:

Figure 5: A map displaying the location of the solar research centre in Glava, Sweden. The red marker is placed in Glava and the image is a screenshot from Google Maps (Google Maps s.a.).

The data set has been analysed by one PhD candidate and one master's student at NMBU. The review of the data uncovered necessary measures to clean the data set. Analysis of selected days, months, and events provided deeper understanding of the data, so that analysis could be generalized to the entire data set of roughly 10 GB of relevant parameters, out of a database in excess of more than 120 GB.

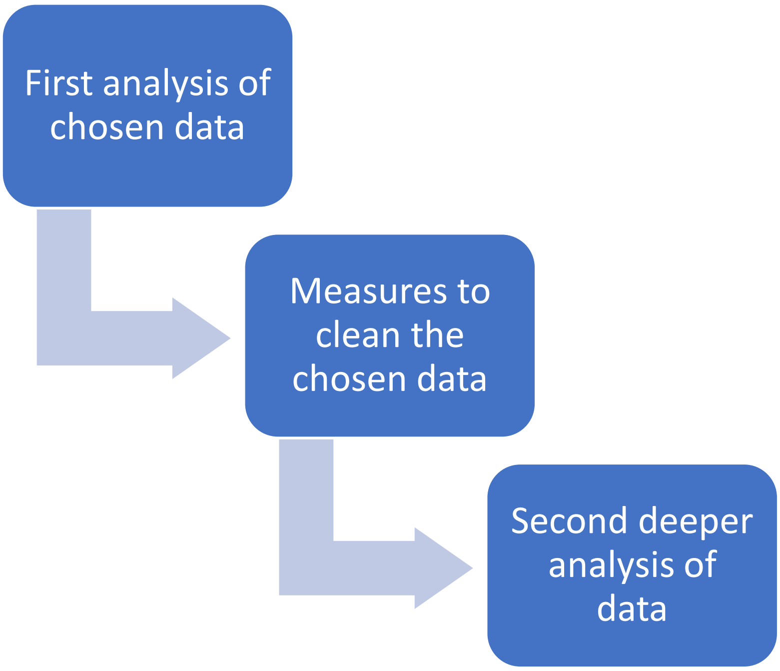
The complete methodology for the thesis looked as follows:

1. Acquiring data
2. Data analysis (see expanded list and explanation below)
3. Presentation of results and evaluation of findings

The data set was acquired from GEC. The selected data set is in its totality ~10 GB, which was exported from their logging program into excel-files. There were also acquired satellite images day-by-day for the period to verify data regarding clouds and snow. Additionally, there was obtained satellite data for simulating irradiance for the select location.

3.2. Data analysis

A simplified version of the data analysis process (number 2 in above list) looked as follows:



This approach is circular, because it proved difficult to determine an approach to the data set before seeing the data set. This means that there have essentially been several rounds of discoveries in the data set that changed how the data set was analysed. As insight was gained, so was the methodology and measures changed accordingly. This is also in accordance with the expert recommendation regarding analysis and quality control of PV data, presented in chapter 2.9 "Quality of Solar PV measurements", as well as the general CRISP-DM methodology for data mining. As CRISP-DM explains, researchers are often moving back and forth between different phases, which was also true for this thesis.

3.2.1. First analysis of chosen data

The first-hand analysis was used to understand the data set and its features. At first glance this meant understanding the different output parameters, and which part of the module each value corresponded with. The data and python script used to evaluate the data was stored on a cloud computer and accessed thought 'Google Colab' in order to let the student and PhD candidate work together.

3.2.2. Measures to clean the chosen data

Cleaning the dataset of incorrect values for the first time was a comprehensive process involving up 10 different checks that disqualified a value each. Each value that failed a test was marked for later analysis.

One way to quantify the quality of the data sets were to use Quality Control Index (QCI) to determine the amount of corrupt data in each set. This was done on a monthly level. QCI is developed by Pecos and tests all numbers for different parameters (Klise K.A. and J.S. 2016). This analysis defined the interval of accepted data between minimum and maximum values, limited the maximum a value could change by each delta and set the amount of false data in a row to be at least 2 before they would be flagged. This list can be expanded and reduced to fit the analysis. Pecos takes the fraction of accepted values divided by total values to give the QCI between 1 and 0.

Periods of time with an unusually large number of disqualified values were analysed to understand the root cause. Given that the understanding that was gained as the work progressed, parts of this work were non-linear: Some results of the analysis of the data set granted meta-understanding of the data set, which improved the robustness of the analysis by changing the method of analysis. For example, PV modelling supplemented or replaced some types of generic quality control checks.

Some extreme values were unnecessary for the research questions, since the inaccurate information did not assist in obtaining any results. However, these values could give information about degradation of measuring equipment, which would be relevant to observe for those in charge of the O&M. As described by chapter 2.9.2 "Quality Assessment", data points that are incorrect should be removed.

3.2.3. Second deeper analysis of data

The second and deeper analysis of the data included visually inspecting the GHI plot of each month. Whenever something that was deemed important were noticed, it was logged and stored in a separate file. Similarly, a heatmap of the GHI for each month was also visually inspected, and notable events were logged. Thereafter, there was an investigation into the different comments. Certain notes were rectified immediately as problems with the script. Other comments were investigated to achieve better understanding of the data set, and some comments lead to the findings that is presented in the result chapter.

The deeper analysis was developed further to create “interactive reports” that presented a large array of data in the browser. The visual representation was interactive by having the option to slice the time periods and to toggle data points in a graphical user interface (GUI).

During the deeper analysis there was also simulated expected power output from satellite data. The simulation used 15-minute rolling windows for output, and there was experimented with different uncertainty bands. Originally, there was used three standard deviations from the mean, but later there was used two standard deviations.

Finally, there were calculated correlation and R2 values for each parameter. The values were places in a colour coded matrix for visual inspection. Three such correlation matrices were constructed: One calculation containing all data points, one calculation containing only the flagged datapoints, and the last one was the sum of the two matrices.

Smaller correlation thumbnails were also employed for inspecting the relationship between two values. These were used in the interactive reports.

In the GUI there were figures for:

- Two-parameter correlation plot
- Availability, performance, and quality scores
- Mean, median, minimum, maximum, standard deviation values for power output
- An irradiance heatmap
- Power output, simulated uncertainty band for power, and simulated clearsky plot
- Monthly deviation trend

3.3. Other comments regarding methodology

The data from GEC was used to address the issue of uptime. Uptime is generally an issue of availability, so it was essential to determine if the solar modules generated maximum possible power for the irradiance available. To understand this, it was fundamental to cross-reference the amount of irradiance at any given point in time. Only when it was certain that there were no clouds blocking the solar radiation, the availability of the solar modules could be analysed.

The sources of data were chosen to have a suitable resolution. To evaluate uptime, it is necessary to have high resolution in the data, since the timestamp must match the event analysed. Determining other present conditions at any point in time will be crucial to make the correct conclusion, and therefore the data had measurements every 6 seconds. The data could naturally have had a higher resolution than every 6th second. That would demand more processing power and time, and there was no such data set available. A lower resolution would mean that the results would have a higher uncertainty, because rapidly moving clouds, shadows, and other unexpected events would be harder to separate from other sources of reduced uptime.

The data sets were treated several times to make sure it would result in a meaningful analysis. One crucial part was to make sure the data from different sources was adjusted to the same time steps. Operational data about the solar modules was gathered by the solar modules itself, while satellites were used to gather satellite data. Finally, data was simulated with use of PVlib Python to generate idealized data for irradiance when assuming optimal clear-sky conditions.

The point of this research is partly to eliminate the need for operators to analyse their own O&M data, and instead use techniques presented in this thesis and following papers to make it manageable for laymen to perform proper O&M without an engineering background. Thus, it was necessary to spend time understanding not only what were likely causes of degradation, but also how to treat data in an effective and robust way. Understanding which time periods were expedient to analyse for the posed research questions are also important.

3.4. Quality assurance of the data set

To achieve scientifically robust results that could be generalized, there was necessary to ensure that the data used for calculation had sufficient quality. The measuring equipment used by GEC includes some innate errors, which results in some incorrect values. For example, temperature values below -60C were removed, since they do not grant any meaningful information regarding the research questions.

Additionally, the data set was evaluated to ensure the legitimacy of the numbers. One part of that involved simulating clearsky conditions to review whether the irradiation measured by the equipment was within a reasonable interval. Measurements above that value would either be cloud focusing or incorrect numbers. Regardless, those values were discarded. Likewise, values of irradiance below 0 W/m^2 were discarded. When discussing maintenance, deviations from the expected is the most interesting. Therefore, the focus has not been on random or chaotic values, but meaningful analysis of the degradation of the solar modules.

Data sets obtained from GEC AS were originally exported as Excel csv-files. The data sets were converted to Python feather-files for faster processing and lower memory demand. Additionally, data points from when the sun was -5 degrees below the horizon were removed from the files to reduce time spent on processing data values during night-time. This action reduced the data sets by roughly 2/3 of the original size. Although this process destroys information, there should not be any uptime to analyse during night-time. It is then assumed that this does not impact the results regarding uptime of the system. However, this action will impact the data set, and could make it harder to detect other faults in the measuring equipment. These faults are interesting for future research, but they are beyond the scope of this paper.

.

4. Results

The results gathered will be presented in the order in which they were obtained. There will first be presented results regarding defect measuring equipment for the initial analysis. Thereafter there are presented results for the interactive reports. The chapter ends with results for the OEE indicators and other analysed parameters.

4.1.1. Corrupt data points

The quality of data presented had several instances of corrupt data. In this paper, “corrupt data” refers to an instance of data points missing, data points assigned an incorrect value, or where the timestamp and the data point mismatches.

- Missing datapoints were generally NaN (Not a Number), where meaningful analysis about the problem would be impossible.
- Data points with incorrect values could happen anywhere, and it is not possible to claim that any point has a correct value without verification through additional sources. However, data points where the irradiance is -500W/m2 for no discernible reason were ruled out as wrong data. The root cause could be faulty measuring equipment, faulty logging etc. These values need to be removed to make reasonable calculations about the remaining data set.
- Datapoints matched with wrong data points would present themselves as uptime during night-time. This happened sometimes due to exporting problems, and sometimes it happened due to logging errors.

4.2. Defect measuring equipment

Below follow examples of defect measuring equipment. The examples do not include every error, but they will serve as a basis for highlighting the problems that have been discovered.

4.2.1. Non-defect measurements of GHI

For the sake of comparison later in this report, figure 6 of GHI and solar height for February 2014 below does not contain any obvious corrupt data points:

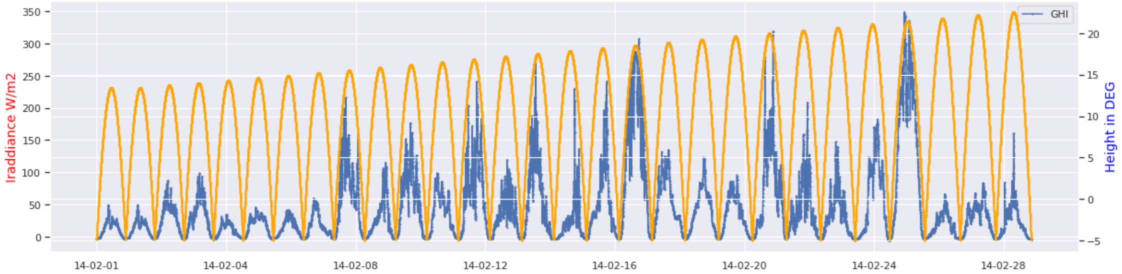


Figure 6: The plot displays GHI from February 2014 without obvious corrupt data.

4.2.2. Missing data points

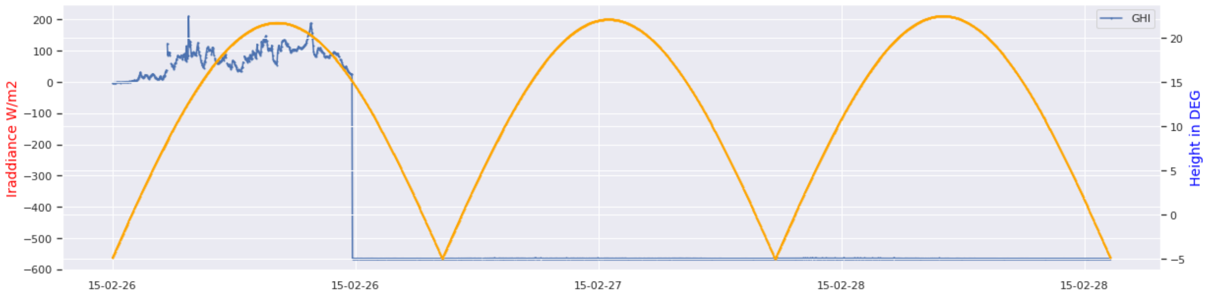


Figure 7: This plot should display the entire February of 2015, but it exists only data points from 3 days.

“Missing data points” means that the data points for a period is non-existent. Figure 7 displays the existing data points for February 2015, which only contains 3 days of data. Additionally, two of the three days have corrupt data. The last two days suggest that the GHI is constantly -550 W/m2.

4.2.3. Inaccurate data points

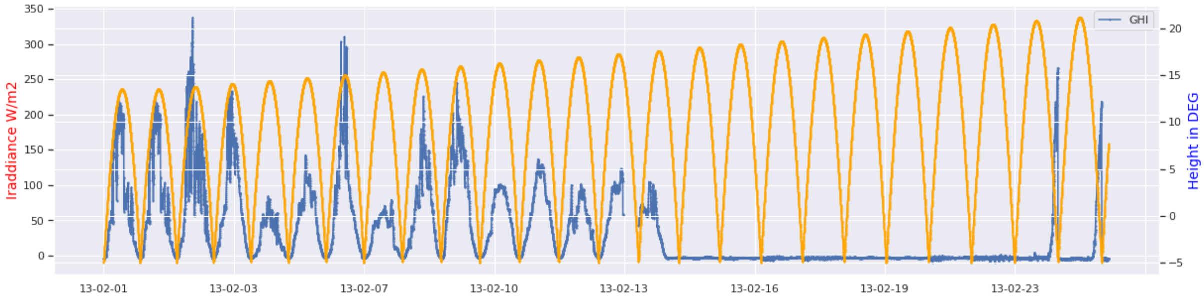


Figure 8: Plot of the GHI for February 2013. The blue line remains close to 0 W/m2 from the 14th.

Figure 8 above shows an example of how a simple visual inspection of a data set can reveal corrupt data points. Even with thick clouds there is expected to be some sort of irradiance absorbed by the measuring equipment. Some data values for day 3 and 6 are also above the expectation, the second plot display solar height and as such is only indicative.

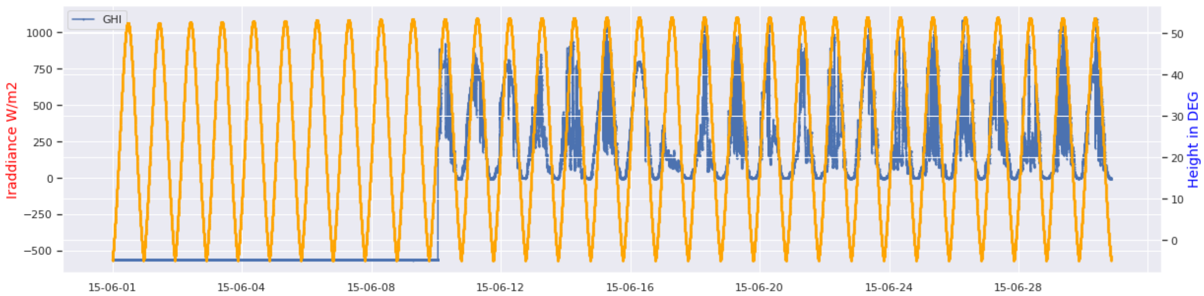


Figure 9: This figure is much harder to visually inspect, since the range of right Y-axis is wider than in normal plots of June.

As shown in figure 9 above, corrupt data can also make the remainder of a month difficult to visually inspect, since the Y-axis is adjusted to the data points plotted. This can naturally be changed by constant height on Y-axis, but that would be less dynamic and would remove information.

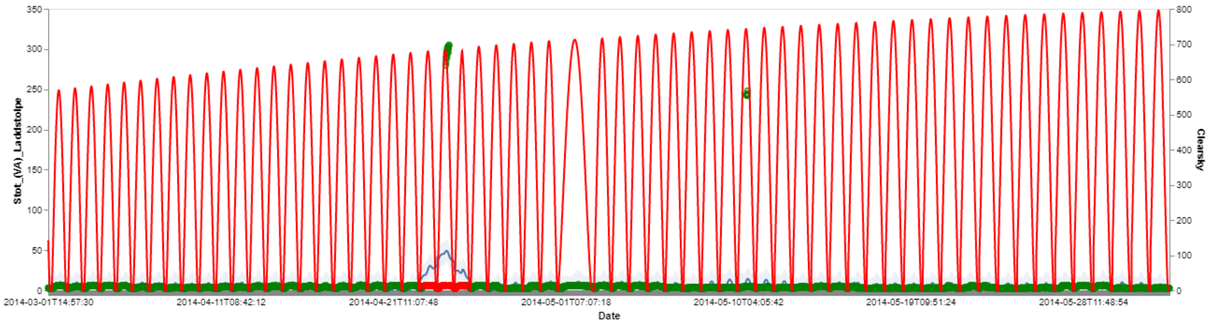


Figure 10: Missing data points for power output thought the Laddstolpe inverter. The green points indicate power production.

Figure 10 above displays two months from the site using a Laddstolpe inverter. The system failed, and there are no usable data points for power available.

4.2.4. Mismatched timestamp with data

The data points presented in figure 11 below for the latter half of the month has been mismatched with the time stamps. The error does not carry over to the next month, since the data sets are exported individually for each month.

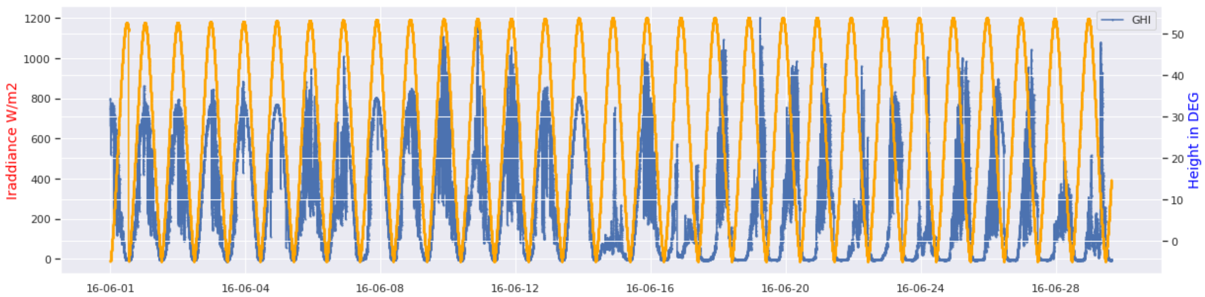


Figure 11: The latter half of this month has a mismatch between the datapoints and the timestamp.

4.2.5. Quality Control Index to evaluate data quality

During analysis there were noted observations regarding the month analysed. One example of such notes and comments is exemplified by table 2 and figure 12 below. The comment was saved with the month it corresponded with. The comment for September corresponds with bad data, since the data did not exist. October had apparently better data, as noted by its QCI. The comment for December also suggests that something is wrong with the dataset, and this could also have been assumed by looking at the QCI.

Table 2: Example of unprocessed comments from visual analysis and accompanying QCI for September, October, and December 2012. The note for December is translated for the purpose of this paper, but it is nonetheless an incorrect explanation. The correct explanation is that the heat map spectrum is compressed due to corrupted (in this case too low) measurements of GHI.

comment_2012_9QCI:-1.0 *- No data*
comment_2012_10QCI:0.929
comment_2012_12QCI:0.843 *- Cf. comment from 'months'. It appears GHI is corrupt, which adjusts the colour spectrum and makes it difficult to see the differences clearly.*

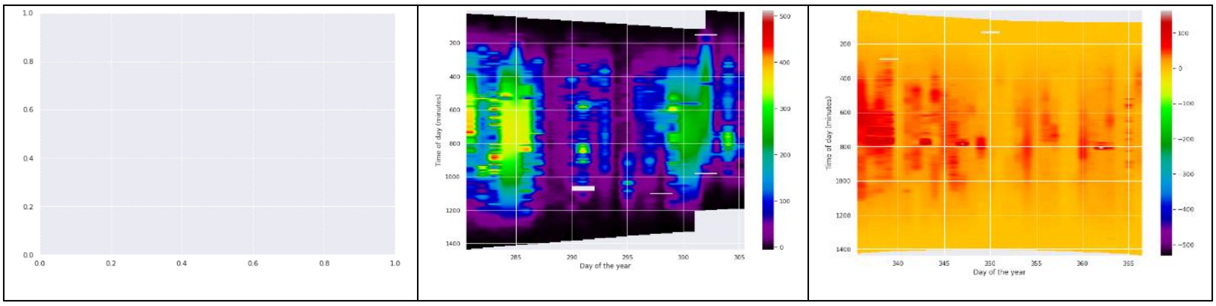


Figure 12: The heatmaps corresponds with the table above. Left to right: September, October, December of 2012. The clock is set backwards in October and shifts the graph.

4.2.6. Indications of forewarning before measuring equipment failure

There were made observations for the time periods before measuring equipment failure. Below follows three heatmaps that displays the behaviour of the irradiance before a failure of the measuring equipment.

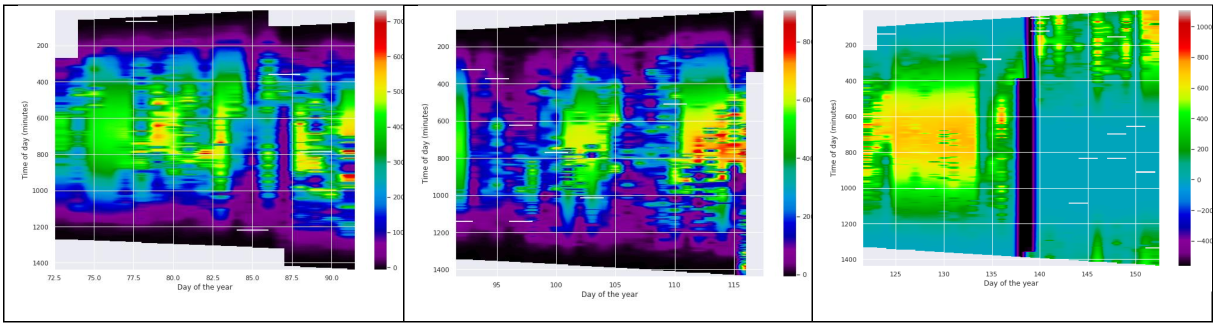


Figure 13: The heatmaps left to right display irradiance for March, April and May in 2016. In May there is a failure of the measuring equipment that corrupts the remainder of that month. The plots are displayed in the addendum as Attachment 3.

There are multiple instances of measurement failures. They last for days or up to several months. Sometimes there are months with more missing data than normal before these failures.

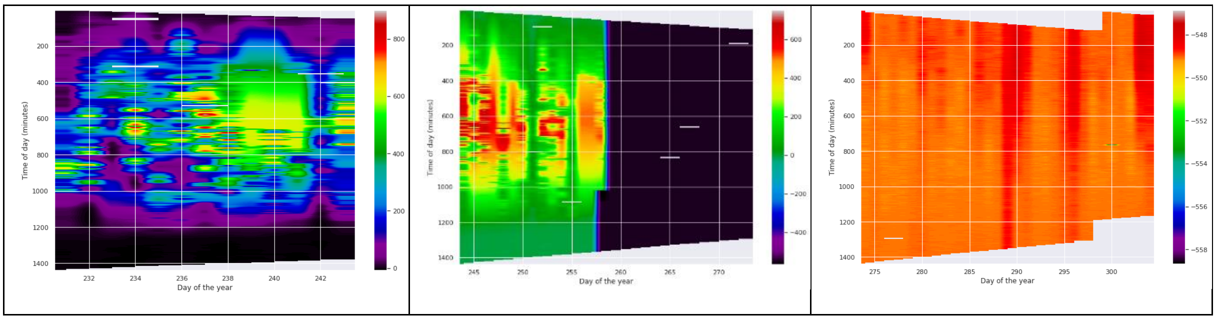


Figure 14: The heatmaps left to right display irradiance for August, September, and October in 2014. The measuring equipment failed in September and lasted for several months. The plots are displayed in the addendum as Attachment 4.

4.3. Interactive reports

4.3.1. Solar irradiance from several angles

There was conducted analysis of irradiance from POA and GHI. POA irradiance proved to be the most accurate model to conform with the simulated power. While GHI needed three standard deviations to accurately match the measured irradiance, the uncertainty band of POA needed only two standard deviations to reach similar accuracy.

4.3.2. Satellite images

Data sets and results were verified by checking satellite imagery for the day or period analysed. In this context, there was tested if snowfall would be noticeable as a failure mode on the irradiance measuring equipment. The satellite images are only a snapshot from one day at a time.

4.3.2.1. Snow affecting measuring equipment

When attempting to find evidence of snow affecting the uptime of the system, a paradox arose. The intention was to verify that that system could operate at peak performance (i.e. under clearsky conditions), and then to identify that it did not operate at peak capacity after snowfall.

However, identifying clearsky after snowfall proved to be a challenge, as snowfall could also cover the measuring equipment that would identify clearsky conditions.

A great example of this was December 2018 in figure 15 below.

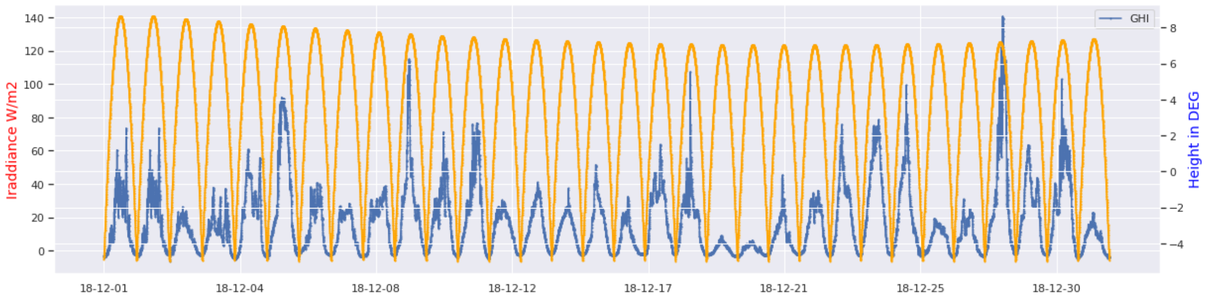


Figure 15: The plot shows GHI for December 2018. It can be argued that either the snowfall on days 6th-8th influences the availability of the system or it does not. This figure displays the difficulty of reading visual data.

From satellite imagery it is visible that there was no snow on December 5th. On December 6th-8th there are cloudy satellite images. On December 9th the satellite images are clear again, and snow is clearly visible, which has fallen between 6th and 9th. These satellite images are presented below as compressed images, but they can be found in full resolution in the Addendum as "Attachment 2".

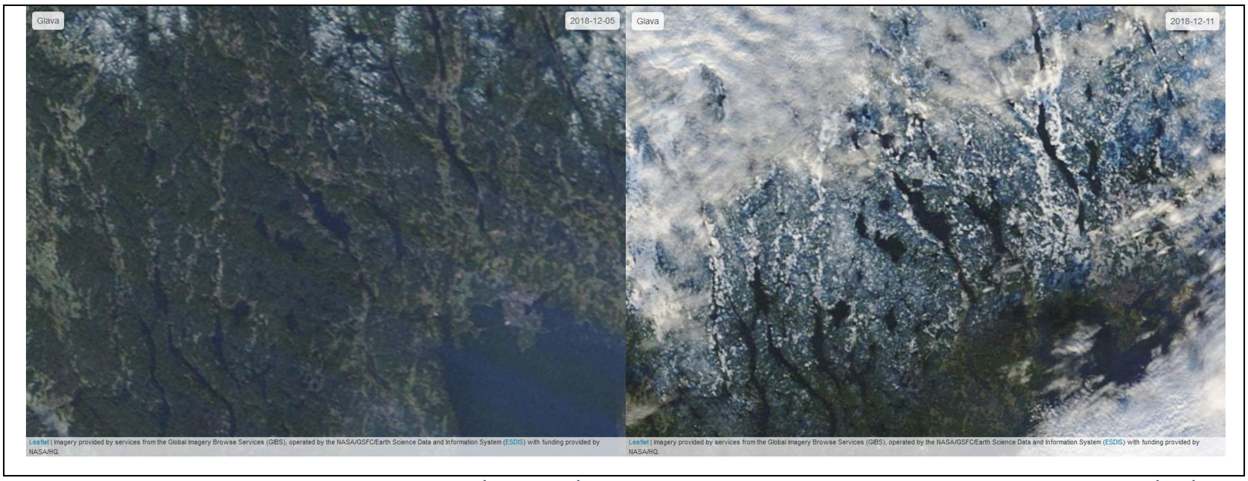


Figure 16: Satellite image from December 5th and 11th, 2018. There fell snow sometime during December 6th-8th.

Investigation of output in certain time periods could be coupled with satellite images to conclude that there has been snowfall. This insight could be used in combination with output data to analyse the impact of snowfall. However, access to high resolution snow coverage data would be an alternative approach.

Another example of this effect is displayed below in figure 17. The first day is January 2nd, where the satellite image shows no snow. Snow falls on January 3rd or 4th, but there is not obvious on January 5th that there is any snow covering the measuring equipment.

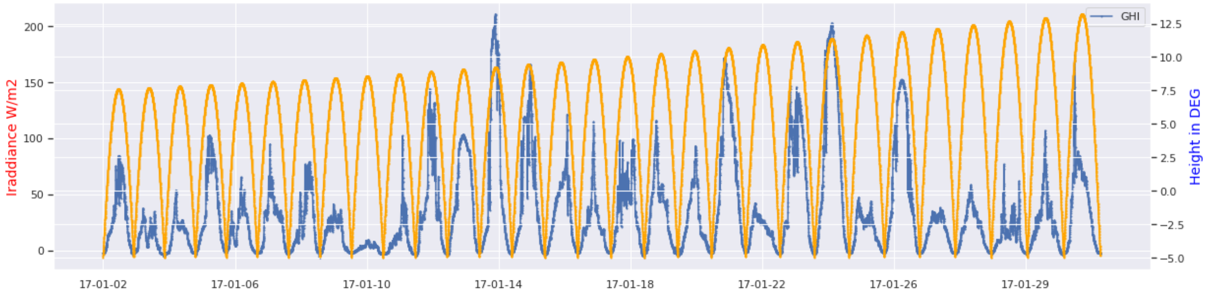


Figure 17: GHI for January 2017, which displays no obvious signature effect on the GHI after snowfall on January 3rd or 4th. The plot begins at January 2nd, due to lack of data from January 1st.

For reference, there was investigated if the irradiance measured on January 15th, 2017 was lower than an equal clearsky day in January another year. Figure 18 below from 2020 shows that January 12th had clearsky, but lower overall irradiance. Satellite images from that date shows that there was no snow.

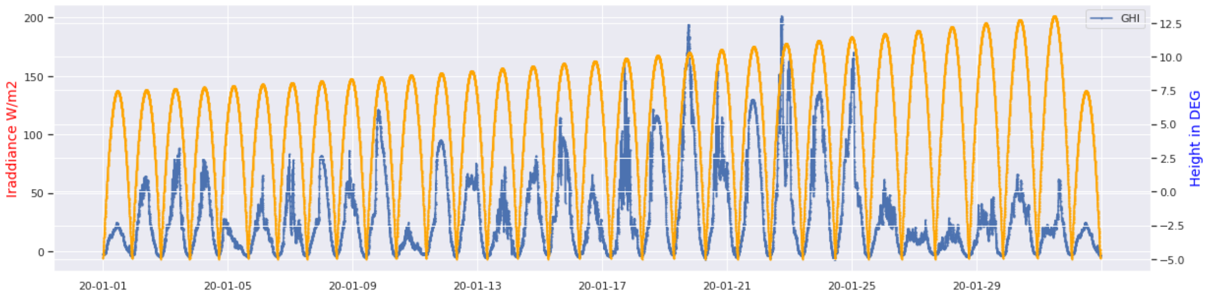


Figure 18: GHI for January 2020. January 12th has clearsky conditions.

4.3.2.2. Faulty satellite images

Satellite images can contain a diagonal cut through the image that reduces the trust in that image. The cause for this issue is unclear. Figure 19 below illustrates this effect. The line is located in the same position in all observed cases.

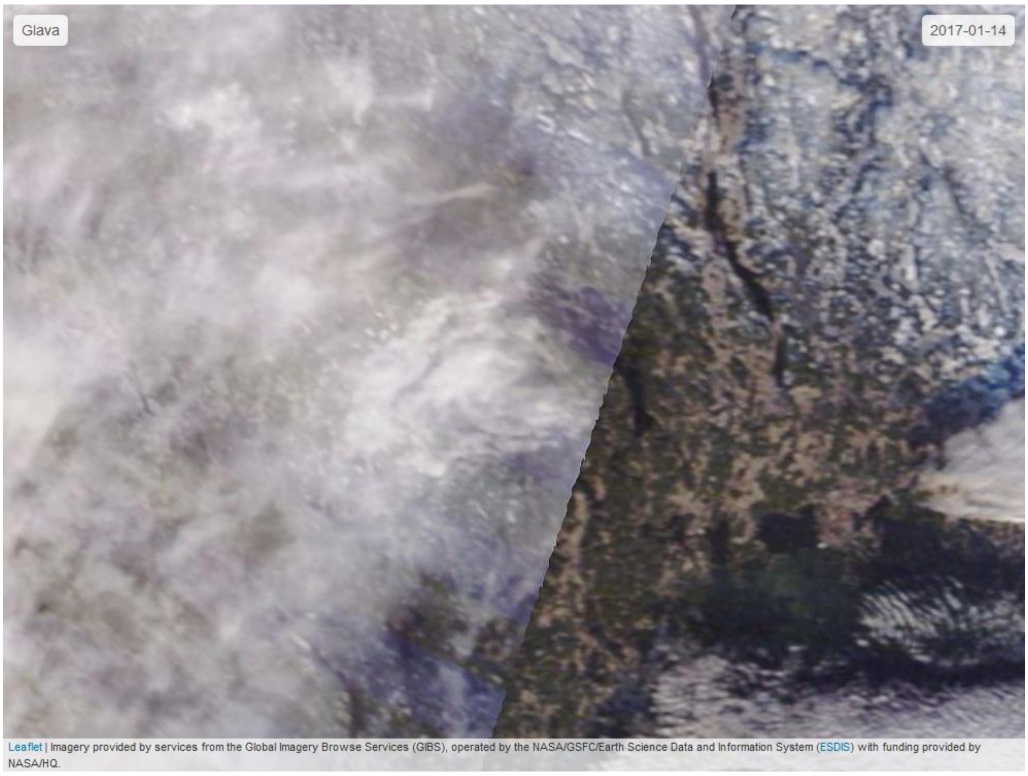


Figure 19: Satellite image for January 14th, 2017. There appears to be two different images combined. This image confirms that the photos are from the same location.

4.3.3. Differences between sites

There was sometimes founds different power output from the 3 sites when comparing over the same time period. The sites use different technologies and dimensioned differently, so the values are not equal. The three sites are presented below in figure XX. For the sake of easy of readability there is only plotted 7 days. The complete figures are presented in the Addendum as "Attachment 1".

The uncertainty band is the light blue areas. The size of the uncertainty band depends on the variability of the satellite data.

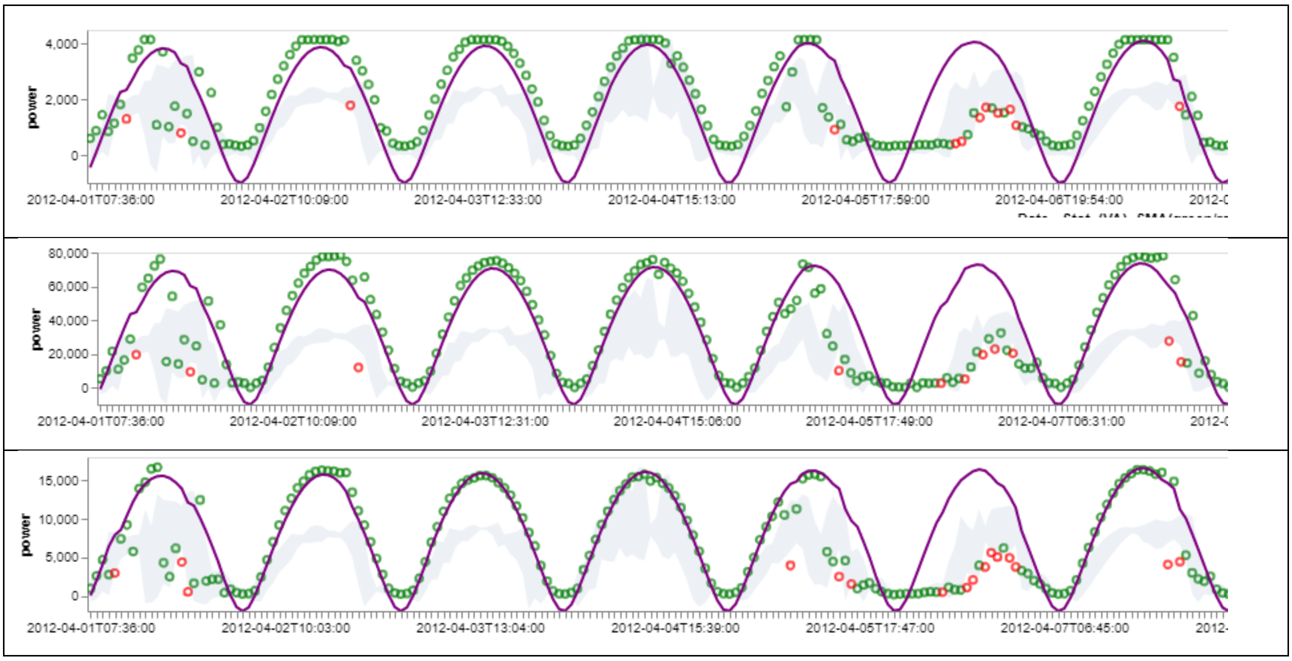


Figure 20: The three plots (20a, 20b, 20c) display the power output, the uncertainty band and the clearsky GHI. Top to bottom: SMA, ABB, Eltek.

The circles are coloured red whenever the power output is lower than the uncertainty bound. The axes have different units of measurement: Power is measured in W against the left axis, and GHI is measured in W/m2 on the right axis. The right axis includes has been cut for readability purposes, and it includes 0-800 W/m2.

Figure 20 is one example of how measurements can vary between systems in the same location. On day 2 there is a flagged point for SMA and ABB, but that point is not registered for Eltek. Likewise, day 5 have 1 flagged value for SMA and ABB, but three flagged points for Eltek.

The uncertainty bound does not seem to align with the measured output for these select 7 days, since most points do not lie within the light-blue interval. Figure 20 is also a visual representation of which data points are flagged and inspected. Data points above the uncertainty bound has been investigated for research purposes, but they are not flagged by the system.

On day 2, 3, and 4 in figure 20a there is visible how the power production follows a rough sine wave most of the day, but it appears to reach a maximum threshold at the peak of the waveform. This can be seen more clearly when this power output in 20a is compared to 20b and 20c.

4.4. Quantifications of OEE parameters and indicators

There was performed an attempt to quantify the OEE parameters introduced in theory section 2.1.1.

Availability was defined as percentage of the period in which the data points are measured above the lower end of the uncertainty bound. This excludes any flagged data point. Availability values varied between the different systems. Four months from 2012 were analysed. The following maximum and minimum values were found for each system:

System	**Minimum value**	**Maximum value**
SMA	0.8528	0.9432
ABB	0.8517	0.9757
Eltek	0.6135	0.9630

Performance was defined as the average value of the entire period compared to the uncertainty band. Any point above the lower bound was defined as 1, and any point below was some number between 0 and 1. The average value for the entire period was denoted as the performance.

Four months from 2012 were investigated. The following maximum and minimum values were found for each system:

System	**Minimum value**	**Maximum value**
SMA	0.9379	0.9767
ABB	0.8857	0.9818
Eltek	0.7524	0.9855

The system using Eltek-inverter has the lowest minimum values. The minimum values for Eltek are both from December 2012, where most power values are ~0 W. The reason for the difference between the different technology is not evident.

Quality was initially defined as the average of all the R-squared values for the different parameters in question. This would result in lower quality for circumstances where the data did not correlate with our model and understanding. However, later quality was defined as period of performance above the upper uncertainty bound.

Five months from 2012-2013 were investigated. The following maximum and minimum values were found for each SMA. (The other systems did not comply with the code):

System	Minimum value	Maximum value
SMA	0.2333	0.5144

4.5. Other analysed parameters

Correlation between parameters could indicate the importance of that parameter for O&M. This thesis investigated windspeed, angle of incidence, precipitation, and pressure. One way to crudely analyse correlation is to plot them all in one map, as in figure 21 below. This gives an overview of which parameters are positively and negatively correlated. Several of these parameters are naturally correlated and will not provide meaningful conclusions alone.

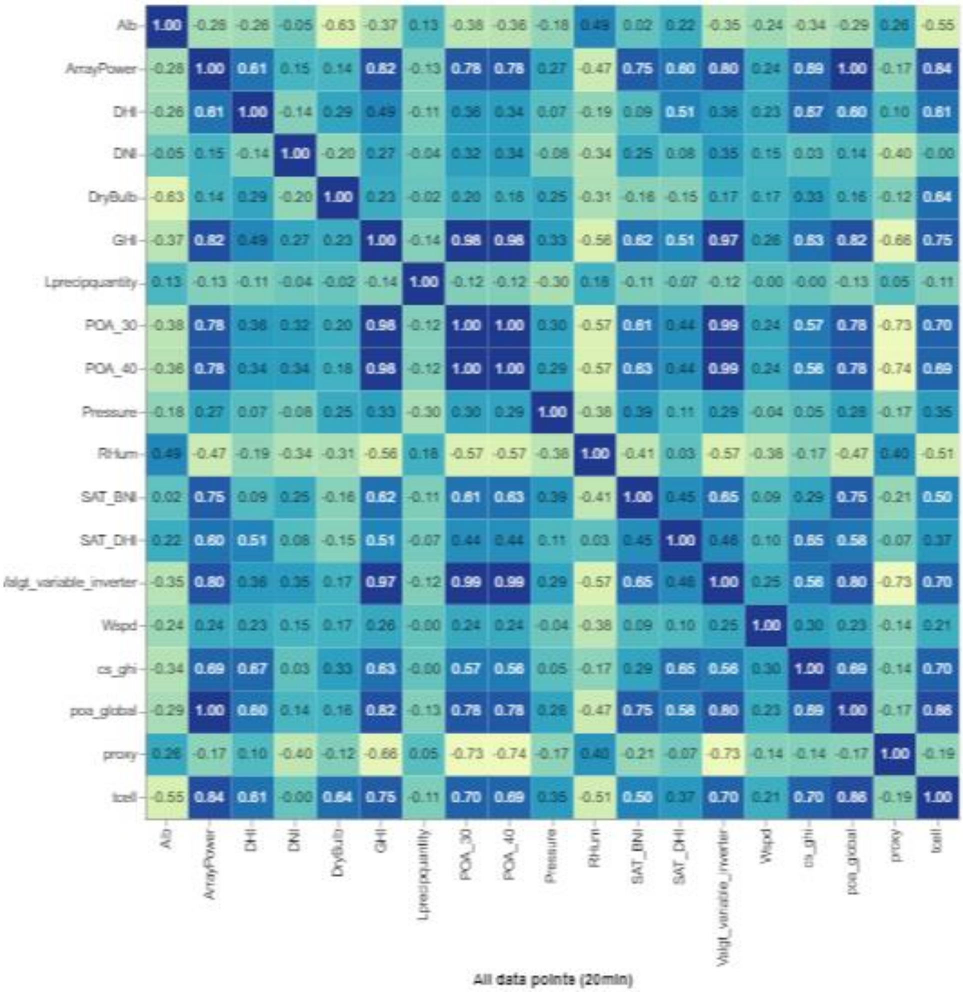


Figure 21: This figure is a plot of the correlation between each parameter.

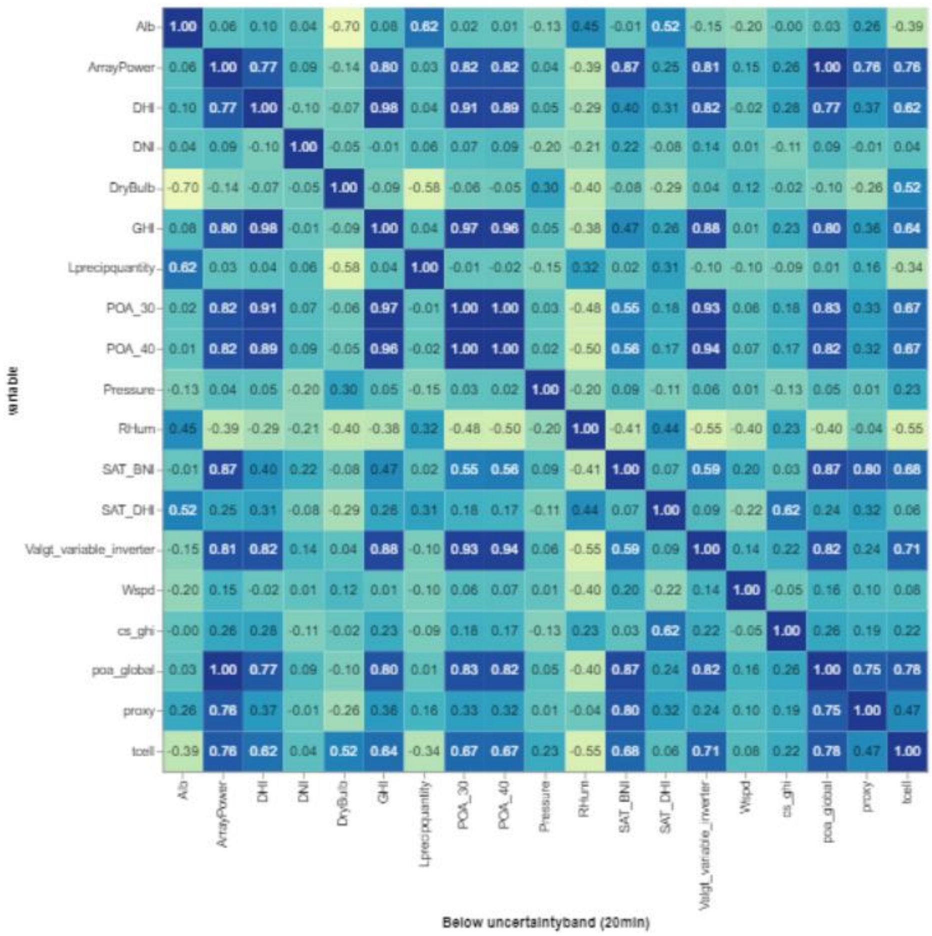
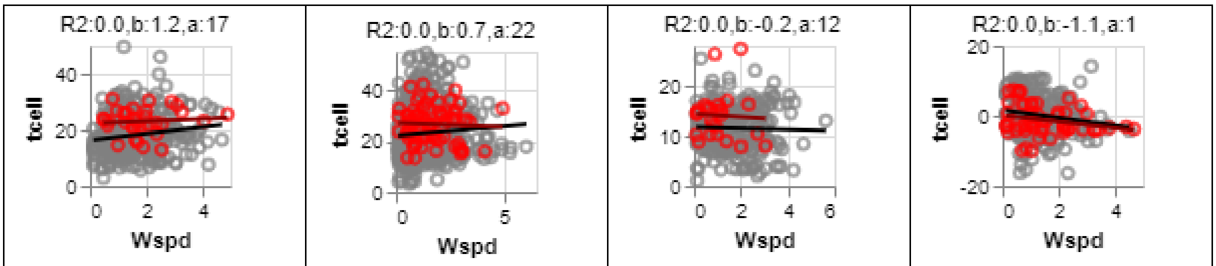


Figure 22: This figure plots the correlation values for every parameter, but only the data points that have been flagged for being lower than the uncertainty band.

Windspeed was plotted against the simulated temperature. The plots below in figure 23 show examples of what the figure could look like.



April 2012 | May 2012 | November 2012 | December 2012

Figure 23: Windspeed plotted against temperature in the solar cell. The two lines are the best fit regression lines. The red circles correspond with the data points that have been flagged for being power production below the uncertainty band. R2 is the R-squared number for each graph. There appears to be no correlation between the windspeed and the temperature in these cases. These examples are from the system using an SMA inverter.

Angle of incidence (AOI) was plotted against GHI. The plots below in figure 24 show examples of what the figures could look like.

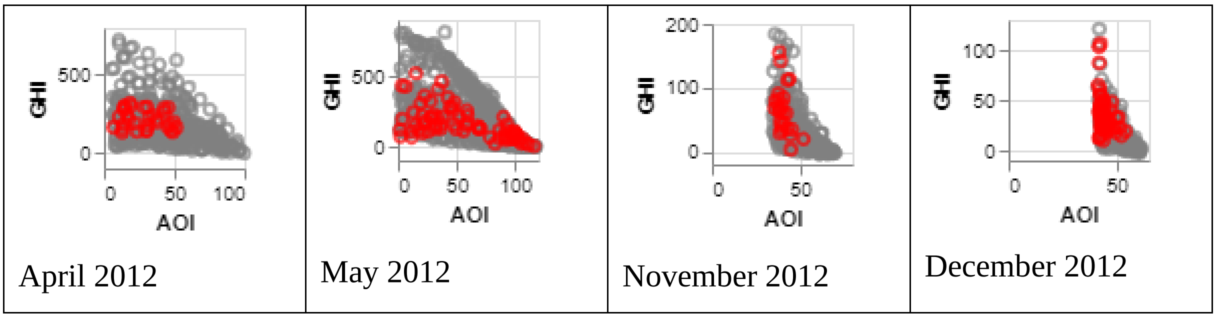


Figure 24: AOI plotted against GHI. The red circles correspond with the data points that have been flagged for being power production below the uncertainty band.

Precipitation was plotted against "sd_v_oc". The plots below in figure 25 show examples of what the figures could look like. Voc is the modelled open circuit voltage of a single module, calculated as check value.

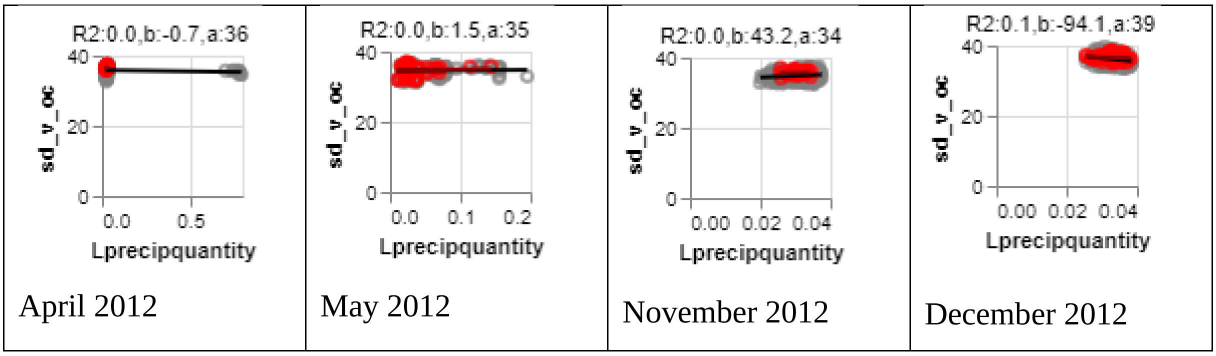


Figure 25: Precipitation plotted against "sd_v_oc". There are different x-axes, so this will skew the data points accordingly. The lines are best fit linear regression. The red circles correspond with the data points that have been flagged for being power production below the uncertainty band.

Pressure was plotted against GHI in figure 26 below.

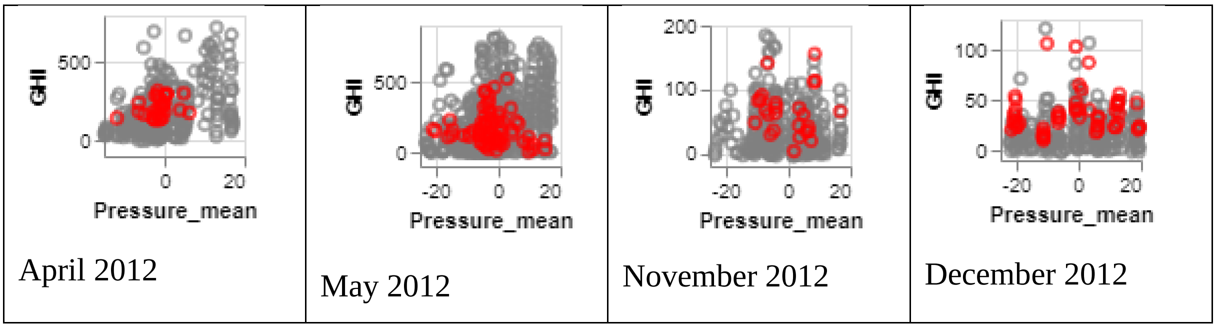


Figure 26: Mean pressure against GHI. There are different Y-axes, so this will skew the data points accordingly. The red circles correspond with the data points that have been flagged for having power production below the uncertainty band.

Maximum power output was studied from the interactive reports. The maximum as-built specifications were defined by GEC on their website. The results and the as-built specifications are listed below. The maximum power output was reached in April for both plants.

Plant name	**Size**	**Module number and producer**	**Inverter type**	**Max power output measured**
Plant no 1	4,6 kW	20 from REC solar	SMA	3,9 kW
Plant no 3	86 kW	400 from REC Modules	ABB	88 kW

5. Discussion

This chapter will discuss the results obtained and approach the findings from different perspectives. The discussion should as a minimum answer the research questions posed in the introduction. Additionally, there has been made observations and gained insight during the research that is worth discussing. The chapter should also be reviewed in tandem with the results for graphs and values, as well as the theoretical basis for context.

5.1. Parameters and indicators

5.1.1. POA as a parameter and indicator for the status of the PV system

The most important parameter for understanding solar PV is the amount of irradiance the modules receive. The irradiance is the source of solar power, and with 0 irradiance there will be 0 power production. Multiple ways to measure and predict irradiance exist, and some of these methods have been explored in this thesis. The results obtained in this thesis indicates that POA40 is the parameter that most closely resembles hybrid power output for this system, shows robustness, and retain consistent relations to other parameters. The output uncertainty band was estimated based on simulated power production.

One should also mention that the spectra from GHI by satellite also ought to be adjusted to the spectra response of the solar cell, and in all likelihood, the values adjusted to fit ground-truth. By analysing the POA irradiance, the most precise estimation for the behaviour of the system may be performed. If operators could only install one type of measurement for irradiance, the results could indicate that POA would be the most precise approach.

However, one measurement of irradiance should be accompanied by other measurements of irradiance. The data set analysed in this thesis has countless instances of missing data for irradiance. Independently of the root cause(s) responsible, these data points cannot be easily obtained. There are a few possible solutions to rectify this issue. As mentioned by Silva et al., it is possible to calculate mean values for a period if the values are missing (Silva, Balanzategui et al. 2019). Another option is to analyse data from other functioning logging equipment for irradiance like GHI, DNI or DHI. Both strategies are inaccurate or not satisfactory, but they may provide some insights into the interval without data points. Additional measurements may further decrease uncertainty and can validate data quality in general.

Furthermore, POA should be analysed within the context of albedo. Albedo can greatly change during winter with the occurrence of snowfall. Analysing the changes in albedo would be helpful, since POA includes ground reflection into the measurement to an extent. Albedo will depend on the local configuration and microclimate, so it may vary between systems.

5.1.2. Temperature as a parameter and indicator

The results regarding power output for the system gives clear indications that the efficiency of the system increases with lower temperatures (for example: winter) and likewise increases for higher temperatures (for example: summer). This is also in accordance with literature, as Pearsall claims the operating temperature is the second most influential parameter to predict module system output (Pearsall 2017).

However, there are some challenges with using temperature alone to predict the output of the system. Temperature is a function of several parameters, many of which can be measured. They are described in detail in Theoretical basis under “Operating Temperature”. In order to perform meaningful analysis regarding the temperature, there need to be context regarding the parameters for the entire system. In descending order of importance, this thesis suggests measuring, time-synchronized irradiance, windspeed, precipitation and pressure.

Irradiance, as explored in the previous section on POA irradiance, will greatly influence both power output and temperature. Module temperature and power output has a feedback loop, in the short and long term. In the short term there is a correlation between module temperature and the power output of the system. In the long term there will be system degradation that leads to lower efficiency of the system, which in turn decreases power output and thus dissipates energy as heat in the system instead. Therefore, there should be data logging of at least one kind of irradiance for the system to complement module temperature.

Measuring the temperature on-site is necessary, however one must decide which type of temperature reading one wants to measure and possibly model.

Furthermore, temperature could also be complemented by measurements of wind, precipitation and pressure, to obtain a more precise model. Of the aforenoted parameters, wind has a more obvious relation with power output, by its relation to module temperature. These parameters may grant additional explanations for circumstances where temperature and irradiance cannot alone explain the behaviour of the system. One may assume a system deviation if neither a combination of these parameters explains the behaviour measured.

The final parameter that could be relevant to evaluate is the efficiency of the system. Efficiency is generally some ratio between output and input. Since the input energy density and output power is available, it may seem tempting to use efficiency as a tool to evaluate system performance. However, this is too early to be useful at this stage in the process of understanding O&M for solar PV. There are too many unexplored conditions and parameters to accurately determine an efficiency that will aid the analysis process. With the current status, efficiency may be a useful parameter in technical terms, however not for coordinating operation of the PV system.

Temperature can be used both as a parameter to aid modelling of system behaviour and as an indicator of degradation. If the owners wish to perform analysis of degradation, there will likely be relevant to study a combination of ambient temperature, module temperature, and cell temperature. This further complicates the issue, but it will lead to more robust results for degradation analysis.
Otherwise, it will be sufficient to measure either the module temperature or the cell temperature to simply increase performance through necessary O&M measures.

5.1.3. Size of periods analysed will affect the result obtained

Plots for correlation between two parameters were shown for all points from four different months in 2012. Choosing a longer time period like a month may hide correlations for a single incident, since added data points for the entire month that does not necessarily coincide with that incident. The following paragraphs explore three topics that highlights why researches must consciously choose an appropriate time period for analysis.

One advantage of analysing the entire period simultaneously is that it provides a trend overview for the entire period. Constant correlations will be highlighted, since the relations will always be true to a degree, due to physical relationship or logical relations. For example, temperature and GHI will always retain a correlation, since sunlight provides heat to the surface. Unlucky researchers could study this correlation for one day with extraordinary conditions where such a correlation between these two parameters may not exist. Therefore, analysing large data sets to discover correlations that are true over time may be advantageous.

One disadvantage of analysing large time periods is natural seasonal changes affect the power output. This will further obfuscate the correlation between parameters. Seasonal changes in input parameters are one of the reasons that solar PV is hard to analyse, since analysing data from different seasons will have different input parameters. Parameters may also be different from random weather changes on the same day in two different years.

On the other hand, the advantage of analysing select intervals within one period is that greater insight into single incidents or events may result. For example, problems that are caused by snowfall may only affect the module for a short time. In this case studying parameters for exactly that time period in order to clearly evaluate the correlation is preferable. Researcher or operators must remember that the values obtained must only be evaluated for the selected period.

5.1.4. Applications and Limitations of correlations found

Two sets of correlations of visualizations were used for the analysis: one large matrix containing every parameter and several small thumbnails plotting two parameters together.

The matrix is a crude and exploratory model that should not be used to make conclusions. Instead, it helps researchers evaluate the correlation between a wide range of parameters of interest. For this thesis two sets of correlation matrices containing all parameters available were used. The first matrix applied all data points to calculate the total correlation, while the other plotted the correlation between the data points that had been flagged. This way there was gained insights into which parameters correlated when the system was underperforming. The difference between the two correlation values were also subtracted from each other to see which correlation changed the most.

Correlation thumbnails provided the ability to analyse correlation between two parameters of interest. The process of selecting periods to analyse through the interactive visualizations will affect the correlation plots, which naturally leads to possible problems of selection bias if the values are translated to other periods where they are not applicable. R-squared was used to evaluate how well a given correlation explained the relation between two selected parameters.

5.2. Strength and weakness of OEE as applied to PV

5.2.1. Discission of definition of availability, performance and quality

As discussed in section 5.1.3 – “Temperature as a parameter and indicator”, evaluating efficiency might be too complex at this stage. One solution might be to nuance the issue by using OEE instead of simple efficiency. This introduces “availability”, “performance” and “quality” to quantify which parts of the production is responsible for the reduced effectiveness. OEE are designed for manufacturing productivity, and is therefore not perfectly applicable to solar PV. The effort made in this thesis defined the terms in – “Quantification of OEE parameters and indicators”. Transferring the definitions from one manufacturing proved both challenging and rewarding:

Availability could be defined as the uptime of the system, which is similar to the manufacturing, and the results successfully flag periods without production or measurement. Unplanned stops in production can be exemplified by snowfall covering the module or measurement equipment.

Performance could be defined as the ability of the parts of the PV system to operate at their best ability. Unfortunately, this term is close related to efficiency, which has already been ruled out for future work. Performance in this thesis were defined as the ratio between the minimum expected power and the measured power. This definition is only partly sufficient for the definition of “performance” in traditional OEE. It is, however, defendable for the sake of evaluating if the system is producing power at the expected level.

Quality is normally defined as the quality of the production units. The ratio is between the approved units and the total units. Analysing to what extent power produced is approved is not easy. It has not been the scope of this thesis. The definition presented will instead be that quality is measured by the amount of time the yield from the system is within the expectancy. Whenever it is not, the model is wrong, or the efficiency is decreased for some unknown reason. Quality is probably the definition that fits PV the least. It is simply an attempt at transferring the approach from OEE to another technology. Furthermore, quality is also the indicator that scores the lowest for the entire PV system.

One advantage of differentiating between availability, performance and quality is that a low value can lead to a different response for each case. Availability issues must be dealt with immediately. Performance issues are visible over time and so must be dealt with when necessary. Quality issues lead the operators to evaluate their models and methodology over the long term.

One limitation of this method for defining the OEE indicators is that performance and quality depend on the uncertainty chosen by the researchers. For this thesis, the size of the uncertainty band was two standard deviations from the expectancy. If the uncertainty band is increased, it will encapsulate more data points, which gives an improved performance and quality score. Definitions are tweaked/tuned along the way, as learning grew.

As mentioned in 4.4 – "Quantification of OEE parameters and indicators", efficiency might be too simple to explain PV behaviour. OEE might also be insufficient, as the results show. Defining quality is challenging when the production units could not be visually inspected or tested. Performance has a debatable definition, so it could have different definition. These are some of the weaknesses that must be considered when transferring OEE concepts to solar PV.

5.2.2. Inaccurate system specifications

Inaccurate system specifications are a general issue for specifications that are backed by warranties. In our specific case, further analysis was needed to handle this issue. Results regarding maximum power output show that the inverter specifications listed on Methodology about GEC's website is not equal. Plant 3 using ABB inverters has allegedly peak output at 86 kW while the results for maximum power output suggests the maximum is 88 kW. This leads to at least two problems:

1. It could mean that the PV system is outsized by its specifications. This could potentially make it challenging to determine the original maximum power output when there are no records of the true maximum output.
2. If this is an ongoing issue, it may mean that other specifications also have different dimensions than suggested.

Typically, PV modules operate with plus tolerances for warranty purposes, and as such the initial performance may be higher than the datasheets.

Likewise, the inverter maximum capacity has been reached for Plant 1, but that maximum is below the suggested maximum. GEC's website claims the system have capacity for 4.6 kW, but the maximum measured is 3.9 kW. The graphs clearly show an inverter cut-off around 4 kW, which remains a constant maximum that may be reached each day. The inverter may have too low capacity to handle the power output of the modules. Unfortunately, this leads to at least two challenges:

1. There is difficult to determine to which extent a system has degraded when the maximum capacity of the module exceeds the inverter maximum.
2. The total power output of the system and its signature might be affected by the maximum measured power being lower than the true maximum generated power. For example, by temporary heat dissipating at inverter cut-off, causing a different thermal operational point.

The challenges presented means that researchers must employ more sophisticated methods to determine the amount of degradation of the system. Maximum power output could be measured in months where the peak capacity is not reached on any day. The maximum power could later be compared to power from similar days to evaluate performance, if other parameters are similar enough to allow it.

5.3. Indications of forewarning before measuring equipment shutdown

During the analysis there was analysed if a trained human mind could identify the reason for a failure of the measuring system. The two examples show what the irradiance appears on the heatmap before the failure. The hypothesis was that a failure should occur due to a cause, and that this cause could be observed as a visual signature from studying the data.

The first example, figure XX, show how there are an increasing amount of missing data in the months before the failure. The second example does not show this effect on the heatmap, although the heatmap appears to have more extreme values than other equal months. The expected signature of the data set is not self-evident if it truly exists.

The hypothesis should be studied further. Enabling operators to stop failures of the measuring equipment will greatly improve the data quality, and thus improve reliability and robustness of the results.

5.3.1. Possibility of detecting signature in data after snowfall

Another hypothesis during this research was that there would be a signature in the data set for how snow affected the irradiance measurement and power output. The hypothesis is not directly tied to the research questions after a shift in perspective, but the results may yet be of future interest. There were confirmed that irradiance data combined with satellite imagery can detect snowfall. Irradiance measurements from the system in question are, fortunately, not affected by snow after the clouds have cleared. The time periods that were identified to have snowfall in this research unfortunately did not have enough available data to compare the insights with power output.

It is entirely possible that there is a signature to snow on modules, but it is not self-evident by visually inspecting only irradiance alone.

This research has identified approaches to confirm snowfall within a time period, and that irradiance measurements are not affected by the snow. This insight may be used further for other periods with snowfall where data for power production is available. With this approach there might be gained further insight into the behaviour of the model after snowfall.

It should be noted, however, that there is some uncertainty in the consensus of the effects of snow on modules. One study suggests that the losses are dependent on the angle of module and the technology of that module (Andrews, Pollard et al. 2013). Therefore, there must be conducted further research into this topic.

5.3.2. Machine learning

One final point that could be considered it a machine learning approach to this issue. It can be argued that data science and machine learning could assist in evaluating the large data set. The results discussed provide several examples for the difficulty of detecting significant values or significant incidents in large data sets.

There is a clear benefit to have multiple ways of analysing data, for example by having different people cooperate and discuss findings. Additionally, there are benefits to have a machine teaching itself how to analyse the large sets efficiently. Analysis of the data does not necessarily demand human attention, although it is a safety mechanism at this early stage of understanding O&M. Some studies suggest that a combination of human comprehension and computer algorithm can derive understanding of how environmental stressors affect degradation in PV modules (Wheeler, Gok et al. 2015).

One weakness of using data scripts for analysis is that several of the instances of degradation happen seldom to the modules. Machine learning can be weak in discovering the causation between output and rare phenomena. It must be guided by human intervention so that the insights learned may be transferred to human knowledge. By for example, choosing time periods where snowfall has been confirmed, the algorithm can be guided to analyse the data set in that context.

Another weakness of using data science directly is that the complex interactions are not yet fully understood in the scientific community. This means that the algorithm can be trained based off of inaccurate or wrong assumptions. This may lead the algorithm to make inaccurate or wrong conclusions. Additionally, machine learning should provide scientists and engineers a greater understanding of the failure mechanisms. The insights gained will be independently beneficial, but understanding the conditions that cause the degradation may allow operators to completely nullify the issue.

5.4. Quality control and assessment

Generally, data should be validated by independent sources to give the results more weight. That was done partly in this paper. For example, to verify data points regarding solar irradiance, there was analysed the measured irradiance on the module from different angles, the power induced by the irradiance in the cell, and satellite imagery to confirm weather conditions. This is only the next-best thing and can potentially lead us to make claims based on bad data sets. In addition, there was taken care of in this paper to quality control the data sets based off human comprehension of what a reasonable value would be.

5.4.1. False positive measurements

Another point that should be considered in analysing data is that some corrupt data will likely be corrupted without being noticed. It might be skewed by a couple of minutes or raised without being flagged as corrupt. This might happen because it is not severe enough to be noticed, or because its self-correcting from a corrupt value to a reasonable, yet wrong, value. False positives will always be a part of data sets of this size. False positives might indicate that a system has a higher OEE than it does, which falsely leads to a suboptimal O&M. If can also disguise root causes that would otherwise indicate an upcoming failure.

5.4.2. Real-time O&M improves system and quality of data

Proper data quality control, as described in chapter 2.9 in theoretical basis, will be essential to provide trusted and reliable results. There are numerous examples of entire months of data that is unable to be used for analysis due to missing or corrupt data. Analysis of data from up to eight years ago might not be relevant now for the operators, but it would provide this research project with more data for analysis.

O&M should be performed in real-time for mainly two reasons: Firstly, this enables operators to rectify issues so that the least amount of power is lost. Secondly, it can protect the module from cascading degradation modes. In addition, the results obtained in this thesis shows that O&M should also be performed in real-time to preserve data quality. Analysis of the state of the system and prediction for its future life will be greatly enhanced by improved uptime of the system.

Likewise, monitoring and analysis should also be done in real-time for the system. That way there can be performed visual inspection in combination with the analysis of operation behaviour. The analysis in this thesis had, in comparison, an academic approach. The aim was not necessarily to address the system as presented by the data, but instead to explore the data set. Thus, there was some leniency to use old data sets.

Pearsall claims that there is not sufficient to look at select times for performance, since natural variations may change the output. He also claims that yearly averages do not necessarily provide enough context unless you understand the year in question (Pearsall 2017). These aspects were attempted overcomed by in-depth analysis of the data set and is part of our delivery.

5.4.3. Resolution of data

The chosen resolution of data will affect the results by changing the data points that are analysed. There was made a conscious decision to obtain data sets with high resolution, so that there could be performed precise analysis on changes of module behaviour under different weather circumstances. The resolution enabled closer and more precise correlation estimation between phenomena and module behaviour. This is also in accordance with Hansen et al. who found that reducing weather-averaging from one hour to 15 minutes reduces the error in energy by a factor of 10 (Hansen, Stein et al. 2012). The thesis also combined data from several sources. If the sources have different sampling rate, the intermediate points from one source must be dismissed since they do not correspond to a point from the other source.

However, a high resolution of a data set demands much more processing power and time. One of the simulations done in this project demanded 10 minutes to visually display the heat map for each month, which equals less than 10% of the entire data source. Measurement equipment able to make log data with higher quality and sampling rate will also be more expensive.

Additionally, data analysis for operational data for PV systems are generally at much lower resolution than the data set evaluated in this thesis. When the research in this thesis used a higher resolution there is possible to detect known issues, as well as unknown issues. One example is that temperature lags behind shifting weather on a small timescale.

Data were down sampled to another frequency for the latest interactive produced reports. This means that precision was sacrificed to make it possible to use the GUI in the web browser. Sampling for a new frequency was also done to match the frequencies of different sources with unequal frequencies. While this is useful for slicing and studying data quickly, it will reduce the accuracy drastically.

5.4.4. Methodology based on incomplete models may yield incomplete results

One issue that should be noted is that our methodology is largely influenced by three factors: the data set available, our understanding of solar modules, and the academically suggested approach. Our understanding of the influence of O&M on solar modules is incomplete, which in turn guides us to make imperfect assumptions when analysing the data set. We strive to make the analysis as rigid as possible, but effects outside our understanding will likely not be investigated, simply because its importance is unknown. In conclusion, this makes the results determined by the understanding of the problem.

Another relevant effect is that the information available determines how data is collected by the operators and researchers. For operators, this means that they will only notice effects that their measuring equipment is designed to measure. Likewise, scientists looking into the effects on the modules will only notice what is known to be an issue. The combination of these two effects can lead to a narrow-focused approach that in turn leads to incomplete descriptions of module degradation. Our analysis of snow affecting the modules are one example of incomplete understanding yielding incomplete answers. Further interdisciplinary studies are necessary.

5.5. Weaknesses of methodology used

One limitation for the methods is that there are room for human error in the processes of figuring out how to approach the data set. Several instances of perceived relationships were in fact human mistakes in the programming, extraction or analysis phase. Avoiding misunderstanding requires resources, i.e. time and effort, which could otherwise have been spent on meaningful analysis.

Another weakness of the approach is that the research was exploratory. There is hard make bold conclusions when a majority of the research is investigation and discovery of opportunities. These opportunities may, however, be used as a foundation for future research. Additionally, exploration of a data set with many variables leads to an increasing chance of randomness to affect the data set.

Furthermore, even though there were multiple technologies analysed and several plants, there were only one geographical location in question. One single location limits the possibility of generalizing findings or insights to other microclimates.

OEE is a concept that is transferred from one technology to another. Adopting the three main indicators proved challenging, and using them may be an inaccurate model that could lead to inaccurate conclusions due to inaccurate or wrong assumptions.

6. Conclusion

The research aimed to identify indicators and parameters that would necessary to evaluate the state of a solar PV system. Through an exploratory analysis of the operational data from a grid-connected solar PV system in Sweden there were identified two parameters, namely POA irradiance and temperature, that contributed the greatest to evaluate the state of the system. Both parameters should be evaluated in combination with other parameters to create the most robust predictions for the future of the system.

The thesis also introduced OEE as a tool to benchmark the PV system. There are proposed definitions that are transformations from the manufacturing industry. The range of availability, performance and quality have been presented, and the indicators and OEE in general show some promise as a tool to evaluate the state of the solar PV system.

From the research there were made observations regarding behaviour that may be a forewarning for equipment failure. There may be a signature in the data set that could prove useful to prevent a failure from decreasing the quality of the data set, so that the reliability and robustness of analysis could be improved. The insights gained by OEE were beneficial on their own terms, and more importantly, system understanding enables operators to reduce or even eliminate issues, and the potential value of humans-in-the-loop for O&M data analysis were illustrated.

The research also aimed to evaluate the quality of datasets and the reliability of the results. There were presented a wide range of instances with low quality in the data set, which have been separated from the analysis. There were also made several considerations regarding the quality assurance relevant to the research. Additionally, there were made considerations for how this research methodologically could be improved.

Finally, the research has identified several necessities that must be satisfied the first year of operation to ensure that the data set is satisfactorily detailed and quality assessed. Having complete data sets are especially important to perform meaningful analysis of the behaviour of the system. These recommendations stemmed from observations that were made throughout the research.

6.1. Future work

This paper is far from comprehensive, and there are obvious issues that needs to be tackled to improve the understanding of O&M and its effects on the lifetime of solar modules. This study aims to provide a system condition assessment. The natural progression from this understanding would be to provide a recommendation of O&M-measures for better OEE and lifetime, by continuing to pursue CRISP-DM as a tool to analyse solar PV systems.

Below follows a suggestion for future work. This chapter does not, naturally, completely cover the extent of unexplored topics in this field, but will be based on results, ideas and perspectives gained throughout the research for this thesis.

6.1.1. Using Machine Learning to assist human evaluation of degradation

Machine learning may provide another tool to aid humans in evaluating degradation in solar PV. The data sets are large, and algorithms could provide new insights, if the process is appropriately monitored and studied. This thesis has suggested snowfall as one weather phenomenon that can be analysed further.

6.1.2. Analysis of materials used in the modules

The material that the solar cell is constructed from will have significant influence on how the environmental stressors affect the module. To ensure that results gathered from this research is generalizable, the topic must be researched further in combination with an analysis of materials science. This is beyond the expertise of the student and will be beyond the scope of this paper.

6.1.3. Economic aspects

The economic aspect has not been a part of the analysis of this paper. It should, however, be a concern for future research and discussion. The motivation for improved O&M is partly financial competitiveness, and it must be evaluated thoroughly the levels of O&M that is the most profitable. This will likely change over time, due to development of technology and improved knowledge. Two proposed strategies could involve:

- Evaluating the spot price for kWh and then compare it to the cost of O&M

- Evaluation of market opportunities by comparing amount of possible O&M in one market and the potential to create business opportunities for that market

6.1.4. Missing parameters

This analysis could be enhanced by having more parameters. Especially voltage and current could have been used in this thesis to improve the analysis. Analysing how much voltage was generated by the cells could give insights into the correlation between temperature and voltage and evaluating the ratio between voltage and current could give new insights into power generation as well as another estimate of the maximum power generated. It may also lead to more robust understanding of where the reactive power stems from.

Reference list

Andrews, R., A. Pollard and J. Pearce (2013). "The Effects of Snowfall on Solar Photovoltaic Performance." Solar Energy **92**: 84-97.
Azizi, A., P.-O. Logerais, A. Omeiri, A. Amiar, A. Charki, O. Riou, F. Delaleux and J.-F. Durastanti (2018). "Impact of the aging of a photovoltaic module on the performance of a grid-connected system." Solar Energy **174**(1): 445-454.
BP plc (2019). BP Statistical Review of World Energy 2018. BP Statistical Review of World Energy, BP. **68th:** 9.
Brown, M. S. (2015). "What IT Needs To Know About The Data Mining Process." Retrieved 21.06.2020, from https://www.forbes.com/sites/metabrown/2015/07/29/what-it-needs-to-know-about-the-data-mining-process/#2065f3a3515f.
Chattopadhyay, S., R. Dubey, V. Kuthanazhi, J. J. John, C. S. Solanki, A. Kottantharayil, B. M. Arora, K. L. Narasimhan, V. Kuber, J. Vasi, A. Kumar and O. S. Sastry (2014). "Visual Degradation in Field-Aged Crystalline Silicon PV Modules in India and Correlation With Electrical Degradation." IEEE JOURNAL OF PHOTOVOLTAICS **4**(6): 1470-1476.
Colli, A. (2015). Failure mode and effect analysis for photovoltaic systems. Renewable and Sustainable Energy Reviews. **50:** 804-809.
de Oliveira, M. C. C., A. S. A. D. Cardoso, M. M. Viana and V. d. F. C. Lins (2017). "The causes and effects of degradation of encapsulant ethylene vinyl acetate copolymer (EVA) in crystalline silicon photovoltaic modules: A review." Renewable and Sustainable Energy Reviews **81**(8): 2299-2317.
EcoSmart. (s.a.). "PV System Performance: GHI to PoA." Retrieved 26.05.2020, from https://ecosmartsun.com/pv-system-performance-3/pv-system-performance-ghi-to-poa/.
Engineering ToolBox. (2005). "AC - Active, Reactive and Apparent Power." Retrieved 10.06.2020, from https://www.engineeringtoolbox.com/kva-reactive-d_886.html.
Fu, R., D. Feldman and R. Margolis. (2018). "U.S. Solar Photovoltaic System Cost Benchmark: Q1 2018." National Renewable Energy Laboratory. NREL/TP-6A20-72399 Retrieved 05.06.2020, from https://www.nrel.gov/docs/fy19osti/72399.pdf.
Gebhardt, P., L. P. Bauermann and D. Philipp. (2018). "BACKSHEET CHALKING - THEORETICAL BACKGROUND AND RELATION TO BASKHEET CRACKING AND INSUALTION FAILURES." 35th European PV Solar Energy Conference and Exhibition Retrieved 12.05.2020, from https://www.ise.fraunhofer.de/content/dam/ise/de/documents/publications/conference-paper/35-eupvsec-2018/Gebhardt_5DO75.pdf.
Geisz, J. F., M. A. Steiner, N. Jain, K. L. Schulte, R. M. France, W. E. McMahon, E. E. Perl and D. J. Friedman (2017). "Building a Six-Junction Inverted Metamorphic Concentrator Solar Cell." IEEE Journal of Photovoltaics **8**(2): 626-632.
Glava Energy Center. (s.a.). "Glava Energy Center - Technologies: Solar Energy." Retrieved 10.06.2020, from https://www.glavaenergycenter.se/en/technologies/technologies-solar-energy/.
Google Maps. (s.a.). "Glava." Retrieved 10.06.2020, from https://www.google.com/maps/place/670+20+Glava,+Sverige/@59.5320103,8.1100132,6z/data=!4m5!3m4!1s0x4643658337955b7b:0x6246c49d922297ca!8m2!3d59.5504211!4d12.562273.
Hansen, C., J. Stein and D. Riley (2012). Effect of Time Scale on Analysis of PV System Performance.
Hatti, M. (2014). Operation and Maintenance Methods in Solar Power Plants. Use, Operation and Maintenance of Renewable Energy Systems. M. A. Sanz-Bobi. Madrid,Spain, Springer**:** 61-93.
Heymi Bahar, D. D'Ambrosio, R. M. & and H. Paillere. (2019). "Tracking Power." Retrieved 07.01.2020, from https://www.iea.org/reports/tracking-power-2019.
Hofstad, K. (2019). "Solenergi." Retrieved 28.01.2020.
International Energy Agency (2014). Performance and Reliability of Photovoltaic Systems. IEA-PVPS**:** 58-59.
International Energy Agency. (2019). "Electricity production from oil, gas and coal sources (% of total)." Retrieved 16.01.2020, from https://data.worldbank.org/indicator/EG.ELC.FOSL.ZS.

Jordan, D. C. and S. R. Kurtz (2012). "Photovoltaic Degradation Rates — An Analytical Review " Progress in Photovoltaics: Research and Applications: 5.
Jäger-Waldau, A. (2019). PV Status Report 2019. PV Status Report. Luxemburg, Publications Office of the European Union: p 1.
Järvelä, M., K. Lappalainen and S. Valkealahti (2020). "Characteristics of the cloud enhancement phenomenon and PV power plants." Solar Energy **196**(January 2020): 137-145.
Kempe, M. D., G. J. Jorgensen, K. M. Terwilliger, T. J. McMahon, C. E. Kennedy and T. T. Borek (2006). "Acetic acid production and glass transition concerns with ethylene-vinyl acetate used in photovoltaic devices." Solar Energy Materials & Solar Cells **91**(4): 315-329.
Klise K.A. and S. J.S. (2016). "Performance Monitoring using Pecos, Technical Report." Retrieved 24.05.2020, from https://pecos.readthedocs.io/en/latest/metrics.html#quality-control-index.
Kottek, M., J. Grieser, C. Beck, B. Rudolf and F. Rubel (2006). "World Map of the Köppen-Geiger climate classification updated." Meteorologische Zeitschrift **15**(3): 259-263.
Köntges, M., S. Kajari-Schröder and U. Jahn (2011). "CRACK STATISTIC OF CRYSTALLINE SILICON PHOTOVOLTAIC MODULES." European Photovoltaic Solar Energy Conference **26th**(_): 1-5.
Köntges, M., G. Oreski, U. Jahn, M. Herz, P. Hacke, K.-A. Weiss, G. Razongles, M. Paggi, D. Parlevliet and T. Tanahashi (2017). Assessment of Photovoltaic Module Failures in the Field. Report IEA-PVPS: p 33-36.
Lyu, Y., A. Fairbrother, M. Gong, J. H. Kim, A. Hauser, G. O'Brien and X. Gu. (2020). "Drivers for the cracking of multilayer polyamide-based backsheets in field photovoltaic modules: In-depth degradation mapping analysis." Retrieved 11.05.2020.
Muñoz-Cerón, E., J. C. Lomas, J. Aguilera and J. de la Casa (2018). "Influence of Operation and Maintenance expenditures in the feasibility of photovoltaic projects: The case of a tracking pv plant in Spain." Energy Policy **121**: 506-518.
National Rewewable Energy Laboratory. (2016). "Photovoltaic Lifetime Project." Retrieved 03.02.2020, 2020, from https://www.nrel.gov/pv/lifetime.html.
Parhami, B. (1997). Defect, Fault, Error, ... , or Failure? IEEE Transactions on Reliability University of California. **Vol. 46:** 450-451.
Pearsall, N. (2017). Prediction and measurement of photovoltaic system energy yield. The Performance of Photovoltaic (PV) Systems; Modelling, Measurement and Assessment N. Pearsall. Newcastle, United Kingdom, Woodhead Publishing Series in Energy: 183-206.
Peel, M. C., B. L. Finlayson and T. A. McMahon (2007). "Updated world map of the Koppen-Geiger climate classification." Hydrol. Earth Syst. Sci(11): 1633-1644.
Pei, T. and X. Hao (2019). "A Fault Detection Method for Photovoltaic Systems Based on Voltage and Current Observation and Evaluation." Energies **12**(9): 1-15.
Silva, J. P., J. L. Balanzategui, L. Martín-Pomares, S. Wilbert and J. Polo (2019). Quality Assurance of Solar Radiation Measurements. Solar Resource Mapping. J. Polo, L. Martín-Pomares and A. Sanfilippo. Madrid, Spain, Springer: 99-137.
SolarPower Europe (2019). O&M Best Practice Guidelines Version 4.0. O&M Best Practice Guidelines, SolarPower Europe: p 12.
Tsao, J. L., N. and G. Crabtree (2006). "Solar FAQs." 9-12.
van Dyk, E. E., J. B. Chamel and A. R. Gxasheka (2004). "Investigation of delamination in an edge-defined film-fed growth photovoltaic module." Solar Energy Materials and Solar Cells **8**(4): 403-411.
Vashishtha, S. (2012). "DIFFERENTIATE BETWEEN THE DNI, DHI AND GHI?" Retrieved 26.05.2020, from https://firstgreenconsulting.wordpress.com/2012/04/26/differentiate-between-the-dni-dhi-and-ghi/.
Vorhies, W. (2016). "CRISP-DM – a Standard Methodology to Ensure a Good Outcome." Retrieved 21.06.2020, from https://www.datasciencecentral.com/profiles/blogs/crisp-dm-a-standard-methodology-to-ensure-a-good-outcome.
Vorne Company. (s.a.). "WHAT IS OVERALL EQUIPMENT EFFECTIVENESS?" Retrieved 21.06.2020, from https://www.oee.com.
Wheeler, N., A. Gok, T. Peshek, L. Bruckman, N. Goel, D. Zabiyaka, C. Fagerholm, T. Dang, C. Alcantara, M. Terry and R. French (2015). A data science approach to understanding photovoltaic module degradation.

Wiesmeier, C., P. Huelsmann, M. Bleuml, P. Schmid, K.-A. Weiss and M. Köhl (2012). "Impact of permeation properties and backsheet-encapsulant interactions on the reliability of PV modules." Renewable Energy: DOI:10.5402/2012/459731.

Zehner, M., T. Weigl, S. Thaler, O. Schrank, M. Czakalla, B. Mayer, T. R. Betts, R. Gottschalg, K. F. A. Behrens, G. K. Langlo, B. Giesler, G. Becker and O. Mayer (2011). Energy Loss Due to Irradiance Enhancement

Blank page

Addendum

Attachment 1: Complete plots for power and clearsky irradiance

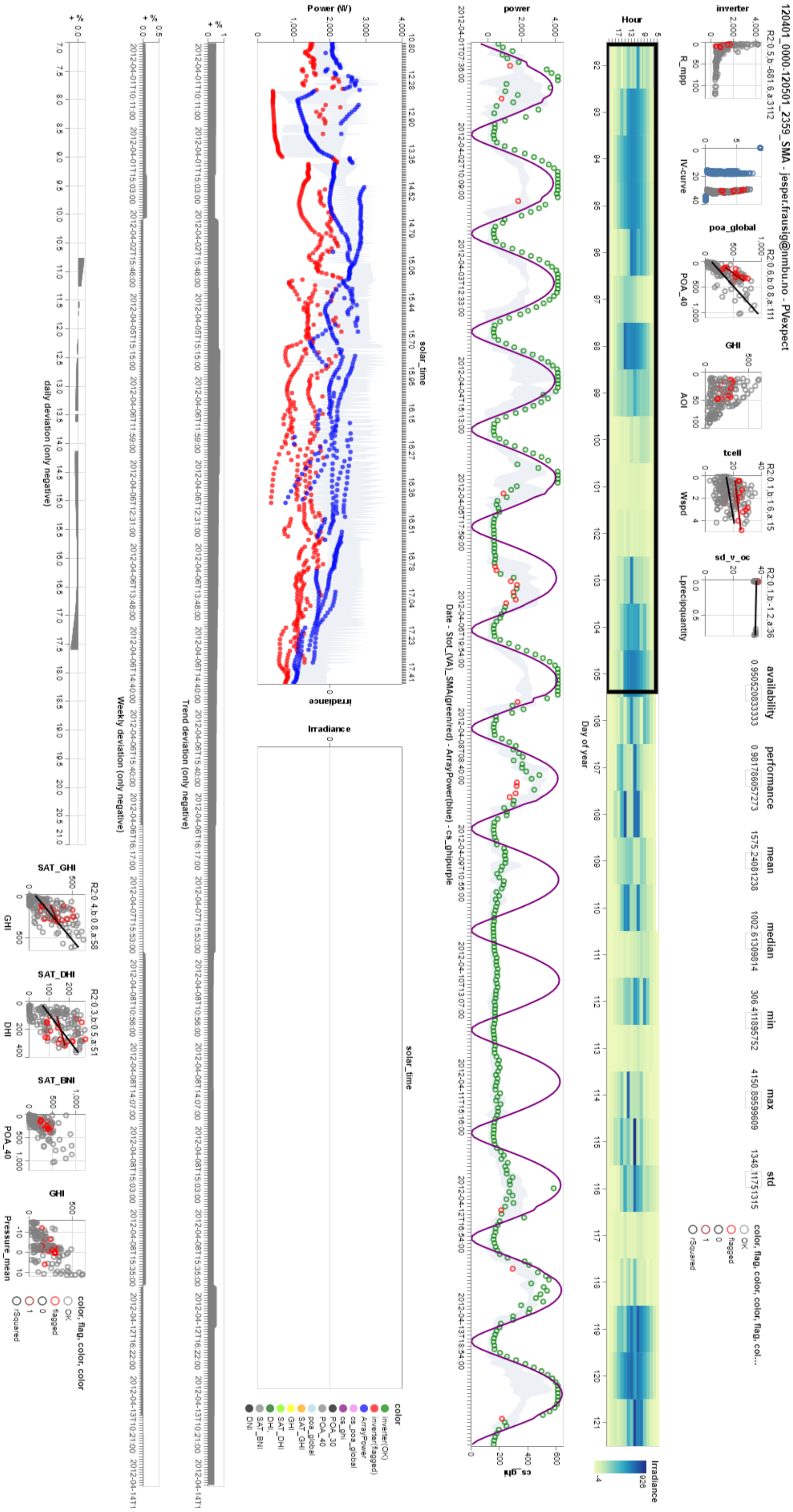


Figure 27: Supplementary plots to Figure XX. This shows the system using SMA-inverter.

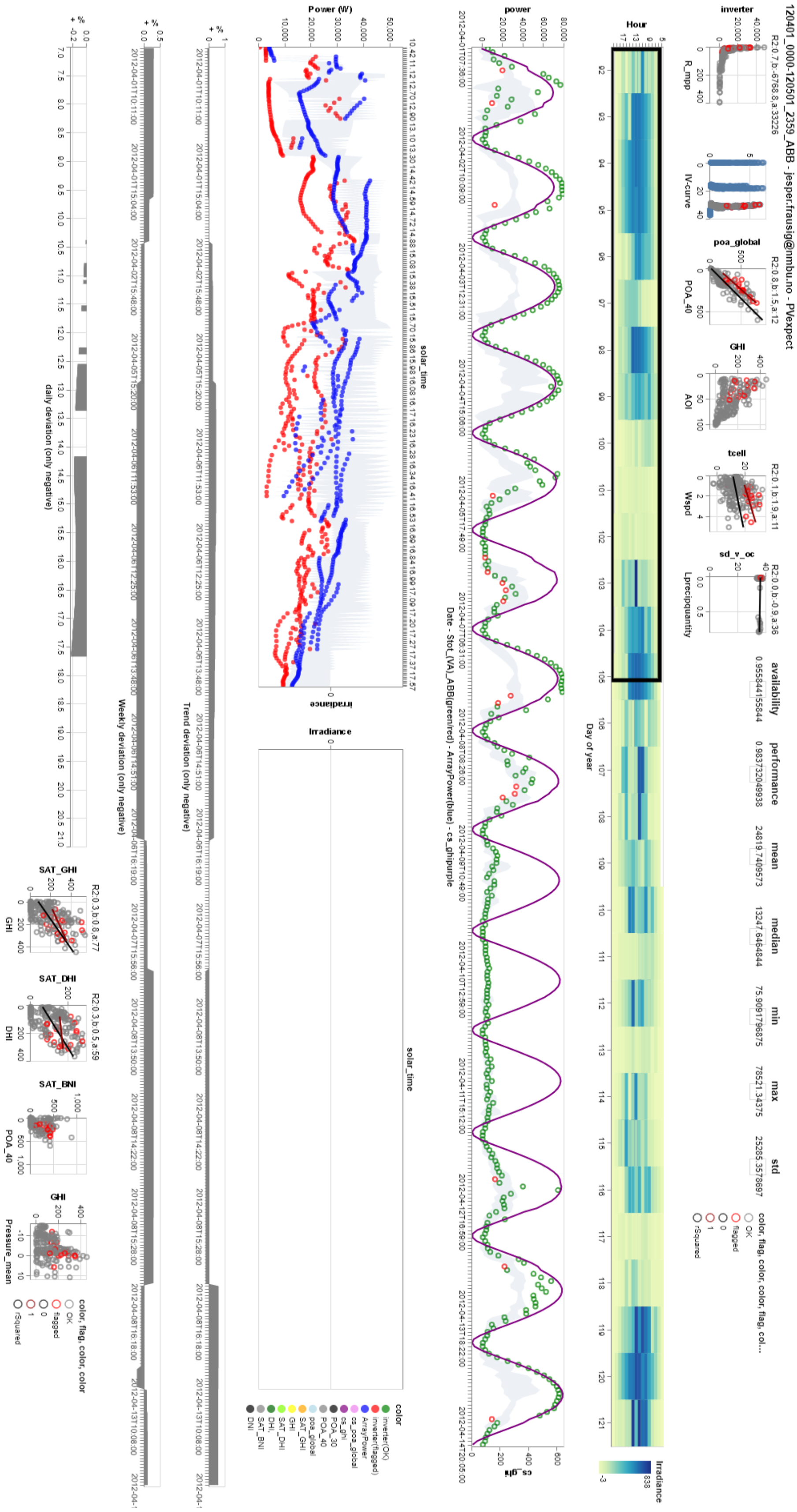


Figure 28: Supplementary plots to Figure XX. This shows system using ABB-inverter.

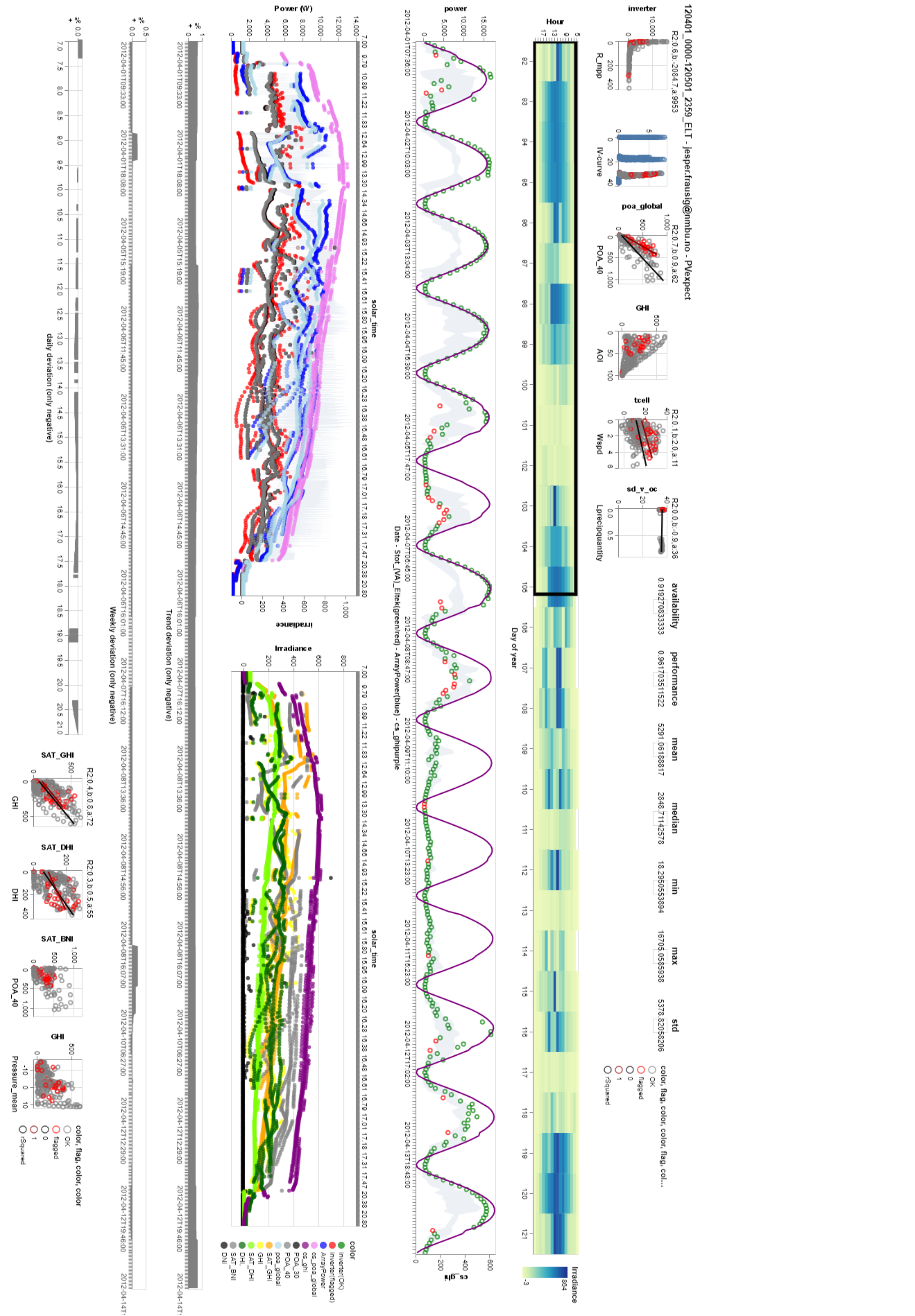
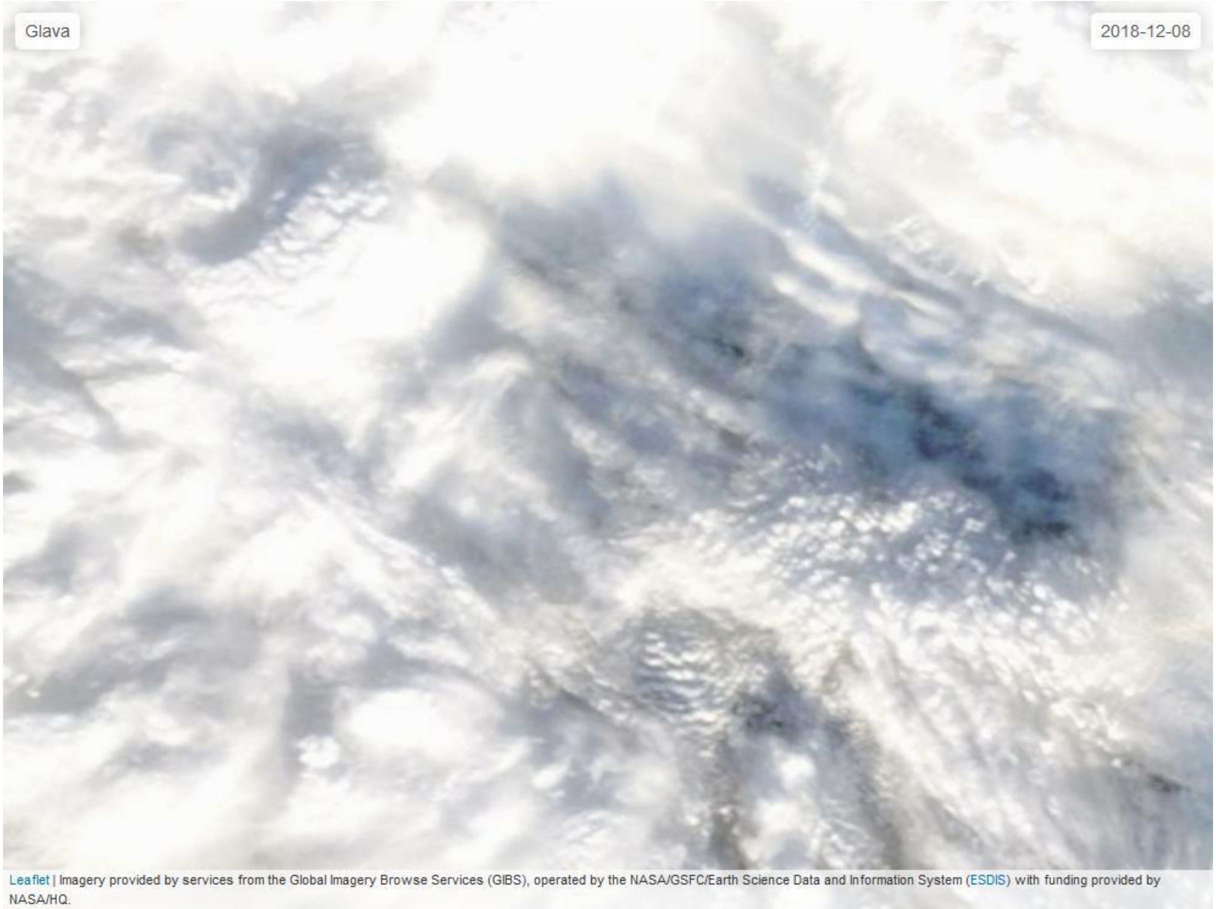
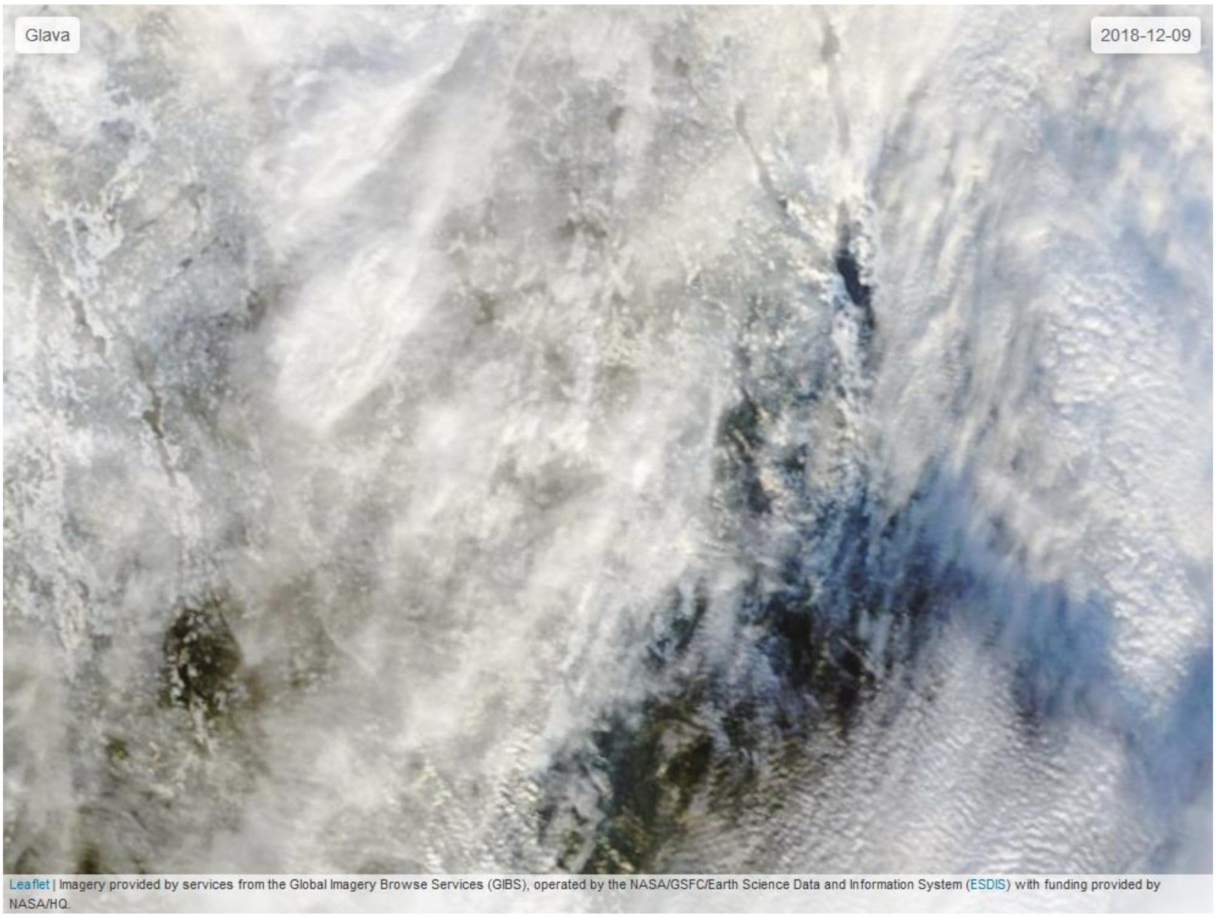


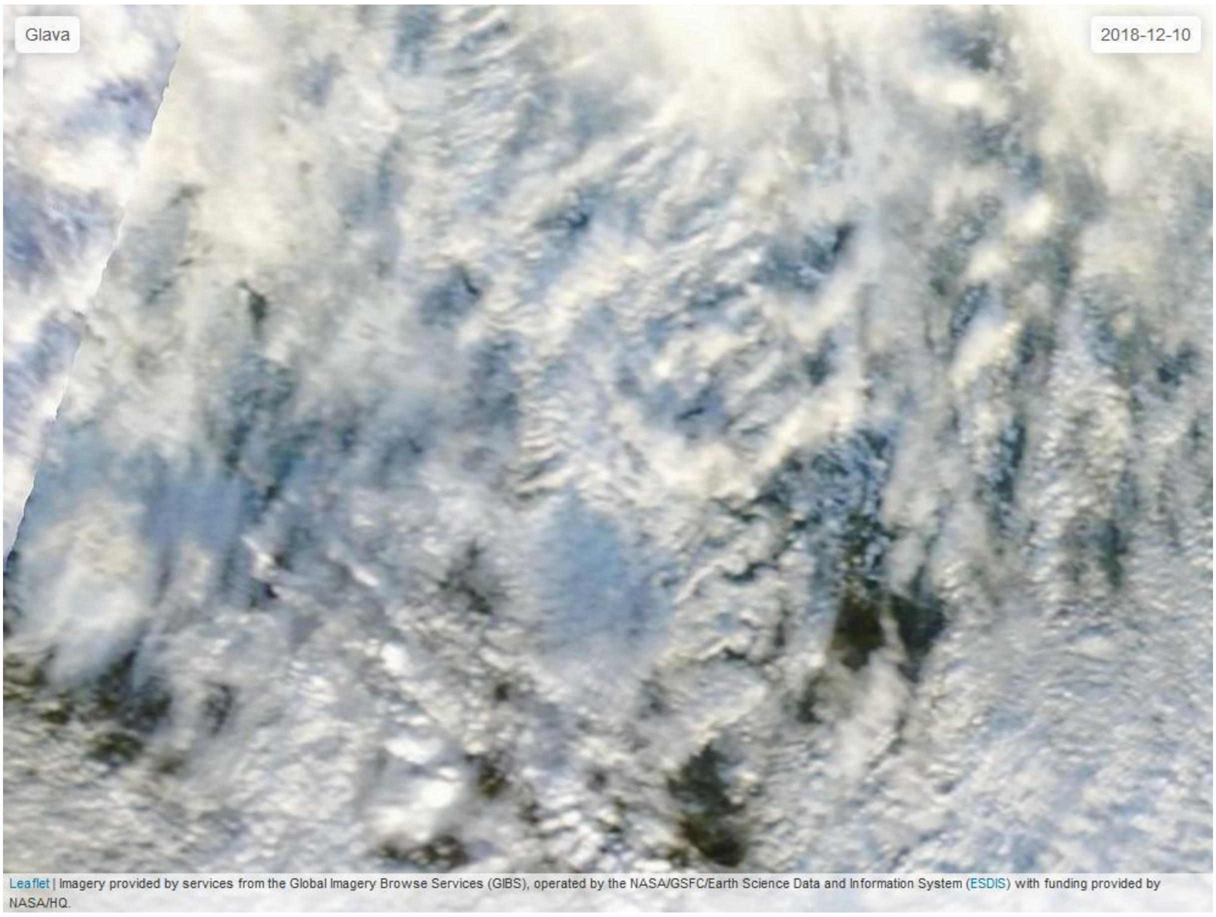
Figure 29: Supplementary plots to Figure XX. This shows the system using ELTEK inverter.

Attachment 2: Satellite images for December 5th-11th, 2018


Glava
2018-12-07
Leaflet | Imagery provided by services from the Global Imagery Browse Services (GIBS), operated by the NASA/GSFC/Earth Science Data and Information System (ESDIS) with funding provided by NASA/HQ.

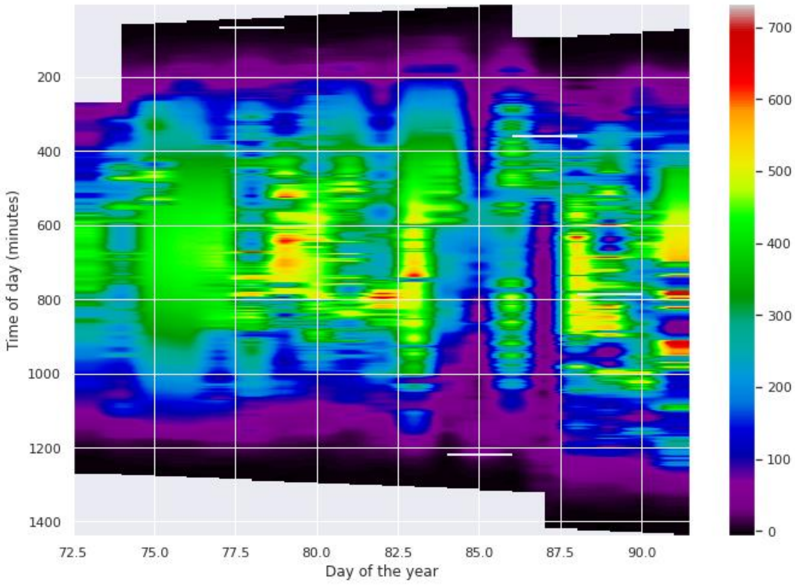

Glava
2018-12-08
Leaflet | Imagery provided by services from the Global Imagery Browse Services (GIBS), operated by the NASA/GSFC/Earth Science Data and Information System (ESDIS) with funding provided by NASA/HQ.

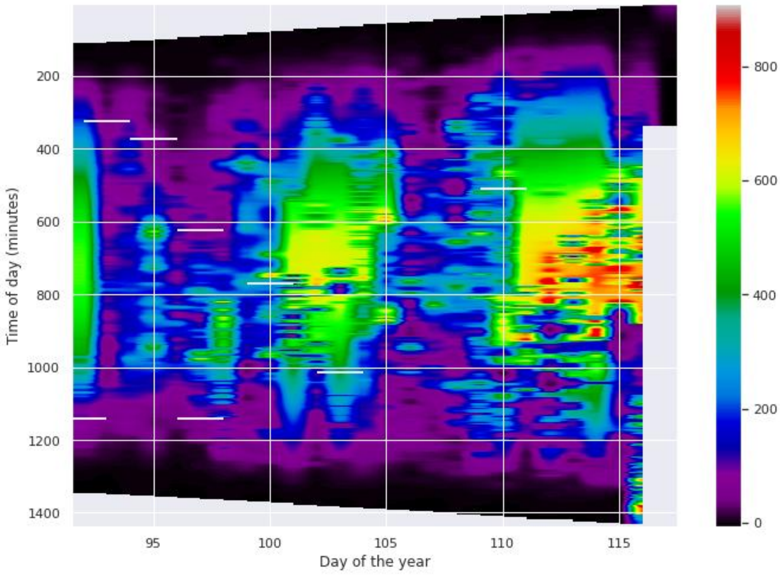

Glava
2018-12-09
Leaflet | Imagery provided by services from the Global Imagery Browse Services (GIBS), operated by the NASA/GSFC/Earth Science Data and Information System (ESDIS) with funding provided by NASA/HQ.

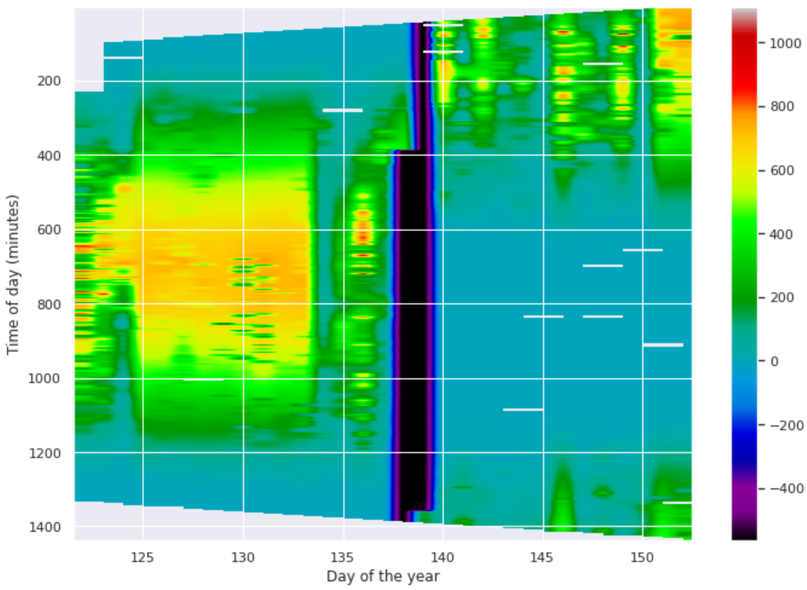

Glava
2018-12-10
Leaflet | Imagery provided by services from the Global Imagery Browse Services (GIBS), operated by the NASA/GSFC/Earth Science Data and Information System (ESDIS) with funding provided by NASA/HQ.


Glava
2018-12-11
Leaflet | Imagery provided by services from the Global Imagery Browse Services (GIBS), operated by the NASA/GSFC/Earth Science Data and Information System (ESDIS) with funding provided by NASA/HQ.

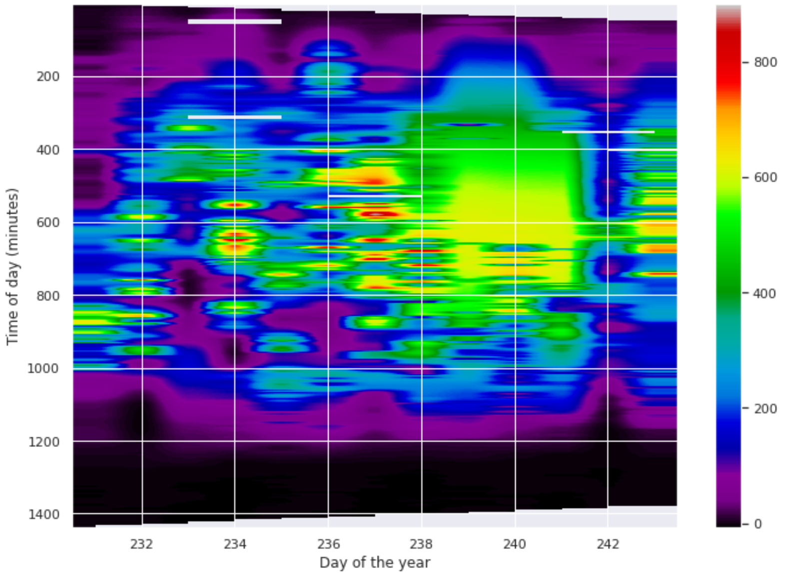
Attachment 3: Heatmaps for March, April, May 2016.

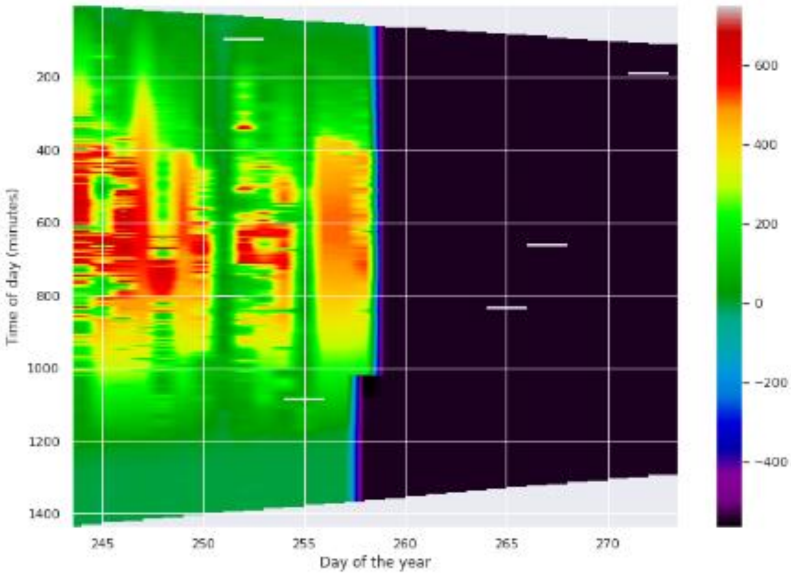


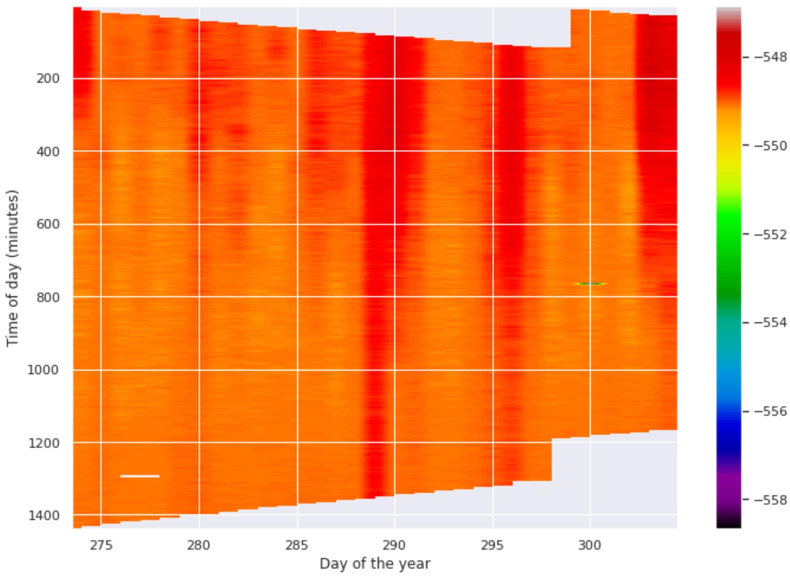




Attachment 4: Heatmaps for August, September, October 2014







Backpage.

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway