UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE GEOFÍSICA

SOLAR IRRADIANCE FORECASTS IN THE ATACAMA DESERT USING ANALOG ENSEMBLES AND A CLOUD ASSIMILATION TECHNIQUE

TESIS PARA OPTAR AL GRADO DE MAGISTER EN METEOROLOGÍA Y CLIMATOLOGÍA

MIRKO FRANCISCO ANDRÉS DEL HOYO PEÑA

PROFESOR GUÍA:
ROBERTO RONDANELLI ROJAS

PROFESOR CO-GUÍA:
RICARDO MUÑOZ MAGNINO

MIEMBROS DE LA COMISIÓN:
MONICA ZAMORA ZAPATA
ROLAND STULL

Este trabajo ha sido parcialmente financiado por CONICYT-PFCHA/MagísterNacional/2017-22170934

SANTIAGO DE CHILE
2022

RESUMEN DE LA TESIS PARA OPTAR AL GRADO DE
MAGISTER EN METEOROLOGÍA Y CLIMATOLOGÍA
POR: MIRKO FRANCISCO ANDRÉS DEL HOYO PEÑA
FECHA: 2022
PROF. GUÍA: SR. ROBERTO RONDANELLI ROJAS
PROF. CO-GUÍA: SR. RICARDO MUÑOZ MAGNINO

PRONÓSTICOS DE RADIACIÓN SOLAR EN EL DESIERTO DE ATACAMA USANDO ENSAMBLES ANÁLOGOS Y UNA TÉCNICA DE ASIMILACIÓN DE NUBES

Chile ha establecido una desafiante meta de capacidad instalada a partir de fuentes de energía renovable para 2050, lo que ha llevado a un aumento de los proyectos de energía solar en los últimos años. Una de las zonas con mayor desarrollo de proyectos solares es el Desierto de Atacama, debido a los altos índices de irradiancia solar y la poca nubosidad de la región. Debido al incremento en el número de proyectos ubicados en el Desierto de Atacama, y las características del mercado de la energía en Chile, existe un alto interés en métodos que sean capaces de pronosticar la irradiancia solar dentro del mismo día, o bien, para el día siguiente, con el fin de obtener una estimación precisa de los ingresos monetarios de la planta y reducir la incertidumbre de la producción de las plantas de energía solar.

En este trabajo, hemos probamos 4 métodos diferentes para pronosticar la irradiación solar en el desierto de Atacama, basados en la predicción meteorológica numérica utilizando el modelo WRF (Weather Research and Forecasting model, por sus siglas en ingles) . Utilizando los pronósticos de WRF y las observaciones históricas de varias estaciones meteorológicas, se crearon ensambles analógos (WRF-AnEn), con el fin de crear un pronóstico de ensamble para la irradiancia solar y reducir el sesgo de las predicciones. Además, también se incluyó en este análisis un pronóstico que usa una técnica de asimilación de nubes basada en datos del satélite GOES-16 (WRFDDA) y creando tambien un ensamble analógico que usa este pronóstico con asimilación de nubes (WRFDDA-AnEn). Los resultados se compararon con diferentes estaciones meteorológicas ubicadas en el Desierto de Atacama y en la costa de la Región de Atacama y Antofagasta en Chile.

Se encontró que el pronóstico del ensamble análogo obtuvo el error medio y el sesgo más bajo entre todos los pronósticos estudiados, creando pronósticos de irradiancia solar más cercanos a la observación. También, se encontró que WRFDDA pudo crear pronósticos de cobertura de nubes más realistas que WRF durante las primeras 4 - 8 horas después de la inicialización del modelo. No obstante, la inclusión de nubes afectó el pronóstico de irradiancia solar WRFDDA, lo que generó un mayor error medio en comparación con WRF.

Sin embargo, y comparando con los pronósticos de irradiancia solar obtenidos del modelo GFS (Global Forecast System, por sus siglas en ingles), se obtuvo una mejora media para el RMSE del 3,8% y del 1,6% para WRFDDA-AnEn y WRF-AnEn para la irradiancia solar. Se encontró tambien una mejora en el RMSE de 0.7% y 0.4% para WRF y WRFDDA en promedio respectivamente para la irradiación solar. La mejora de WRFDDA-AnEn sobre WRF-AnEn puede explicarse por el pronóstico de cobertura de nubes más preciso de WRFDDA, lo que ayudó a una mejor selección de análogos en el período histórico, creando pronósticos de irradiación solar más precisos.

RESUMEN DE LA TESIS PARA OPTAR AL GRADO DE
MAGISTER EN METEOROLOGÍA Y CLIMATOLOGÍA
POR: MIRKO FRANCISCO ANDRÉS DEL HOYO PEÑA
FECHA: 2022
PROF. GUÍA: SR. ROBERTO RONDANELLI ROJAS
PROF. CO-GUÍA: SR. RICARDO MUÑOZ MAGNINO

SOLAR IRRADIANCE FORECASTS IN THE ATACAMA DESERT USING ANALOG ENSEMBLES AND A CLOUD ASSIMILATION TECHNIQUE

Chile has set an ambitious goal of capacity installed from renewable energy sources for 2050, which has led to an increase of solar energy projects in the past years. One of the areas with the highest development of projects is the Atacama Desert, due to the high solar irradiance values and the low cloudiness in the region. Due to the increase in the number of projects located in the Atacama Desert, and the characteristics of the energy market in Chile, there is a high interest in methods that are able to forecast the solar irradiance during the same day or for the next day, in order to obtain a more precise estimation of the revenues and reduce the uncertainty of the energy output of the solar energy plants.

In this work, we tested 4 different methods to forecast solar irradiance in the Atacama Desert, based on numerical weather prediction using the weather research and forecasting model (WRF). Using the WRF forecasts and the historical observation from several meteorological stations, analog ensembles were created (WRF-AnEn), in order to create an ensemble forecast for the solar irradiance and reduce the bias of the predictions. Moreover, a forecast using a cloud assimilation technique based on GOES-16 satellite data was also included in this analysis (WRFDDA) and an analog ensemble using this forecast (WRFDDA-AnEn) was also created. The results were compared against different meteorological stations located in the Atacama Desert and in the coast of the Atacama and Antofagasta Region in Chile.

It was found that the analog ensemble forecast obtained the lowest mean error and bias between all the forecasts, creating solar irradiance forecasts that are closer to the observation. It was also found that WRFDDA was able to create more realistic cloud coverage forecasts than WRF during the first 4-8 hours after the initialization of the model, although affecting the solar irradiance forecast, which created a higher error compared to WRF.

Nonetheless, and compared to solar irradiance forecasts from the Global Forecast System (GFS), on average there is an improvement of 3.8% and 1.6% for WRFDDA-AnEn and WRF-AnEn on the RMSE of the solar irradiance forecasts. Only an improvement on the RMSE of 0.7% and 0.4% were found for WRF and WRFDDA on average respectively. The improvement of WRFDDA-AnEn over WRF-AnEn may be explained by the more accurate cloud cover forecast of WRFDDA, which may help the selection of more similar analogs in the historical period, creating more accurate solar irradiance forecasts.

Para mis padres, gracias por todo.

Acknowledgments

I would like to thank the following people, without whom I would not have been able to complete this work.

To my supervisor Roberto Rondanelli, thanks for your patience and support during these past years. I would also like to thank Ricardo Muñoz and Monica Zamora who helped me at different stages of this work with their insight and opinions. A special thanks to Prof. Roland Stull, who received me in his lab and gave me all the inspiration and tools to continue and prepare this work.

Finally, I would like to thank my family, friends and all the people from the Department of Geophysics of the University of Chile and the Center for Climate and Resilience Research; their words and support were incredibly important to (finally) finish this journey.

The research has received financing from the Center for Climate and Resilience Research (FONDAP 15110009), and the grant CONICYT-PFCHA/MagísterNacional/ 2017-22170934.

Table of Content

List of Tables

List of Figures

Chapter 1

Introduction

The Chilean energy policy *Energía 2050* [1] proposed a goal of renewable energy generation of 60% by 2035 and 70% by 2050. The transition from fossil fuels to renewable energy sources such as wind and solar requires not only a big initial investment to install several wind/solar farms but also intensive research and development, so the energy transition does not affect the stability of the current Chilean electric system/market. Figure 1.1 shows the *Energia 2050* plan and the expected contribution on each energy source in the future.

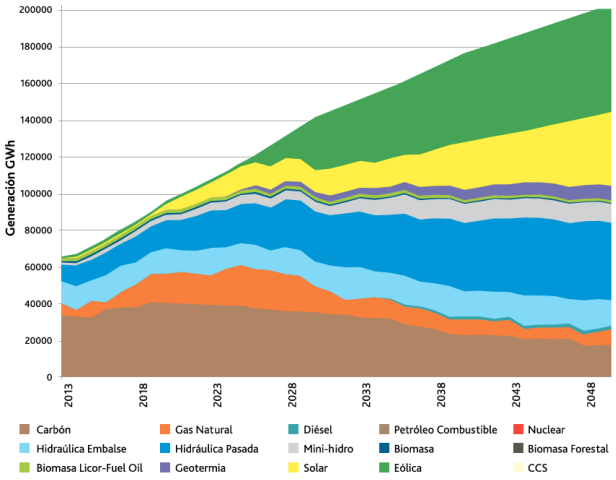


Figure 1.1: Future projections of energy sources for Chile. Source : *Energía 2050* [1]

The long-term goal proposed by *Energía 2050* shows that solar energy will be one of the most relevant energy sources that will help to reach the goal proposed by 2050. One reason that may explain this is the high solar irradiance values throughout the country [2, 3, 4]. Indeed, studies show that the highest solar irradiance values on earth are found in the Atacama Desert [5]. The rising interest in recent years in solar energy is reflected in the creation of several research centers, and an increase of Research and Development related to solar energy, also helped by public initiatives that contributed to the estimation and characterization of the solar irradiance resource in Chile, such as the *"Explorador Solar"* [6].

However, solar energy in Chile still has technical and economic barriers that difficults the expansion of this technology. Haas et al. [7] indicated that the most important barriers are the lack of economical incentives; the volatility of solar panels, equipment and energy prices; the absence of surface data information from meteorological stations; the deficiency of good forecasting systems; the environmental conditions in some areas of Chile and the need of water to clean the solar panels.

Despite these barriers, by February 2021, Chile currently has more than 3.4 GW[1] of installed capacity of photovoltaic solar plants. Also, more than 3.3 GW of installed capacity in photovoltaic projects is currently under development. These values show the increasing interest in solar energy in recent years.

The vast majority of the large solar farms in Chile are located between the Tarapacá Region and the Central part of the country. Nevertheless, the Atacama Desert and the areas close to it have always been one of the most preferred areas for solar energy, due to the high solar irradiance values and low cloudiness in the central valleys located a few kilometers inland from the coast. Currently, almost 1.6 GW of capacity installed in solar farms are installed/under development in areas inside or close to the Atacama Desert.

Nonetheless, and regarding the barriers related to the environmental conditions, the solar resource variability during the summer and the effect of environmental variables such as aerosols, the temperature, wind speed and water vapor, and the cloud cover variability at the coast are still being one of the major challenges to accurately estimate the solar irradiance in this region. The impact of these variables on the energy production is important, and their consideration is vital in order to obtain an accurate long or short-term prediction of the capacity factor of a solar farm.

Given the energy production variability of a solar farm, energy companies use forecast tools in order to estimate the energy production in different time scales [8]. The energy forecast is usually created using numerical weather models or using statistical methods. Figure 1.2 shows the different time scales used for both methods, in which it is possible to observe that numerical weather models are used mostly for temporal horizon up to 10 days and spatial scales between 1-380 km. Meanwhile, statistical methods are used for short time scales, with spatial scales between 10 meters and 10 kilometers. The area in red shows the ideal spatial resolution and temporal horizon for solar energy forecasting, showing an ideal temporal horizon between 1 hour and a couple of days, and a horizontal resolution between 1 and 10 kilometers.

[1] `https://www.cne.cl/wp-content/uploads/2021/02/RMensual_ERNC_v202102.pdf`

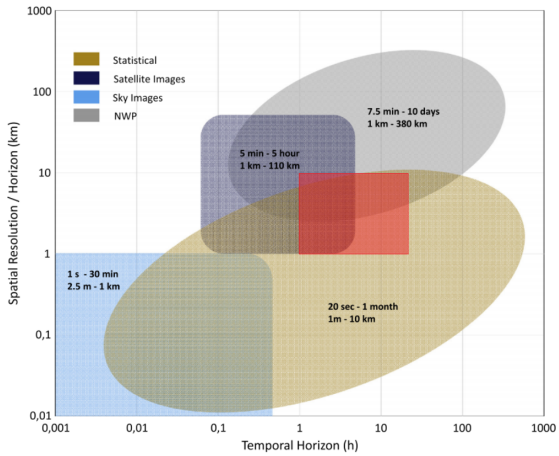


Figure 1.2: Temporal and horizontal resolution recommended for different methods to obtain solar irradiance forecasts. Source: Antonanzas et al. [8]

Several studies have shown the feasibility of using numerical and statistical models to forecast the energy production of solar farms [9]. Still, numerical weather prediction models have taken more relevance during the last years due to the low errors obtained when forecasting with a temporal horizon of 1 or more days [10], which is useful in cases where energy companies want to estimate the future energy that it will be needed to buy/sell in the market, based on the future energy output of the solar farm in the future.

However, solar forecasts using numerical weather prediction models possess several disadvantages that have been studied in recent years. In particular, large errors due to cloudiness, especially in low clouds [11] and thin clouds (cirrus) [12]. On the other hand, the lack of public and historical weather databases with environmental variables such as vertical column water vapor/ozone and aerosols has impacted the accuracy of solar irradiance forecasts [13]. Most of the time, these variables are the large contributors to the systematic errors in solar irradiance forecasting [13].

For the Atacama desert, the unique regional climate presents a challenge for weather forecasting. In particular, the presence of low clouds off the coast of the Atacama desert, the unique regional wind flow in the interior [14] and the intense rainy season in the Altiplano [15] are not features usually well represented/modeled by weather forecasting models [16, 17, 18]. For solar energy forecasting, for example, cloud cover variability in the Atacama desert is a topic of interest for solar farm developers, as this variable is important in order to estimate the solar irradiance at the surface. Figure 1.3 shows the seasonal cloud cover for the Atacama Desert between -72.8°W and -67.5°W, in which it is possible to observe the difference between the annual mean cloud cover for different parts in the region.

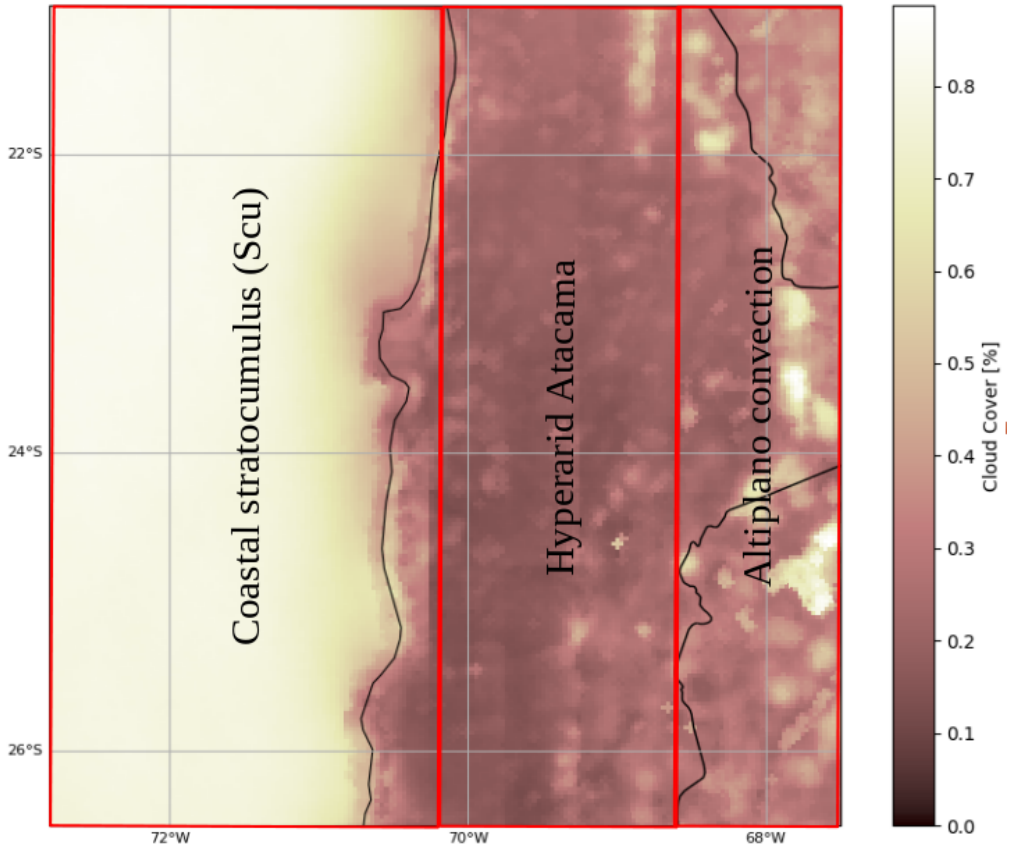


Figure 1.3: Mean Cloud Cover [%] between June-2017 and February-2019 obtained with GOES-16 between 8 AM and 8 PM UTC

Based on the annual cloud cover shown in Figure 1.3, it is important to consider the different types of clouds present in the Atacama Desert, in order to obtain an accurate forecast of the cloud cover and the solar irradiance.

One of the most distinct types of clouds present in the Atacama Desert are the stratocumulus clouds located on the coast of the Atacama Desert. The large-scale subsidence found at the coast of the Atacama Desert, the characteristic topography found along the coast of Chile and the cold waters of the Pacific Ocean at this latitude help the formation of a dense and persistent cover of stratocumulus clouds. Different publications have explained the mechanism of the formation, dissipation and challenges in modeling the stratocumulus cover in front of the Atacama Desert, such as Muñoz et al. [19] and Abel et al. [16].

The formation and persistence of the stratocumulus deck off the Atacama Desert coast is a complex ocean-atmosphere process, linked to the cold sea waters in front of Chile and the subsidence due to the presence of the Southeast Pacific High. The large-scale subsidence creates a stable layer with a distinct thermal inversion just above the boundary layer, setting the top of the stratocumulus deck. On the other hand, the longwave radiation cooling at the top of the stratocumulus cloud enhances the thermal inversion above the cloud top, maintaining and preserving the stratocumulus cloud.

Stratocumulus clouds at the coast of the Atacama Desert are present during the whole

year, with a maximum cloud coverage during the austral winter and spring, and a minimum between January and March [19]. On the other hand, the cloud cover fraction is higher during the night-morning, and reach a minimum during the afternoon, as the solar irradiance increase. Additionally, the stratocumulus horizontal coverage changes during the day. During the night and morning the cloud deck is usually close to the coast, next to the coastal mountain range. As the day passes and the solar irradiance increase, the stratocumulus cloud deck moves to the west, as the dissipation of the clouds closer to the coastline increase. The cycle continues after the sunset when clouds start forming again closer to the coastline.

In order to illustrate the spatial and temporal variation of the stratocumulus cloud cover, Figure 1.4 shows the cloud cover for October 26th, 2018 between 12 UTC and 18 UTC. During this day, the stratocumulus cloud cover persists in front of the Atacama/Antofagasta Region for several hours. As the day continues, the coastline starts to clear during the evening and the clouds start to move out to sea. This day shows the typical stratocumulus spatial/temporal cloud cover variability. An accurate representation of the cloud cover variability of the stratocumulus clouds is important to obtain precise forecasts of the solar irradiance for coastal sites.

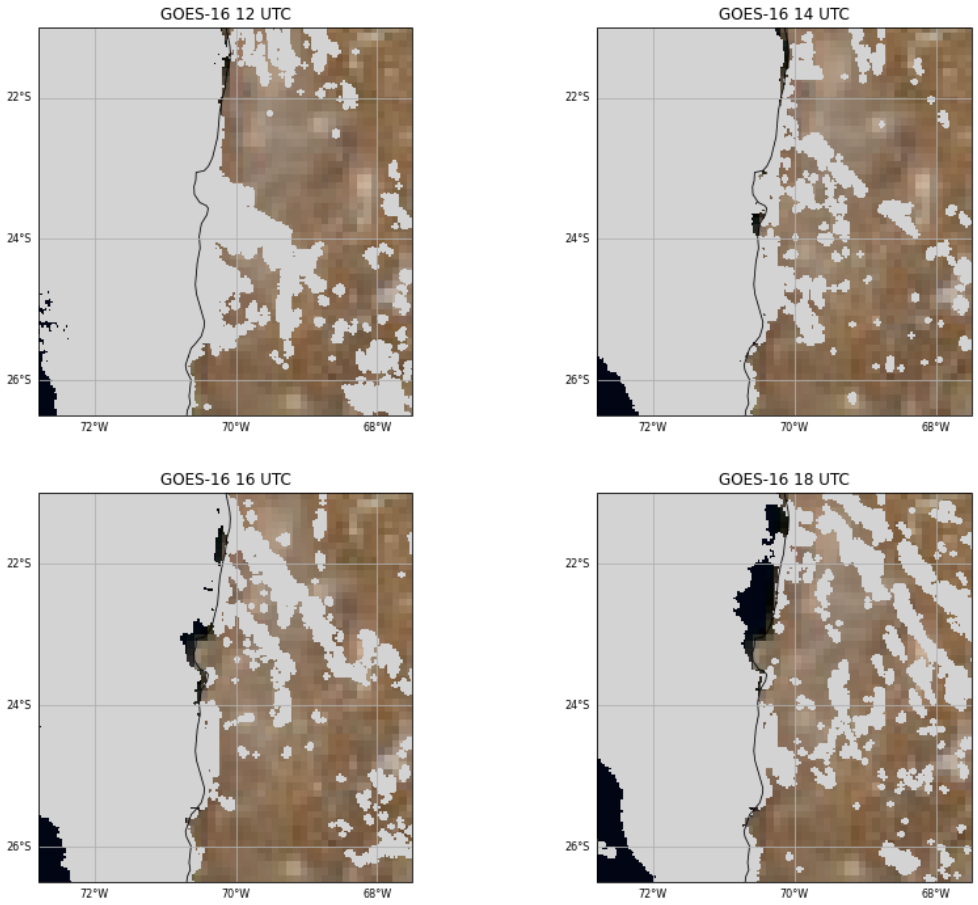


Figure 1.4: Cloud cover during October 26th, 2018 obtained with GOES-16 between 12 UTC and 18 UTC. Cloud areas are marked in white.

The cloudiness due to the intense convection in the Altiplano region is also important in

order to understand the challenges of forecasting the cloud cover and the solar irradiance in the Atacama Desert. The convection during the warm season, linked to the South American Monsoon regime or the "Bolivian Winter", is produced by the high humidity, east winds on the eastern side of the Andes and the surface warming during the warm season. The intense convection induces the formation of a high-altitude pressure system over the Altiplano, known as the "Bolivian high". The upper-level circulation of this high-level pressure system produces easterly winds, which transport mid and high clouds (cirrus clouds) to the west part of the Altiplano and the Atacama Desert region.

Thin cirrus clouds may impact solar irradiance forecast estimations, given the difficulties of estimating the temporal-spatial variation and the solar irradiance attenuation produced by using numerical weather prediction models [20]. Figure 1.5 shows the cloud cover according to GOES-16 for February 10th, 2019. During this day, intense convection is found in the Altiplano, together with cirrus clouds that pass through the hyper-arid region of the Atacama Desert. The main characteristic of the cloud cover during the intense convection episodes in the Altiplano is that clouds are present in most of the Atacama Desert. Although these events of intense convection are less frequent during the rest of the year, they are responsible for most of the cloud cover frequency in the interior of the Atacama Desert, and hence, one of the main contributors to solar irradiance decrease in summer.

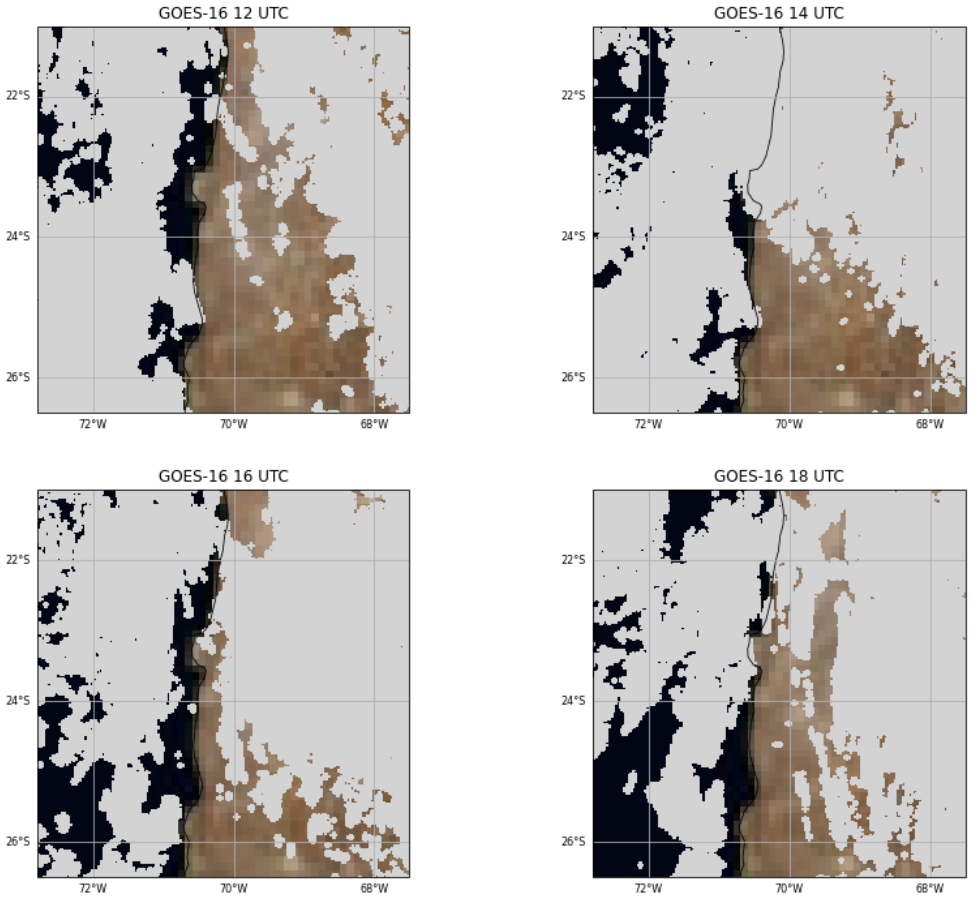


Figure 1.5: Cloud cover during February 10th, 2018 obtained with GOES-16 between 12 UTC and 18 UTC

To decrease the systematic errors of solar irradiance forecasts, especially the one produced by incorrect cloud cover or cloud optical depth, multiple solutions have been proposed in recent years. For example, the improvement of weather databases and meteorological stations that are used to validate and correct the forecasts, the improvement of the initial condition by using new assimilation techniques, and the use of statistical methods to post-process the solar irradiance forecasts have impacted positively the quality and accuracy of solar irradiance forecasts. The most used statistical methods are Kalman filters, PPM (Perfect Prog Method), MOS (Model Output Statistics), and Analog Ensembles. The use of each statistical method relies on the variable being studied and the frequency in which the event occurs [21].

Among the most used statistical methods, the analog ensemble is one of the most preferred methods given their flexibility and low computational resources required [22]. The analog ensemble consists of using past forecasts and historical measurements to create ensemble forecasts with lower bias than using standalone forecasts. The past forecasts or analogs are selected based on a similarity filter, selecting the historical periods that are the most similar to the future forecasts. Then, analogs are compared with on-site measurements or any long-term reference database, to then create an ensemble of the past forecasts, which will help to assess the quality and uncertainty of the variable that is being forecasted. In renewable energy forecasts, studies conducted by Alessandrini et al. [23], Haupt et al. [24], Cervone et al. [25], Zhang et al. [26] showed how the analog ensemble created using meteorological stations and/or operational wind/solar data can both reduce the mean error of the forecasts and can also be used to obtain an insight into the uncertainty of the forecasts. [23] for example found a reduction of 7-10% in the mean average error for solar forecasts in a region in Italy, which varied based on the time and period of the year.

Given the unique regional climate of the Atacama Desert and how important is to obtain accurate solar irradiance forecasts for owners/developers of solar farms and to accelerate the transition to an electric market with renewable energies, the purpose of this work is to produce and analyze different methods to obtain solar forecasts in the Atacama Desert, in order to obtain one with the lowest possible error. The solar irradiance forecasts will be produced using WRF-Solar, an alternative version of the WRF model used specifically for solar irradiance forecasts. The forecasts will be compared against a network of on-site meteorological stations using different statistical indexes, in order to quantify the quality of the forecasts at each site between June 2017 and March 2019.

With the aim of decreasing the error between the WRF forecasts and the on-site measurements, a cloud assimilation technique will be used alongside the WRF forecasts, with the purpose of producing accurate cloud cover forecasts, which may improve the quality of solar irradiance predictions.

Also, an Analog Ensemble will be obtained using WRF forecasts, in order to create solar irradiance ensemble forecasts that could increase the quality and accuracy of the solar irradiance forecasts in the Atacama Desert.

The data and methodology used in this study can be found in Chapter 2. The main results and the validations of the solar forecasts can be found in Chapter 3. Finally, the main conclusion and a discussion of possible future work can be found in Chapter 4.

Chapter 2

Data and methodology

In this study, the solar irradiance forecast for the Atacama Desert was performed using the numerical weather prediction model WRF-Solar [27]. The difference between WRF and WRF-Solar is that the latter includes specific solar irradiance parametrizations for solar irradiance forecasts, such as the effect of aerosols on the absorption/scattering or the effect of shallow clouds on the solar irradiance.

48-hour forecasts were created using WRF-Solar, initialized using the Global Forecast System (GFS) [28]. The resolution of GFS used to initialize the forecast is 0.25°. Then, 3 nested domains with a horizontal resolution of 27 km, 9 km and 3 km were created. Figure 2.1 shows the WRF domain used in this study. The vertical resolution was set to have 50 vertical levels. From the 50 vertical levels, 15 levels are under the first kilometer above the ground, with the objective of obtaining a good representation of the low cloud cover. Each forecast was initialized at 12 UTC.

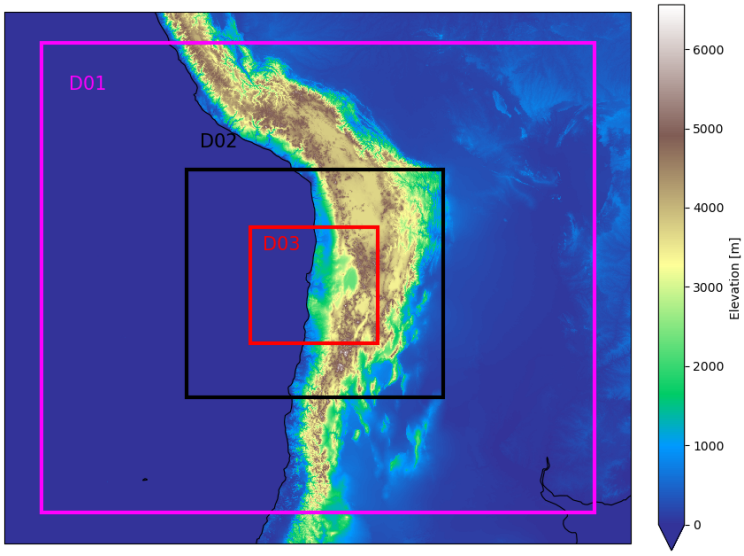


Figure 2.1: WRF domains proposed in this study. The colorbar represent the elevation in meters.

Several combinations of WRF parameterizations were tested in order to obtain the best configuration (not shown in this study). However, it was found that the best configuration was the standard WRF configuration (CONUS). The WRF configuration used in this study is shown in Table 2.1. Minor adjustments were applied to the model, based on the work of Muñoz et al. [29]. This is, the soil humidity of the land cover areas labeled as "Desert" were reduced from 2% to 0%, and the roughness length of the desert surface was reduced to 0.001 m, as presented by Martínez [30]. These changes showed an improvement in the temperature, humidity, and wind speed on the simulation performed by Muñoz et al. [29] and Martínez [30] for the same geographical region as the present study.

Table 2.1: WRF-Solar configuration proposed in this study.

WRF Solar v3.6	
Horizontal resolution	27, 9, 3 km
Vertical resolution	50 vertical levels, 15 of those in the first kilometer.
Time step	10 s
Nesting method	Two-way nesting
Solar Radiation (long and shortwave)	RRTMG
Microphysics	Thompson
Cumulus	Tiedke
Boundary layer physics	MYJ
Surface layer physics	Eta similarity
Initial/Boundary conditions	GFS 0.25°

The verification of the WRF forecast was performed for both the solar irradiance and the cloud cover. The solar irradiance forecasts were verified using solar irradiance observation, obtained from the Dirección Meteorológica de Chile (DMC) network, the database from the Ministry of Energy of Chile (MINENERGIA), and the CRC1211 Database from the University of Cologne [31]. A brief description and the location of the stations used in this study are shown in Figure 2.2 and Table 2.2. In total, 6 met stations were considered, 1 from the DMC database (Taltal), 1 from the MINENERGIA Database (Crucero) and 4 from the CRC1211 Database (Cerro Calate, Salar Llamara, Caleta Loa, and Quebrada Mani). The meteorological stations were chosen based on their proximity to the coast and to cover different areas of the Atacama Desert. Also, all stations have at least 18 months of data, in order to create the analog ensemble forecasts and to perform the verification with a period as long as possible.

Table 2.2: Meteorological stations used to validate the solar irradiance forecasts in this study. Mean GHI was calculated only for daytime periods.

Met station	Latitude	Longitude	Altitude	GHI [W/m^2]	Period
Taltal	-25.41	-70.48	37	472.9	2017/06/01 - 2019/03/01
Crucero	-22.27	-69.57	1183	640.7	2012/08/17 - 2019/03/01
Caleta Loa	-21.42	-70.06	19	521.3	2017/10/01 - 2019/01/14
Cerro Calate	-21.41	-69.86	1148	610.3	2017/06/08 - 2019/03/01
Quebrada Mani	-21.11	-69.07	2399	595.2	2017/09/25 - 2019/03/01
Salar Llamara	-21.36	-69.54	789	618.7	2017/09/24 - 2019/03/02

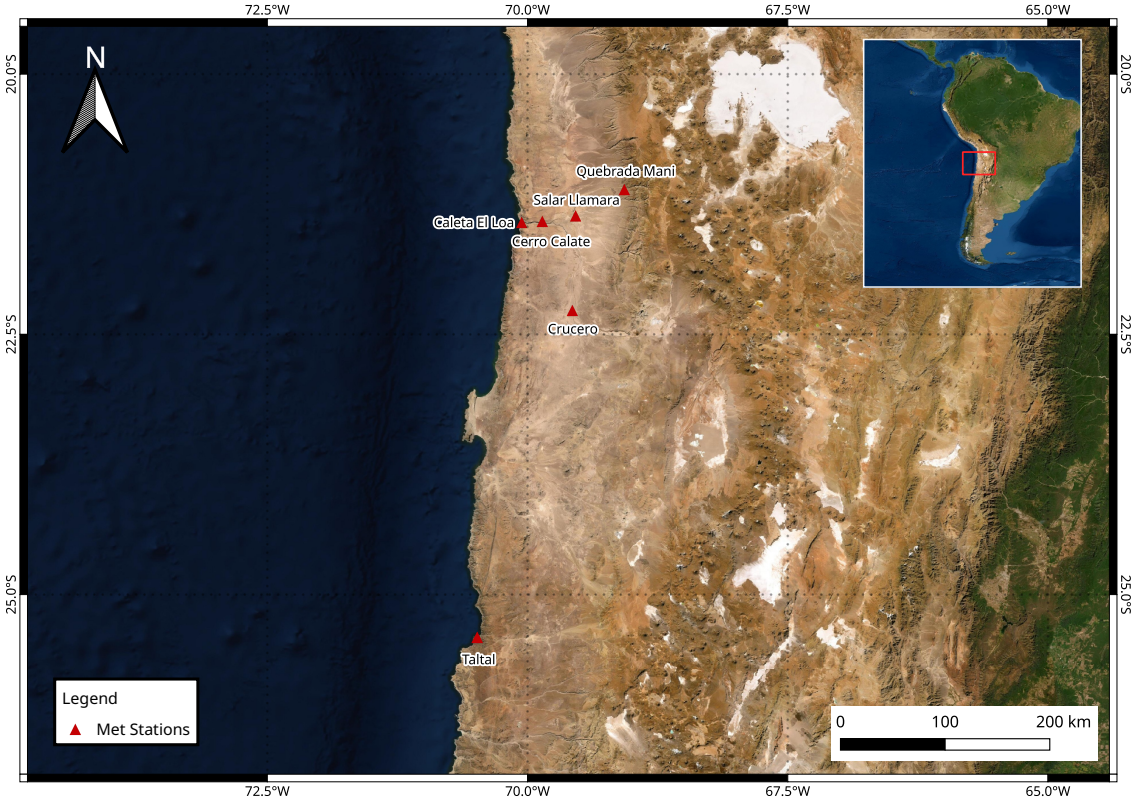


Figure 2.2: Map with the meteorological stations used in this study.

At each met station shown in Table 2.2, the 15-min mean solar irradiance data at 2 meters was obtained. Missing and erroneous solar irradiance data was excluded, in order to avoid incorporating spurious data into the study. Additionally, the 15-min average angle between the sun and the surface (solar zenith angle) was calculated for each met station and for the entire measurement period, using the python package PVLIB [32].

One of the challenges of modeling solar irradiance is that it relies on the cloud cover, as this variable is responsible for most of the scattering/absorption of solar irradiance. Given the quasi-permanent presence of stratocumulus on the coast close to the Atacama Desert, it is important to produce forecasts that are able to reproduce this cloud cover as accurately

as possible, in order to obtain good results. Hence, it is important to evaluate and validate the cloud cover forecasts in the Atacama Desert

To verify the cloud cover (CC) obtained with WRF, GOES-16 imagery was obtained for the entire Atacama Desert and the portion of the Pacific ocean in front of it. The cloud cover mask from GOES 16 imagery was obtained between June 2017 and February 2019 and consist of a binary variable with two option: Clear or Cloudy skies (0 and 1 respectively). Given that the WRF cloud cover (CC_{wrf}) is a variable that represents the grid coverage in presence of clouds (percentage of the grid with cloud, between 0 and 1), multiple cloud cover limits (CC_{lim}) were tested, in order to determine if a given WRF cell was considered cloudy or clear. This is represented in the following equation:

$$CC = \begin{cases} 0 \text{ (Clear)}, & \text{if } CC_{wrf} < CC_{lim} \\ 1 \text{ (Cloudy)}, & \text{otherwise} \end{cases} \tag{2.1}$$

Then, the verification was performed using different skill scores. For the solar irradiance, the RMSE, MAE, and Bias were calculated using the observation data. These skill scores were chosen based on the recommendation given by Yang et al. [33] for solar irradiance forecasts. The RMSE, MAE, and Bias were calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{\mathrm{i}=1}^{n}(f_{\mathrm{i}} - o_{\mathrm{i}})^2} \tag{2.2}$$

$$MAE = \frac{1}{n}\sum_{\mathrm{i}=1}^{n}|f_{\mathrm{i}} - o_{\mathrm{i}}| \tag{2.3}$$

$$Bias = \frac{1}{n}\sum_{\mathrm{i}=1}^{n} f_{\mathrm{i}} - o_{\mathrm{i}} \tag{2.4}$$

Where f_{i} and o_{i} are the forecasts and the observation values respectively. To assess the quality of a forecast Yang et al. [33] recommends the RMSE over MAE/Bias, given that RMSE is more sensitive to large errors produced in the model. For solar irradiance modeling, the larger errors are produced mostly by the difference in cloud cover or aerosol optical depth [34, 35, 36]. Nonetheless, the MAE and Bias are also presented to show more traditional skill scores and also to show the differences between each forecast and the observations.

The RMSE, MAE, and Bias were calculated for the entire forecast period and also every 15 min. Only daytime observations/forecasts were included in the RMSE/MAE/Bias calculation, by filtering out the periods with zenith angle (θ_z) ≥ 85, following the recommendation of Yang et al. [33]. Also, an additional verification was performed by taking only cloudy sky periods, in order to assess the quality of the solar irradiance forecasts during cloudy conditions, which are the most difficult periods to forecast accurately.

On the other hand, given that the cloud cover is a binary variable, the verification was based on the Heidke Skill Score (HSS) [37]. The HSS score was chosen because is equitable, which means that it gives an expected score to random forecasts, even if the forecasts and observation have a similar probability distribution. The HSS is also a transpose symmetric score, which means that changing observations and forecasts in the score formula does not change the skill score, which indicates that the HSS is difficult to hedge. Given the contingency table shown in Table 2.3, the HSS kill score is calculated as:

$$HSS = \frac{2(a\mathrm{d} - bc)}{(a+c)(c+\mathrm{d}) + (a+b)(b+\mathrm{d})} \tag{2.5}$$

Where a, b, c and d are the counts corresponding to the events matching the criteria shown in Table 2.3. In this Table, CC_{GOES16} correspond to cloud or clar sky according to GOES-16 and CC_{wrf} correspond to cloudy or clear sky according to WRF.

Table 2.3: Cloud Contingency table used in the study.

Cloud Forecasted	**Cloud Observed**	
	Yes ($CC_{GOES-16} = 1$)	No ($CC_{GOES-16} = 0$)
Yes ($CC_{wrf} > CC_{lim}$)	a	b
No ($CC_{wrf} \leq CC_{lim}$)	c	d

Hogan et al. [38] studied different skill scores to perform cloud cover verification and found that HSS has most of the desirable properties to verify cloud cover forecasts. Although Hogan et al. [38] shows that the HSS score has disadvantages related to bias and the use of different cloud limits (CC_{lim}), the HSS skill score was selected given the familiarity in forecast verification [37, 39]. For HSS, a perfect forecasts would have a score of 1 and a forecast without skill would have a score of 0 or lower.

WRFDDA: WRF forecasts using a cloud assimilation technique

Given the complexity of modeling the stratocumulus cloud cover in front of the Atacama Desert, another WRF forecast using an alternative cloud assimilation technique was created using the methodology described by Mathiesen et al. [40]. This forecast (herein called WRFDDA) uses cloud satellite data in order to correct the cloud cover forecasts obtained with WRF, by filling or removing clouds using the cloud top temperature and the cloud top pressure derived by GOES-16.

The WRFDDA forecast was performed as described in Mathiesen et al. [40], using the same code to perform the cloud assimilation. For this method, a normal WRF forecast was set with the configuration shown in Table 2.1, and after a spin-up of 2 hours the simulation was stopped to correct the cloud cover using the GOES-16 data. This correction was based on an empirical approach that takes into account the height of the temperature inversion (H_{inv}). H_{inv} is an important variable as the formation of the stratocumulus cloud cover depends on

the vertical mixing profile and the cloud top height, which are variables that depend on the height in which the thermal inversion is located [41]. A cloud height base model was used based on the study by Mathiesen et al. [40], which is shown below:

$$z_{b,base} = \begin{cases} 0.45 \cdot H_{inv}, & \text{if } H_{inv} < 750m \\ 0.28 \cdot H_{inv} + 127.5, & \text{otherwise} \end{cases} \tag{2.6}$$

Where $z_{b,base}$ is the height of the base of the cloud in meters and H_{inv} is the height at which the thermal inversion is present.

Then, based on the base cloud and the temperature inversion height obtained from Equation 2.6, a 3D binary array is created, which the code used to create or destroy clouds in each of the grid cells in the WRF simulation. To create clouds, the code sets the water vapor mixing ratio to 1.1 times the saturation water vapor mixing ratio in each grid point. To destroy clouds, the water mixing ratio is set to 0.75 times the saturation water vapor mixing ratio. According to Mathiesen et al. [40], the values are characteristic of stratocumulus clouds. After the water vapor mixing ratio from WRF is corrected, the new cloud cover is replaced and the WRF simulation is restarted. To avoid further numerical instability in the WRF simulation, the cloud microphysics is turned off for one hour after the WRF simulation is restarted. Also, before incorporating the new cloud cover into the WRF simulation, the water mixing ratio is smoothed along the horizontal and the vertical axis, in to avoid large gradients that could cause numerical problems. Figure 2.3 shows the cloud creation/destruction using this technique for May 5th, 2018.

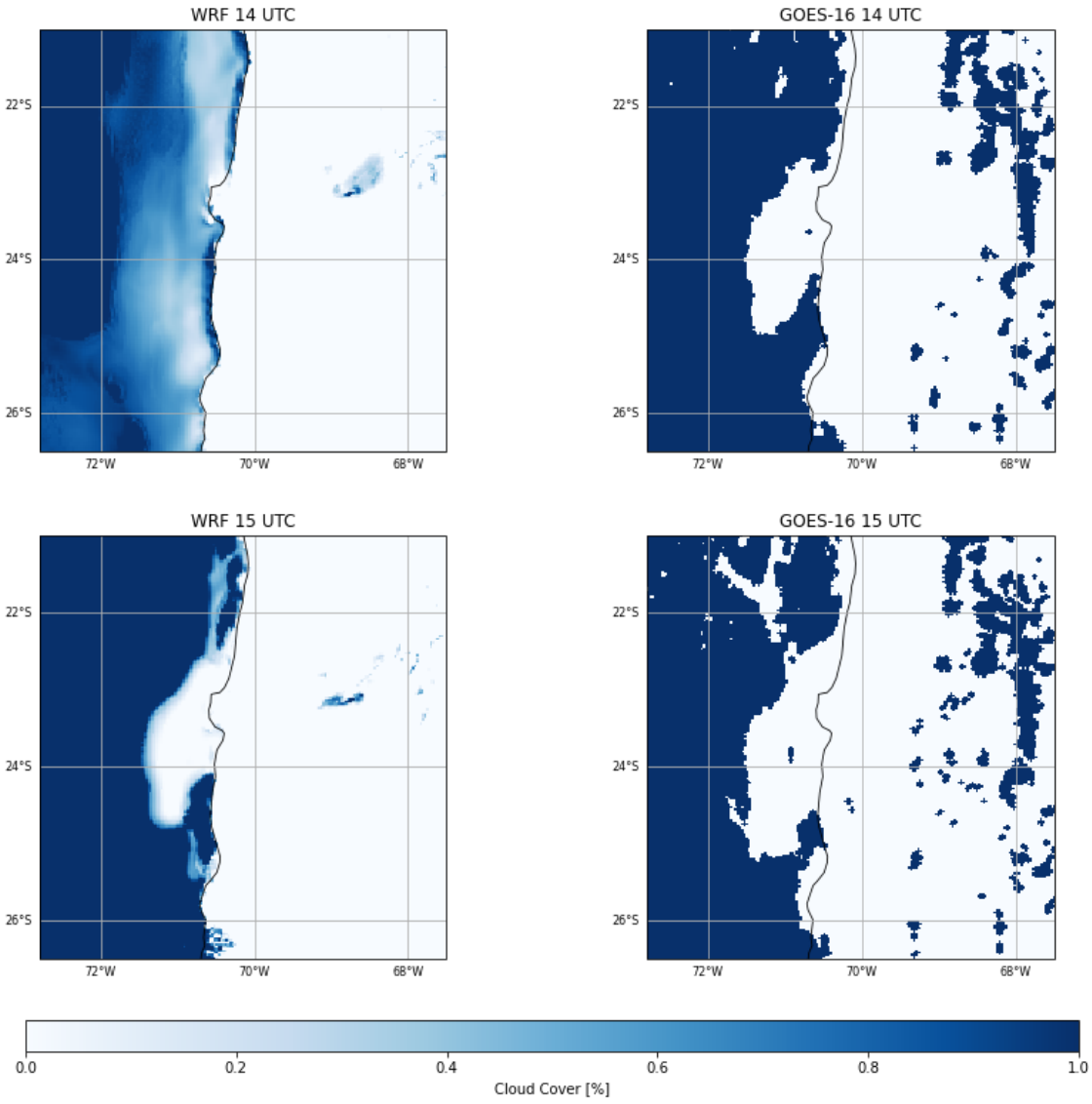


Figure 2.3: WRF cloud cover after and before the cloud assimilation technique for May 5th, 2018.

Given the lack of cloud measurements during the last years in the Atacama Desert region and that there is some similarity between the vertical profiles found in Mathiesen et al. [40] and the vertical profiles found in studies performed by Muñoz et al. [19], the input variables and the empirical model to estimate the cloud base were the same as the ones used in Mathiesen et al. [40]. The only difference between this study and the one performed by Mathiensen is that the cloud top temperature was obtained from GOES-16 instead of GOES-13. Also, given that this empirical model relies on the height of the temperature inversion in the boundary layer, it is expected that this method will have little to no effect on medium to high clouds, such as cirrus.

The verification of WRFDDA was performed using the same skill score used with the forecasts obtained with WRF. This is, the solar irradiance forecasts were validated by calculating the RMSE, MAE, and Bias using equations 2.2, 2.3 and 2.4 respectively. Also, the

cloud cover forecasts were verified and compared with the ones obtained with WRF using the HSS skill score.

AnEn : Analog ensemble forecasts

Additionally, analog ensemble forecasts were created using WRF and WRFDDA. The analog ensemble forecast is a technique that is used to create probabilistic forecasts from a deterministic forecast, based on the historical forecasts database and observational data. The analog ensemble (herein called AnEn) approach consists of obtaining multiple forecasts similar to the current model run (analogs), which then are sorted and matched by a similarity metric with their concurrent observations. Then, an ensemble forecast is created by taking the most similar and best analogs forecast. Figure 2.4 shows the AnEn approach given a historical database of observations and forecasts.

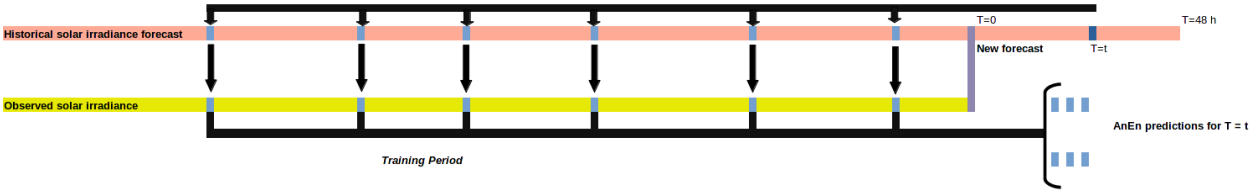


Figure 2.4: Analog ensemble forecast methodology.

The analog ensemble approach has many advantages compared to the conventional ensemble forecast. For example, the ensemble members are not limited to technical restrictions such as computational cost or time, as the number of ensemble members on the AnEn approach depends solely on the number of analogs that are selected. However, the number of analogs chosen can affect the quality of the forecasts, by selecting low-quality analogs that are not representative of the time that is being forecasted.

Another advantage of the AnEn approach is that it is possible to obtain information related to the uncertainty of the forecasts in less time than the normal ensemble forecast and that the AnEn forecasts do not require post-processing calibration, as the AnEn forecast is already based on the historical observations.

In this study, AnEn forecasts were created using historical solar irradiance from the meteorological stations in the Atacama Desert and historical cloud imagery obtained from GOES 16, together with the WRF/WRFDDA forecasts obtained between 2017 and 2019. The analog forecasts were created using the RAnEn code, provided by Hu et al. [42]. The RAnEn was set using an operational mode, in which the searching space for the first AnEn forecasts was set to the first 500 days, which was increased as the number of AnEn forecasts increased. Hence, if the historical period consist of 300 forecasts of 48 hours initialized once at the start of each day, the analog ensemble forecast for day one will be created by searching analogs in those 300 past forecasts. Then, for the second day, the analog ensemble forecast will be created using the first 300 forecasts and the forecast for the day before (day one), and so on.

Three variables have to be set up before obtaining the AnEn forecasts: the number of analogs (N_v), the temporal space window ($\tilde{t}$) and the weight of each variable (w_i) used to

create the similarity criterion used to select the analogs. These three variables are used together to create a similarity criterion, which is shown in the following equation:

$$\|F_t, A_t, \| = \sum_{\mathrm{i}=1}^{N_v} \frac{w_{\mathrm{i}}}{\sigma_{f_{\mathrm{i}}}} \sqrt{\sum_{j=-\tilde{t}}^{\tilde{t}} (F_{\mathrm{i},t+j} - A_{\mathrm{i},t'+j})^2}, \tag{2.7}$$

Where F_t are the forecasts that are going to be corrected, A_t are the analog forecasts obtained at time t' before the forecasts F_t and $\sigma_{f_{\mathrm{i}}}$ is the standard deviation of the time series forecasts for the variable that is being forecasted. The temporal window $\tilde{t}$ is used to search and identify forecast trends in the historical database that are similar to the forecast at the time t. For example, a temporal window $\tilde{t}$ of 2 will search forecasts with similar trends to F_t using the forecast between t-2 and t+2.

The analogs for both AnEn forecasts, created with WRF and WRFDDA (herein called WRF-AnEn and WRFDDA-AnEn), were obtained using a similarity criterion based on the solar irradiance (GHI), the cloud cover (CC) and the zenith angle (θ_z). The weight of each variable (GHI, CC and θ_z), N_v and $\tilde{t}$ were obtained using a brute force optimization, in which the analog ensemble was run using different configurations, with the goal of minimizing the root mean square error (RMSE) between the forecasts and the observation. The number of analogs (N_v) was tested between 1 and 50, and the time windows ($\tilde{t}$) between 1 and 80 time steps (15 min and 20 hours) for all sites. After that, the weight of each variable (GHI, CC, and θ_z) was set between 0 and 1, increasing by 0.05, with the condition that all the weights must sum 1. Table 2.4 shows the configuration of each analog ensemble (WRF-AnEn and WRFDDA-AnEn) used in this study.

Table 2.4: AnEn weights for using WRF and WRFDDA forecasts for each met station.

Met station	Forecast	N	$\tilde{t}$	Weights		
				GHI	CC	Zenith
Taltal	WRF	15	67	0.5	0	0.5
	WRFDDA	14	57	0.4	0.3	0.3
Crucero	WRF	15	40	0.1	0.2	0.7
	WRFDDA	15	25	0.3	0.4	0.3
Caleta Loa	WRF	15	43	0.1	0.4	0.5
	WRFDDA	15	2	0.1	0.7	0.2
Cerro Calate	WRF	15	31	0.1	0.1	0.8
	WRFDDA	15	75	0.8	0.1	0.1
Quebrada Mani	WRF	15	24	0.6	0	0.4
	WRFDDA	12	29	0.1	0.7	0.2
Salar Llamara	WRF	15	40	0.3	0	0.7
	WRFDDA	15	6	0.3	0.3	0.4

Solar irradiance, solar zenith angle, and cloud cover obtained from the met stations were used to obtain the analogs. Given that the cloud cover obtained by GOES-16 is a binary

variable, only 1 or 0 were compared with WRF cloud cover. The reason for this was to simplify the calculation of the weight for the cloud cover, as the use of different CC_{lim} means that multiple iterations have to be made, which can take more computational resources than simply using the cloud cover obtained with WRF/WRFDDA. Also, given that the similarity is based on a distance metric, it was found that leaving the WRF/WRFDDA cloud cover unchanged, was more advantageous compared to the transformation of the variable to binary, given that the similarity criterion was based on how much of the grid cell was covered by clouds, as opposed to the situation in which the grid cell was cloudy or not.

Finally, and as with the other 2 forecasts (WRF and WRFDDA either), the AnEn forecasts were verified using the observational data. To compare all 4 forecasts (WRF, WRFDDA, WRF-AnEn, and WRFDDA-AnEn), the verification was performed using the same historical period, which accounted for the dates between 2018-10-25 to 2019-02-28. Additionally, and as described previously, all periods with $\theta_z \geq 85°$ were filtered out, and an additional verification taking into account only cloudy period was performed.

Chapter 3

Results and discussion

3.1 WRF versus WRFDDA : Cloud cover forecast verification

First, cloud cover periods obtained with WRF and WRFDDA were verified against GOES-16 data using the HSS score for different forward lead times (FLT) using a cloud cover limit (CC_{lim}) of 0.4, which is shown in Figure 3.1 and 3.2. It is important to notice that the validation of WRF and WRFDDA against GOES-16 cloud cover data was only performed during daytime periods, as GOES-16 cloud mask has problems identifying low cloud cover during nighttime, due to limitations of the infrared sensor in GOES-16 [43]. In general, it was found that WRFDDA cloud cover forecasts showed a better performance compared to WRF, with HSS scores greater not only in the interior of the Atacama desert but also close to the coast.

As can be observed in Figure 3.1, there are multiple zones between 70°W and 72°W where the HSS score is zero or negative for the WRF forecast, meaning that the forecasts show no skill or perform worse than a random forecast. Nonetheless, WRFDDA HSS scores (Figure 3.2) close to the coast show higher HSS scores, with fewer negative values than the ones obtained with WRF, especially between 4-8 hours after the initialization, where the HSS score difference is over 0.2 for an FLT of 6 hours and a $CC_{lim} > 0.4$ for Taltal and Caleta Loa. This shows that there is an improvement in the forecast when cloud information is assimilated into the forecast.

The inclusion of cloud cover information from GOES-16 and the corrections on the vertical cloud mixing ratios seem to have a positive effect on the cloud cover forecasts, which persisted for some hours after the inclusion of the new information. As can be observed in Figure 3.1 and 3.2, areas with negatives HSS scores are present throughout the entire forecast for WRF, but not for WRFDDA. It is important to notice that although an improvement of the forecast is observed in the coast, the HSS scores obtained with WRFDDA and WRF are far from the HSS score of a perfect forecast (HSS = 1). The reason for this may be related to the difficulties of WRF/WRFDDA to reproduce the temporal and spatial variability of the stratocumulus cloud layer.

This improvement of the forecast is also observed in the interior of the Atacama desert. Comparing Figures 3.1 and 3.2 it is possible to observe a slight improvement of the HSS score in zones between the coast and the interior, around 70° W between FLT's 4 and 8. Besides that, no clear differences are observed for the remaining hours.

To assess the quality of the cloud cover forecasts, the difference between WRF and WRFDDA cloud cover forecast was calculated for each met station. This is shown in Figure 3.3, where the HSS scores difference is presented between WRF and WRFDDA, for different FLT's, cloud cover limits and for each met station. The HSS scores for the met station with WRF and WRFDDA is also presented in Appendix A. From Figure 3.3 is it possible to observe that HSS scores obtained with WRFDDA for met stations close to the coast (Taltal, Caleta Loa, and Cerro Calate) are higher than those obtained with WRF. The improvement found for WRFDDA in coastal met stations is also higher close to the initialization of the model, and it decreases as the forward lead time increase.

On the other hand, the increase of the HSS score for sites at the interior is lower compared to sites closer to the coast. Although the increase in the HSS score in sites like Salar Llamara and Crucero is over 0.1-0.15 for some cloud cover limit cases, the improvement is only observed for FLT between 4 and 6 hours. HSS score improvements of 0.1-0.14 were also observed in sites like Crucero for FLT between 12 and 36 hours.

Based on these results, it is possible to observe that the inclusion of GOES-16 cloud cover data on the WRF model produced an overall improvement on the cloud cover forecast on the Atacama desert. This improvement on the forecasts, verified by the increase of the HSS score, is greater for sites closer to the coast and for FLT between 4 and 8 hours. The greater impact of WRFDDA on the cloud cover forecast on sites close to the coast is explained by the fact that only low clouds were added on WRFDDA, which are mostly found on the coast rather than in the interior of the Atacama Desert. Additionally, the reason why the improvement is only present on FLT's between 4 and 8 hours could be because WRF is not able to sustain and reproduce the stratocumulus cloud cover properly, even with the inclusion of the GOES-16 cloud cover data in the model.

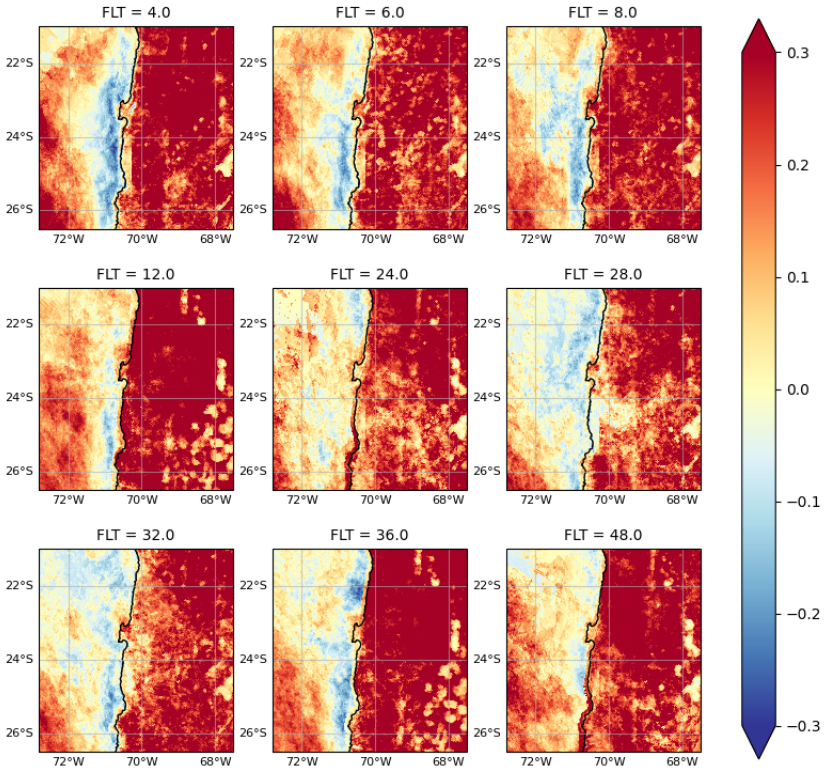


Figure 3.1: Heidke Skill Score (HSS) for the cloud cover forecast (CC $\geq$ 0.4) obtained with WRF for forecast lead times between 4 and 48 hours

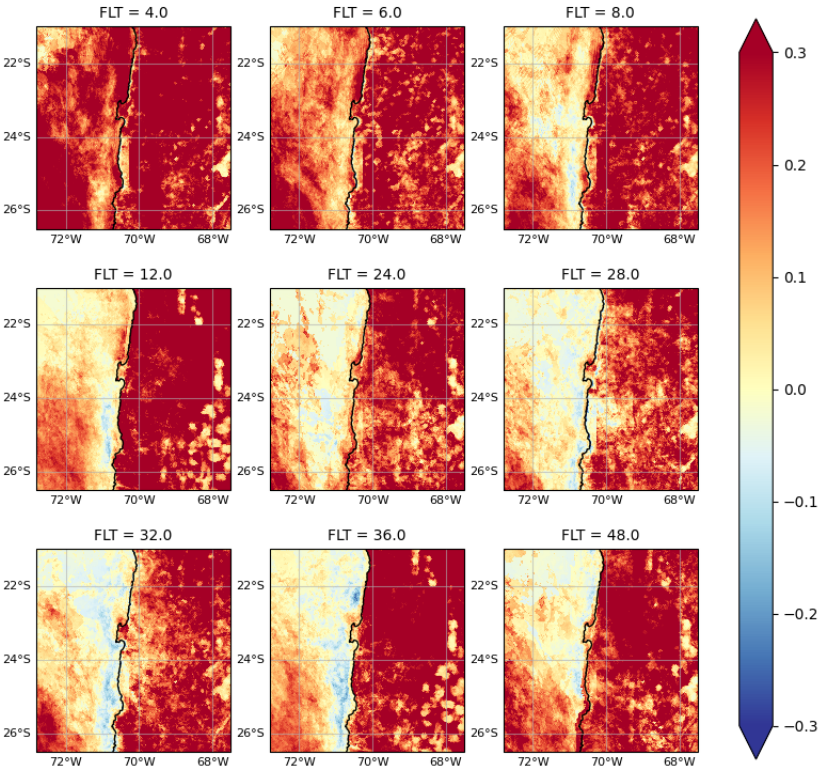


Figure 3.2: Same as Figure 3.1 but for WRFDDA

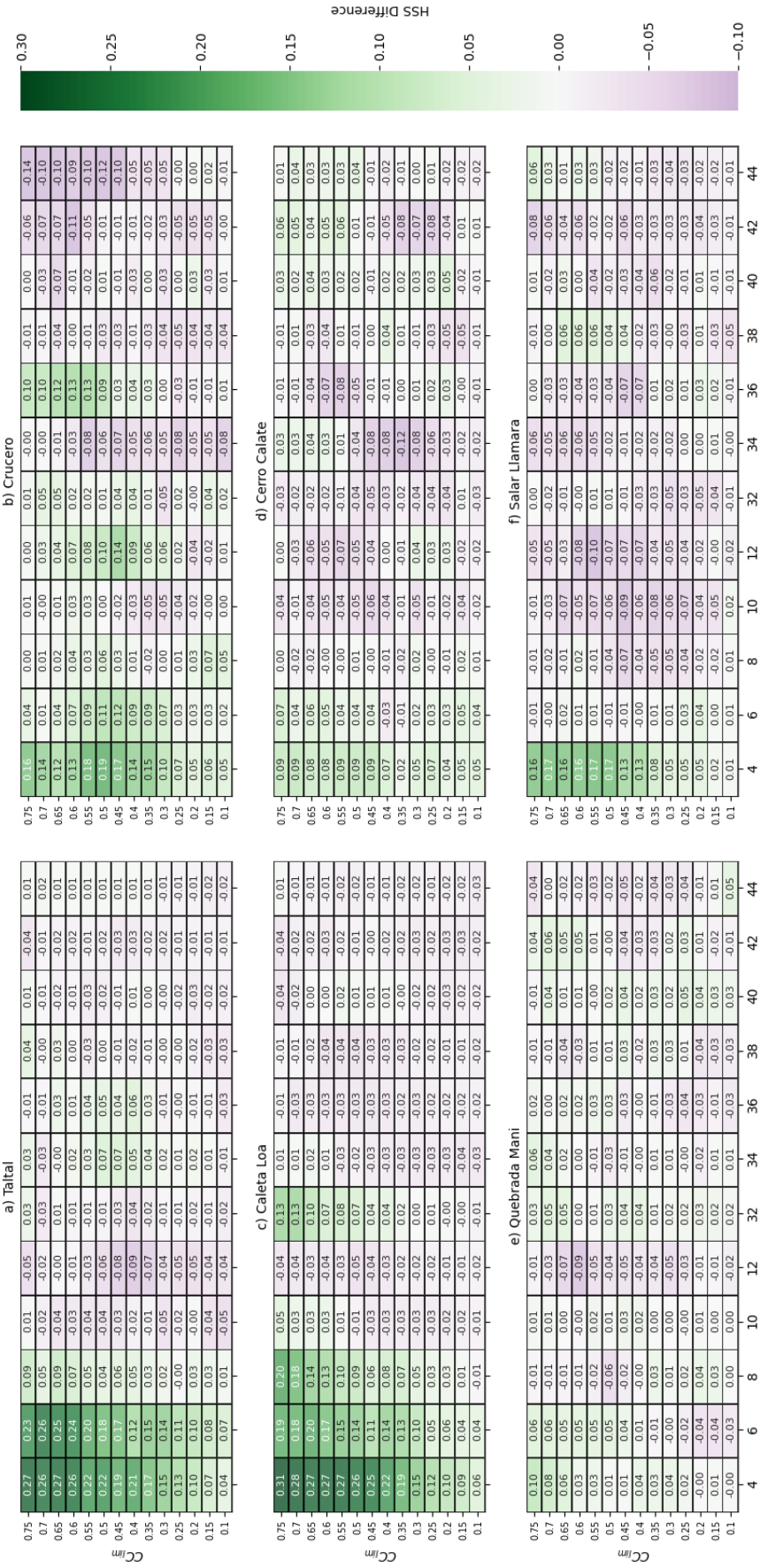


Figure 3.3: HSS score difference between WRF and WRFDDA forecasts for different forward lead times (FLT) and cloud cover limits (CC_{lim})

3.2 Verification during clear and cloudy periods

In order to evaluate the performance of each solar irradiance forecast (WRF, WRF-AnEn, WRFDDA, and WRFDDA-AnEn), Taylor plots were created for each site, which are shown in Figure 3.4. The results shown in this section consider only daytime periods and consider cloudy and clear skies. Additionally, the forecasts obtained with WRF were compared against GFS, in order to have a base forecast to compare and to observe if there is any improvement by using WRF to model the solar irradiance.

From Figure 3.4 it is possible to observe that the performance of each method varies along with the different sites. For sites close to the coast, such as Taltal or Caleta El Loa, the results are more spread and far away from the observed values than the results obtained in sites located in the interior. For sites located at the interior, like Crucero or Salar Llamara, it is possible to observe a good agreement between the forecasts and the observed values, especially for the forecasts obtained with the analog ensemble approach.

When clear-sky and cloudy periods are both considered in the verification, the performance between all the forecast models is close to each other for all the sites located at the interior of the Atacama Desert. This could be due to the low cloudiness and clear sky conditions prevalent at the interior, which minimizes the errors produced by the scattering or absorption of clouds. For sites like Crucero, Salar Llamara, and Quebrada Mani, the major advantage of the analog ensemble approach is that this technique can reduce the bias between the model and the measurement, reducing the overall difference between the measurements and the forecasted values.

On the other hand, a clear difference is observed in sites where cloudiness is frequent (Taltal, Caleta Loa, and Cerro Calate). The difference between the measurements and the forecast may be caused by the way that scattering/absorption is calculated in the presence of clouds. In the case of Cerro Calate and Caleta El Loa, an improvement in the correlation and the RMSE for the WRF-AnEn and WRFDDA-AnEn forecasts is observed compared to the other forecasts. Moreover, in the case of Taltal, a clear improvement is observed for the correlation and the root mean square error for the WRFDDA-AnEn, but also a decrease of the standard deviation of the forecasts was observed (Figure 3.4.c).

One reason that may explain the improvement observed for WRFDDA-AnEn on sites close to the coast may be the improvement on the initial cloud cover conditions for WRFDDA. This, because the more accurate cloud cover representation obtained with WRFDDA-AnEn may be related to a better selection of representative analogs in the historical period, especially in periods with cloudy skies. By taking into consideration the effect of low clouds on the historical solar irradiance data and by selecting analogs that are representative of cloudy conditions, it is possible to correct the error produced by the incorrect parametrization/model used to calculate the solar irradiance in presence of clouds, producing more accurate solar irradiance forecasts. Hence, although the cloud cover forecasts or the cloud optical depth may be different compared to the on-site measurements, the AnEn may correct the effect that clouds have on solar irradiance by using similar cloudy days in the historical period of the observations. It is important to notice that this correction is only possible if the forecast is able to predict the periods in which a cloud is present at a given site. Based on

the results shown in the previous sections, WRFDDA produces cloud cover forecasts that are more accurate than WRF, which may explain the reason why WRFDDA-AnEn produces better results than WRF-AnEn.

A diurnal comparison between the forecast and the measurement is presented in Appendix B, where the RMSE, MAE, and Bias versus FLT for each station can be found. Figures in Appendix B show how the analog ensemble approach has the lowest MAE, RMSE, and Bias for most of the met stations, corroborating the results observed on the Taylor diagrams.

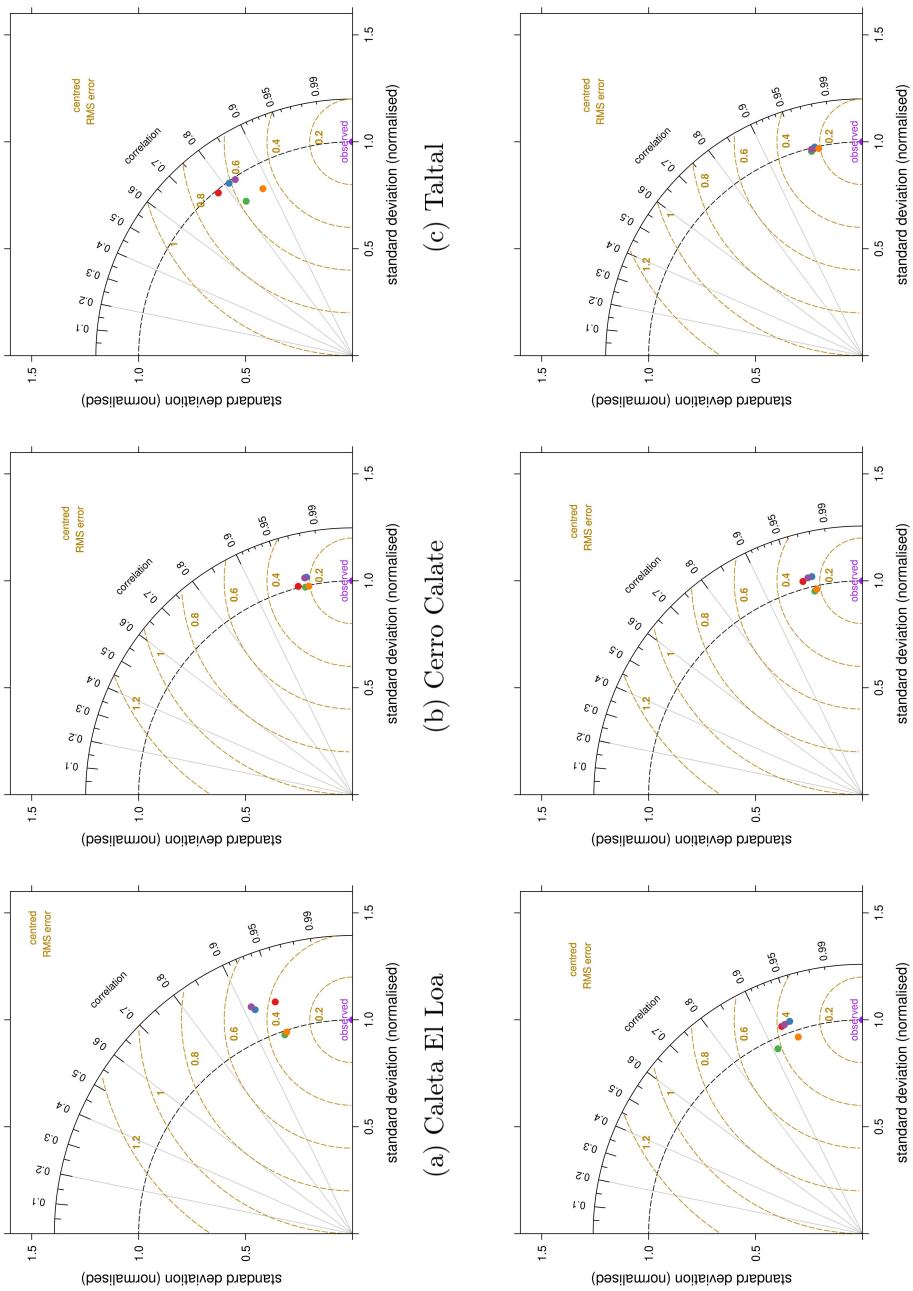
Table 3.1 shows the RMSE and MAE results for each site. Here, we can see that the WRFDDA forecasts is the best method. The improved cloud cover forecasts and the selection of accurate analogs made this method the best suitable among all the alternatives. This is true not only for sites close to the coast but also for sites in the interior of the Atacama Desert.

The results in Table 3.1 also show a decrease in the MAE between WRF and WRFDDA for coastal sites (Caleta Loa, Taltal and Cerro Calate), but an increase in the RMSE. The reason for this could be that although WRFDDA is able to produce accurate spatial cloud cover forecasts, the WRF radiative transfer model and the cloud properties in those periods may be not accurate enough to reproduce the surface solar irradiance. Given the high dependency of solar irradiance on clouds, small changes in cloud properties/coverage can produce large differences between measurements and estimations.

Nonetheless, the analog ensemble approach is able to correct the model bias, and produce results with RMSE and MAE lower than GFS or other WRF forecasts. The only exceptions are for WRF-AnEn forecasts of Cerro Calate and Quebrada Mani. This could be produced by an incorrect number of analogs or/and weight selections, which may lead to larger errors compared to the WRF forecasts. An improved method to select weights/number of analogs such as the one presented by Hu et al. [44] and the use of a historical database with a longer period is recommended to avoid this issue.

Table 3.1: Comparison between WRF, WRFDDA, WRF-AnEn, WRFDDA-AnEn, and GFS. The results are for clear and cloudy skies during the daytime between 4 and 48 hours after the forecast initialization.

	MAE [W/m^2]					RMSE [W/m^2]				
Site	WRF	WRF-AnEn	WRFDDA	WRFDDA-AnEn	GFS	WRF	WRF-AnEn	WRFDDA	WRFDDA-AnEn	GFS
Caleta Loa	115.3	50.1	107.4	47.2	100.8	122.1	86.2	127.2	83.3	99.1
Cerro Calate	44.8	31.9	43.1	27.2	42.8	64.0	65.2	66.2	60.4	74.9
Quebrada Mani	76.5	70.3	77.6	53.2	76.8	102.8	125.7	110.5	93.5	113.8
Salar Llamara	54.8	33.6	54.6	30.5	49.5	67.1	62.8	71.6	59.9	76.6
Taltal	130.5	101.4	109.7	82.7	130.4	161.7	142.2	147.2	118.0	174.9
Crucero	33.0	37.6	34.3	29.6	42.3	69.3	74.8	74.1	63.9	72.7
Mean	75.8	54.1	71.1	45.1	73.8	97.9	92.8	99.5	79.8	102.0



(a) Caleta El Loa (b) Cerro Calate (c) Taltal

(d) Quebrada Mani (e) Salar Llamara (f) Crucero

Figure 3.4: Taylor diagrams for the global solar irradiance in a) Caleta el Loa, b) Cerro Calate, c) Taltal, d) Quebrada Mani, e) Salar Llamara and f) Crucero between October 2018 and February 2019. Each point represents a forecast model.

Lastly, a comparison between the analog ensemble forecasts was performed, by analyzing the rank histogram (15-min observation/forecasts) of WRF-AnEn and WRFDDA-AnEn. This is shown in Figure 3.5, in which the rank histogram for each station and each model is shown, together with the Missing Rate Error (MRE), which is the fraction of observations higher/lower than the highest/lowest-ranked predictions above or below the expected missing rate of $1/(n+1)$. The negative/positive values of MRE show how over/under dispersive is the ensemble forecasts. From Figure 3.5, it is possible to observe that for Quebrada Mani, Salar Llamara, and Crucero the rank histogram shows over-estimation of the lowest and highest frequencies, especially for Quebrada Mani. This means that the analog ensemble spread is small, so the ensemble member often falls in the highest/lowest members of the forecasts. On the other hand, Caleta Loa and Cerro Calate show a positive bias, which means that the observations are closer to the ensemble members with the lowest values.

By comparing both analog ensemble forecasts, it is possible to observe that the MRE decreased between WRF-AnEn and WRFDDA-AnEn for all met stations, except Caleta Loa and Cerro Calate. The reason for this may be related to the underestimation of the cloud cover effect in both forecasts and the lack of analog ensemble members representing cloudy periods. A solution for this may be the use of an extended historical period for both met stations, which may lead to an increase of analog members that are representative of the lowest ensemble members. This may also help in sites such as Quebrada Mani, where the analog members most of the time were closer to the extremes of the analog spread.

Nonetheless, and based on the rank histograms, the results show an improvement of WRFDDA-AnEn over WRF-AnEn for Taltal, Crucero, Salar Llamara. The distribution for these met stations is more uniform and the MRE values decreased significantly. Taltal for example presented an MRE of 4.57% and a rank histogram with a clear overestimation of the lowest/highest bins for WRF-AnEn. This is no longer the case for WRFDDA-AnEn, in which the MRE decreased to -0.49% and the rank histogram is more uniform.

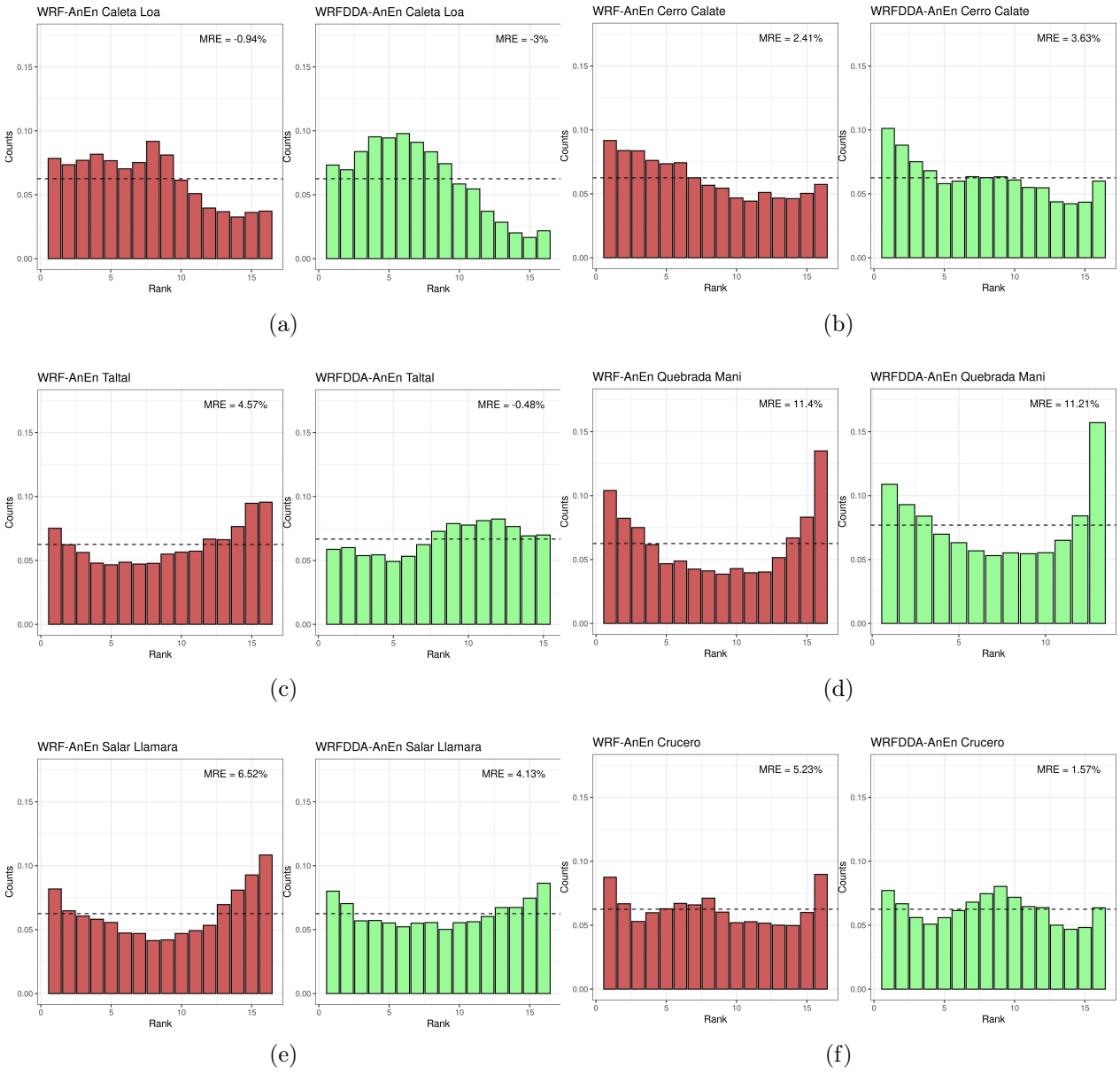


Figure 3.5: Rank histograms using analog ensemble with WRF (red) and WRFDDA (green) for each met station in the study.

3.3 Verification during cloudy periods

The previous section showed the results of the forecasts verification during clear and cloudy skies. Although clear sky verification is important to assess the difference between the values obtained by the forecasts and the measurements during all periods, forecasting during cloudy periods is more important, given the difficulties of modeling the solar irradiance in presence of clouds.

Figure 3.6 shows the Taylor Diagrams of the solar irradiance forecasts verification for each met station. This verification is based on the 15-min values and only considers daytime values under cloudy conditions. From this figure, it is possible to observe the improvement of the analog ensemble approach in sites close to the coast compared to the other forecasts (WRF and WRFDDA). Nevertheless, the improvement of the analog ensemble approach is lower on sites in the interior of the Atacama Desert compared to coastal sites. This may be related to the low frequency of cloudy events in the hyper-arid region of the Atacama Desert.

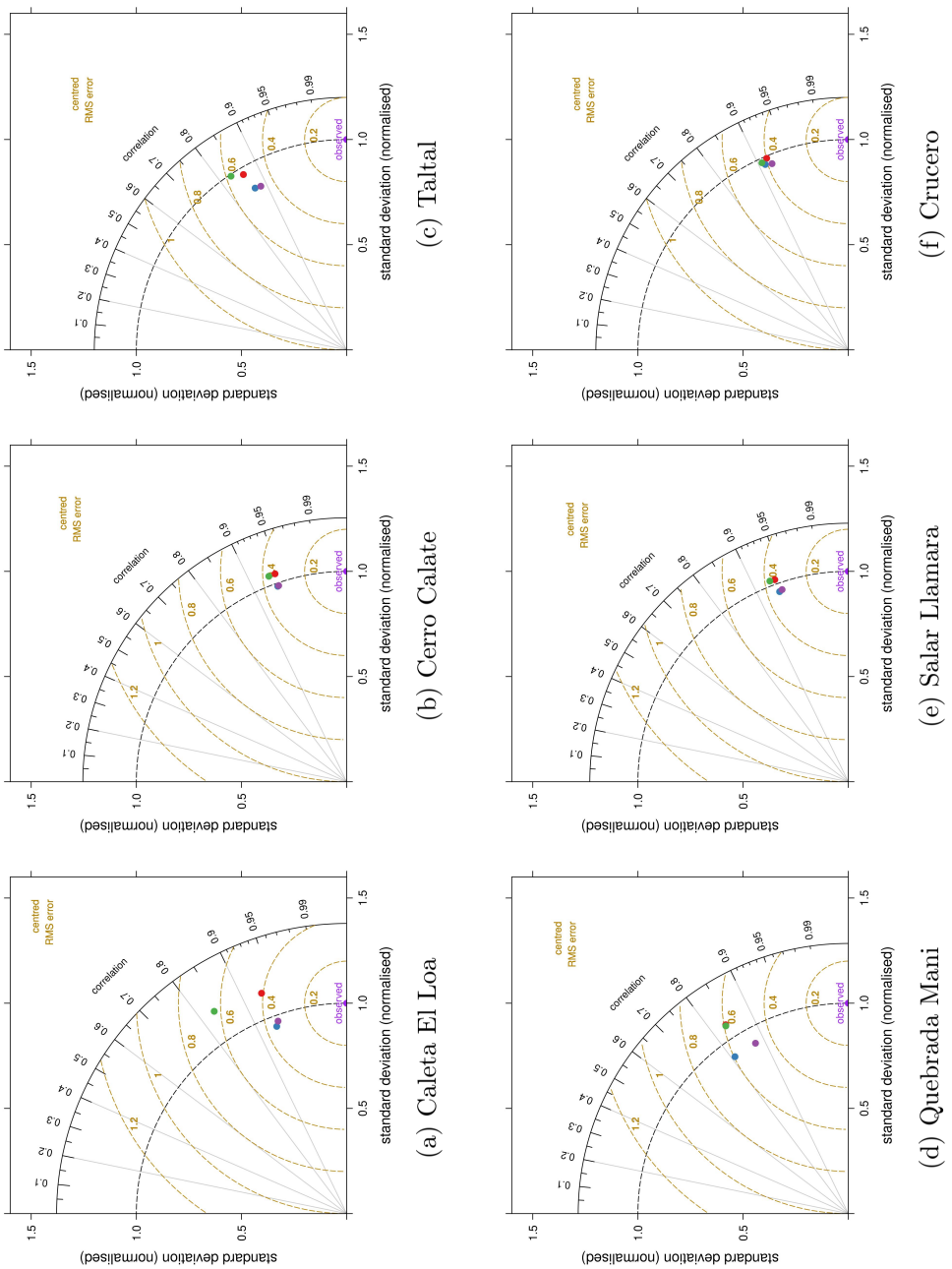
For example, for sites close to the coast such as Caleta Loa, Taltal, and Quebrada Mani (3.6 a, c and d respectively), there is a significant difference between WRF/WRFDDA and the forecasts obtained with AnEn. In particular, forecasts obtained with AnEn and WRFDDA are closer to the observation compared to the other forecasts. For sites at the interior of the Atacama Desert (3.6 b, e, and f), the forecasts are more similar to each other, but again, the forecasts produced with WRFDDA and AnEn have higher correlation values, lower RMS and have a standard deviation more similar to the observations.

Table 3.2 shows the RMSE and MAE obtained with cloudy skies for all the met stations. As in the previous scenario with clear and cloudy skies, the WRFDDA forecasts with the analog ensemble approach obtained the best results along with all the WRF forecasts. From this table and the Taylor Diagrams in Figure 3.6, it is possible to observe that RMSE and MAE from WRFDDA is higher compared to the other forecasts. The results obtained for WRFDDA may be explained by the effect that GOES-16 cloud data assimilation produced on WRFDDA forecasts during the first 12 hours. The WRFDDA approach may not be adding clouds with an accurate optical cloud depth in some areas, which may explain the overestimation of the solar irradiance. This is also observed in the time series presented in Appendix C. In these series, it is possible to observe that WRFDDA is always overestimating the solar irradiance during the first 12 hours of forecasts, even in periods when clouds are present.

Table 3.2: Comparison between WRF, WRFDDA, WRF-AnEn, and WRFDDA-AnEn. The results are for cloudy skies during the daytime between 4 and 48 hours after the forecast initialization.

	MAE [W/m^2]				RMSE [W/m^2]			
Site	WRF	WRF-AnEn	WRFDDA	WRFDDA-AnEn	WRF	WRF-AnEn	WRFDDA	WRFDDA-AnEn
Caleta Loa	130.4	75.3	149.7	72.0	172.1	110.4	200.2	105.7
Cerro Calate	84.3	70.9	86.0	68.5	140.7	123.3	146.3	121.3
Quebrada Mani	159.3	151.5	158.3	120.7	246.0	234.2	245.2	182.0
Salar Llamara	89.5	68.2	90.7	63.8	144.5	120.3	150.6	115.6
Taltal	133.9	118.8	135.3	112.1	207.6	168.2	207.3	158.6
Crucero	87.2	91.7	89.1	84.9	161.9	158.3	165.6	144.8
Mean	114.1	96.1	118.2	87.0	178.8	152.4	185.9	138.0

Although the WRFDDA forecasts seem to be overestimating the solar irradiance in some sites, the analog approach is able to correct the overprediction of the solar irradiance and create more accurate forecasts. Hence, despite the fact that WRFDDA is overestimating the solar irradiance, the more accurate cloud cover of this forecast allows selecting analog more similar to the observations in the historical periods, leading to more accurate forecasts of the solar irradiance.



(a) Caleta El Loa (b) Cerro Calate (c) Taltal

(d) Quebrada Mani (e) Salar Llamara (f) Crucero

Figure 3.6: Same as 3.4 but for cloudy sky periods.

3.4 Case Study: Taltal

Given its position close to the coast, Taltal was the surface meteorological station selected to assess the quality of the solar irradiance forecasts obtained in this study. In this section, a cloudy day was selected in order to compare all the solar forecasts obtained, to show how each one is reported, and also to discuss possible errors and solutions that could be implemented in the future.

Figure 3.7 shows the cloud cover according to GOES-16, WRF and WRFDDA for 2018-10-26 initialized at 12:00 UTC. From this Figure, it is possible to observe the stratocumulus cloud cover in front of the coast of the Antofagasta/Atacama Region, and also some cirrus clouds passing through the north part of the region, especially during October 26th. In this Figure, it is also possible to see the difference in the stratocumulus cover between WRF and WRFDDA. For WRF, the stratocumulus cover is not as dense compared to GOES-16 and WRFDDA and it does not cover as much area compared to GOES-16 or even WRFDDA. Hence, it seems that the dissipation of the stratocumulus cover according to WRF is much faster than in reality. On the other hand, the stratocumulus cloud cover is always close to the coastline according to WRFDDA, which is more similar to GOES-16. Although WRFDDA's stratocumulus cover is more similar to GOES-16, some differences are presented in the coastline, which can be difficult to model properly and that may also affect the cloud cover and solar irradiance forecast for Taltal. Nonetheless, from Figure 3.7 it is possible to observe that in terms of spatial coverage, the cloud cover from WRFDDA is more similar to GOES-16 than WRF.

Figure 3.8 (a), (b) and (c) show the solar irradiance forecasts for 2018-10-26 initialized at 12:00 UTC for WRF/WRF-AnEn and WRFDDA/WRFDDA-AnEn respectively. From this Figure, it is possible to observe that for cloudy days, the analog ensemble is able to capture the variability of the solar irradiance, given that the observation is within the limits of the analog ensemble interval found for that day. Although the difference between the minimum and maximum ensemble is large, the mean value of the ensemble is close to the observations.

Also, looking at the forecast obtained with WRF and WRFDDA (Figure 3.8.b and 3.8.c), it is possible to observe the difference between the solar irradiance forecast obtained from both methods. The solar irradiance obtained using WRFDDA is closer to the observation than the forecast obtained with WRF. This, as mentioned in the previous section, could be produced by the correction of the cloud cover for WRFDDA, which may have affected the solar irradiance values from the WRFDDA and WRFDDA-AnEn forecasts. As can be seen in 3.8.c, a sudden decrease of the solar irradiance is found at FLT = 3 hours, the time in which the GOES-16 data was incorporated into the forecasts. From this time, it is possible to observe that the difference between the observations and the WRFDDA-AnEn is lower than WRF-AnEn.

Between 4 hours and 10 hours after the initialization, the solar irradiation calculated for Taltal according to the observation was 4.57 kWh/m^2. Meanwhile, the solar irradiation and the relative error for the forecasts was 5.18 (12.9%), 5.15 (12.2%), 3.48 (-24.2%) and 4.12 (-10.3%) kWh/m^2 for WRF, WRFDDA, WRF-AnEn and WRFDDA-AnEn respectively. These values show that during the first day of forecasts, the WRFDDA-AnEn produces the solar

irradiance forecast with the lowest error while having the lowest spread according to Figure 3.8.c and the observations.

Between 4 and 48 hours after the initialization, the solar irradiation according to Taltal met station was 4.32 kWh/m^2/day. On the other hand, the solar irradiation (relative error) according to the forecasts was 7.21 (66.8 %), 6.94 (60.5%), 4.52 (-4.6%) and 4.83 (-11.6%) kWh/m^2/day for WRF, WRFDDA, WRF-AnEn and WRFDDA-AnEn respectively. These values show that the Analog Ensemble forecasts have the lowest error compared to WRF and WRFDDA. The lower error for WRF-AnEn compared to WRFDDA-AnEn may be due to a slight overestimation of solar irradiance on the second day for WRFDDA-AnEn.

It is worth noting that all forecasts produce different forecasts for cloudy conditions. From Figure 3.8, WRFDDA forecast starts to overestimate the solar irradiance a few hours after the incorporation of the GOES-16 data. Meanwhile, WRF overestimate the solar radiation for the entire forecast period and WRF-AnEn forecast seems to be underestimating the solar irradiance compared to WRFDDA-AnEn. The difference between all the forecasts may be explained by the accuracy of the cloud cover of WRF and WRFDDA and by how WRF-AnEn and WRFDDA-AnEn considers cloud covers to select their analogs.

As shown in Figure 3.8.d, the cloud cover forecast (1 is cloudy and 0 is clear) shows that the WRFDDA cloud cover is more similar to GOES-16 compared to WRF, specially during the afternoon and night. As discussed previously, although the difference between the solar forecast and the measurements increase throughout the day for WRFDDA, it seems that this is not the case for WRFDDA-AnEn, as the forecasts seems to be closer to the measurements compared to WRF and WRF-AnEn.

Moreover, and as shown in Table 2.4, the weights that minimize the RMSE for the WRF-AnEn for Taltal do not take into account the cloud cover as a variable used to find analogs in the historical series. The reason for this may be caused by the cloud cover forecast of the WRF model, which is worse than the WRFDDA forecasts as shown in the previous section. Given that the cloud cover of the WRF forecasts is not as reliable as the WRFDDA forecasts, the analog ensemble approach using the WRF forecasts is only correcting values based only on the solar irradiance values and the time of the day (zenith angle), which is connected with the cloud cover diurnal cycle for coastal sites at the Atacama Desert. Hence, WRF-AnEn may be producing This may explain the overestimation of the solar irradiance on the 10 hours of forecast compared to WRFDDA-AnEn.

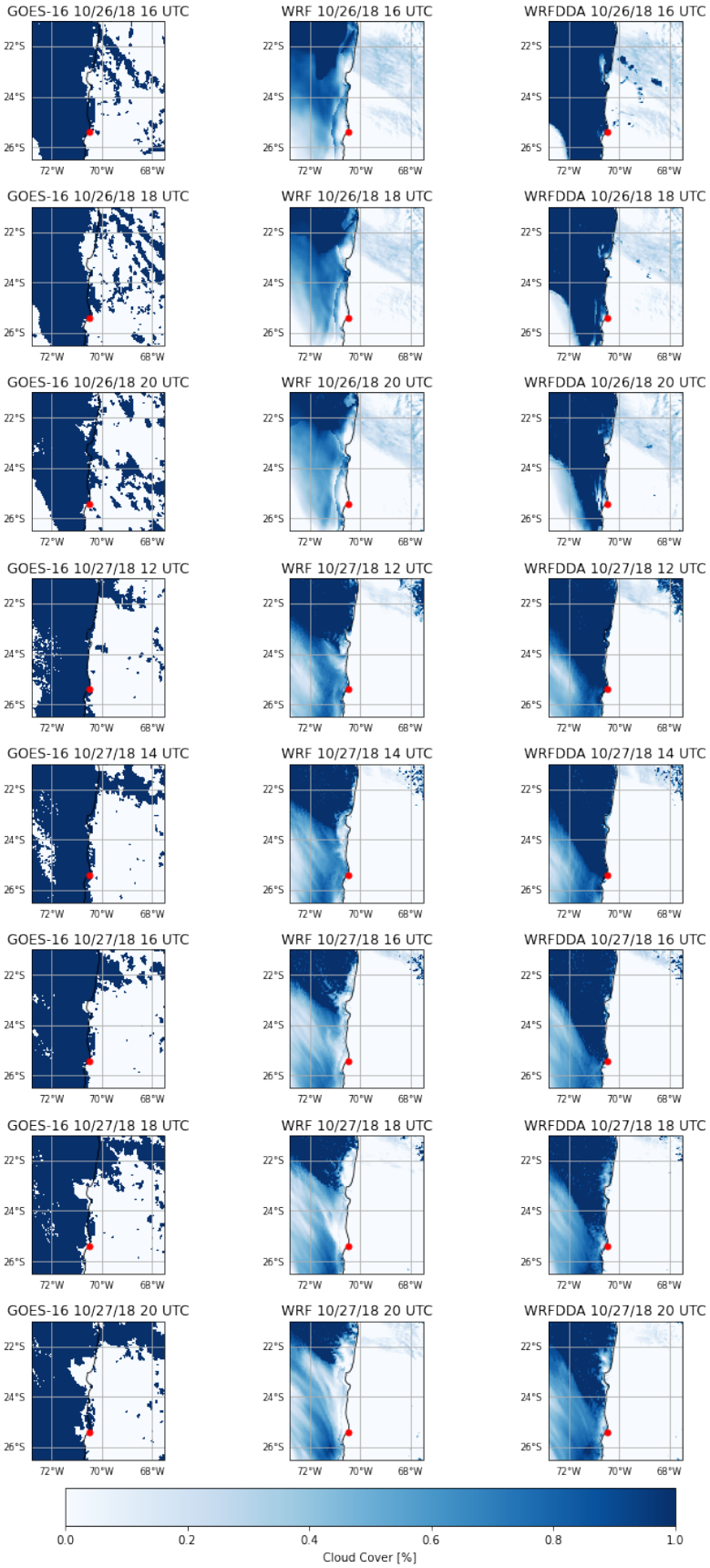


Figure 3.7: Cloud Cover for 26-10-2018 and 27-10-2018 according to GOES-16, WRF and WRFDDA. The red dot in the map corresponds to the Taltal Meteorological station.

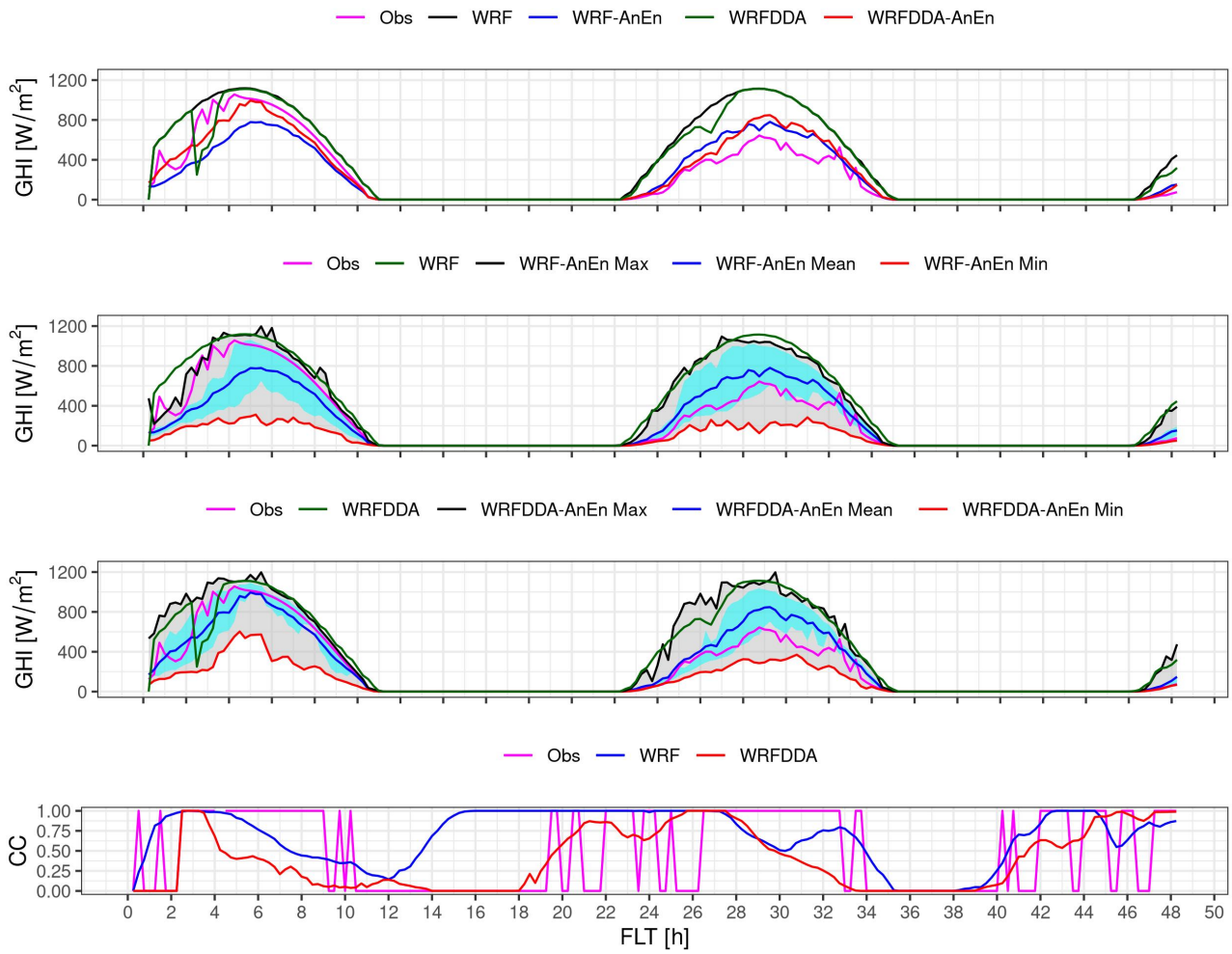


Figure 3.8: Solar irradiance forecast between WRF, WRFDDA, WRF-AnEn and WRFDDA-AnEn for October 26th and 27th, 2018

3.5 Case Study: Crucero

The meteorological station Crucero is located in a region with a high development of solar energy projects. The hyper aridity of this area and its low cloudiness makes this part of the Atacama region the perfect spot for solar photovoltaic or concentration energy projects. However, during the summer, clouds from the intense convection in the Altiplano can reach this area, reducing the solar irradiance during January, February and March.

Figure 3.9 shows the cloud cover for the Atacama Desert during an intense convection event on February 10th and 11th, 2019. In this figure is possible to observe cloud cover values over 0.2 for most of the Altiplano region and the hyper-arid zone according to WRF and WRFDDA. Clouds of different thicknesses are present in the region during this day. For this Case, WRFDDA may not be relevant, given that most of the clouds present during these events are medium to high clouds. Hence, the solar irradiance forecasts for these events rely on the accuracy of each method to predict the solar irradiance in cloudy periods, without the benefit of the correction by observed clouds.

Figure 3.10 shows the solar irradiance forecast for this day using the 4 methods proposed in this study, together with the cloud cover forecast. In this case, it is possible to observe that both WRF-AnEn and WRFDDA-AnEn are able to forecast the solar irradiance variability produced by cloudy periods and that the solar irradiance observed is within the minimum and maximum solar irradiance in the ensemble forecast. The ensemble forecast created could be used by solar plant owners to estimate the uncertainty of the forecasts and help the plant operation for the following hours or days.

For this case, the mean daily solar irradiation forecast and the relative error (%) between 4-48 hours after the initialization is 6.70 kWh/m^2/day for the observations and 7.31 (9.3%), 6.43 (-4.1%), 6.33 (-5.6%) and 6.45 (-3.8%) kWh/m^2/day for WRF, WRFDDA, WRF-AnEn and WRFDDA-AnEn respectively. The lowest error was obtained with the WRFDDA-AnEn method, followed by WRFDDA. Although the WRFDDA forecast for the first day underestimated the solar irradiance in Crucero, for the second day the solar irradiance forecast of WRFDDA was closer to the observations than the other methods.

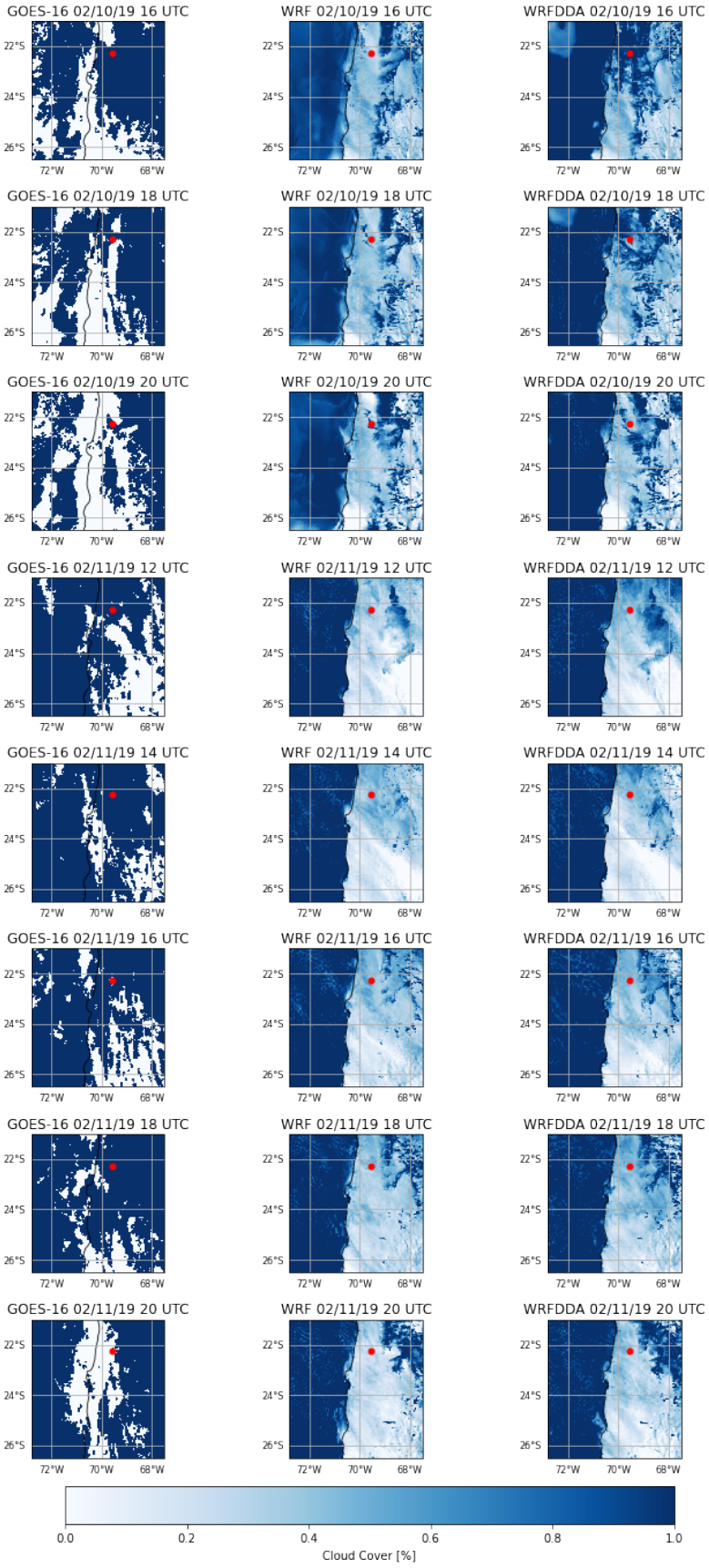


Figure 3.9: Cloud Cover for 10-02-2019 and 11-02-2019 according to GOES-16, WRF and WRFDDA. The red dot in the map corresponds to the Crucero Meteorological station.

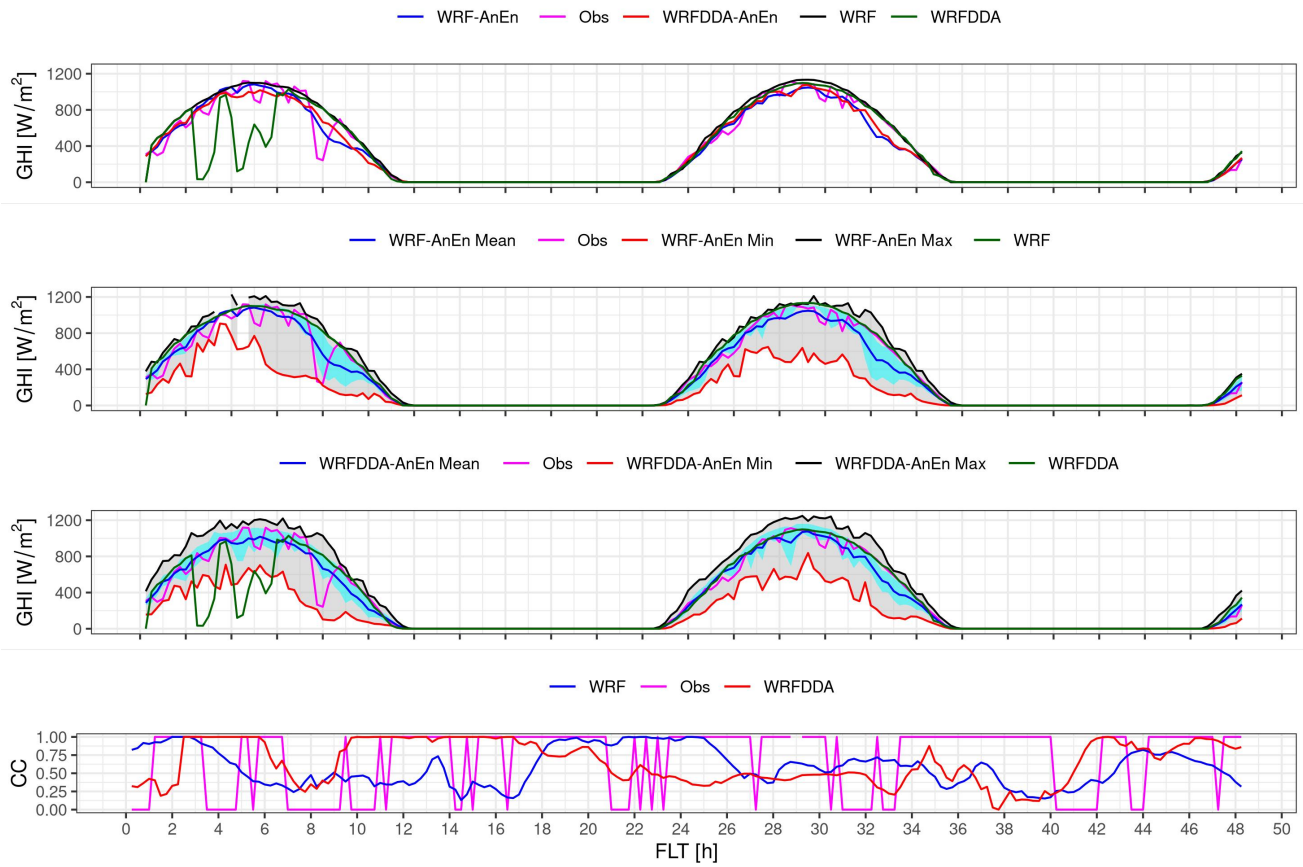


Figure 3.10: Solar irradiance forecast between WRF, WRFDDA, WRF-AnEn and WRFDDA-AnEn for February 10th and 10th, 2019

Chapter 4

Conclusions

In the previous sections, 4 different solar irradiance forecasts were obtained and compared to observational data in 6 different sites in the Atacama Desert. The results found in this study show different outcomes depending on the location of the site, the forwarding lead time, and the method used to obtain the forecasts. Overall, it was found in the verification that forecasts produced using the analog ensemble technique obtained better results than forecasts created using only WRF. Also, a small improvement on the cloud cover forecasts was obtained using a cloud assimilation technique, which leads to more accurate solar irradiance forecasts when it was used together with the analog ensemble method.

According to the verification of the cloud cover forecast, it was found that the inclusion of cloud data from GOES-16 produced an improvement compared to the case in which the forecasts relied only on the initial/boundary condition given by GFS. The results show an improvement of the HSS score for periods closer to the initialization and the assimilation of cloud data, which impacted the solar irradiance forecast during those hours.

The reason for this is that accurate cloud cover forecasts in AnEn configuration that relies on cloud cover to select historical forecasts, will select analogs that are more similar to the present observations. For example, it was found in Taltal that the WRF-AnEn configuration with the lower RMSE was the one without cloud cover as the variable used to select analog members, which means that the cloud cover forecasts were not reliable enough to select analogs that could correct the solar irradiance forecasts. Hence, it was found that accurate cloud cover forecasts are important not only to produce a reliable solar irradiance forecast with WRF, but also to select good analogs members that can correct the forecasts.

About the improvement of WRFDDA, it was found that the improvement found was limited to a specific period (4-8 hours after the initialization), as the forecasts started to behave similarly to WRF as the model keep running till the end of the forecasts. The reason for this behavior could be related to the vertical profile used in the cloud assimilation to detect and correct the cloud cover. This, because the vertical profile used to create/destroy clouds during the assimilation may not be the same as the climatology of the vertical profile of the coast of the Atacama Desert. It is recommended that a new vertical profile climatology using radiosondes or cloud water mixing ratio vertical profile together with cloud top height

measurements is used to build an empirical vertical profile more suitable for the coast of the Atacama Desert.

Additionally, the coarse grid size and the uncertainty given by the use of a model such as GFS to initialize the WRF model may have also been responsible for the differences found in the cloud cover forecasts compared to the GOES-16 observations. Given how sensitive are forecasts to initial/boundary conditions, it is recommended that other global, weather forecasting models or regional forecast models can be tested along with GFS.

Nonetheless, the improvement of the cloud cover forecast obtained with WRFDDA impacted positively on the solar irradiance forecasts, and the forecasts created using the analog ensemble method. For example, taking clear and cloudy skies into consideration, and using GFS as a base case for solar forecasting, on average the improvement of the RMSE for WRF-AnEn and WRFDDA-AnEn is 3.84% and 1.59% respectively. On the other hand, the improvement on average of the RMSE for WRF and WRFDDA is only 0.72% and 0.43% respectively. Although the improvement may not be considered substantial, a 2-3% of improvement may be significant to photovoltaic plant owners, which are looking for ways to maximize revenues and reduce uncertainty on their solar irradiance forecasts.

In summary, the forecasts presented in this study are able to predict the solar irradiance variability in the Atacama Desert. Some recommendations are presented below, with the objective of improving the results presented in this study:

- To increase the length of both the historical observation database and the forecasts database. This, in order to obtain more accurate analogs that can correct and predict more accurately the solar irradiance forecasts. The latter is more critical for sites in the interior, in which the frequency of cloudiness is rarer, limiting the selection of proper analogs.
- To incorporate cloud measurement in each site, in order to select more accurate analogs based on the cloud cover. Given the grid size of GOES-16 and the horizontal variation of the cloud cover, it is important to incorporate high frequency and on-site measurement of cloud cover, in order to increase the quality of the analog selected. Cloud measurements using ceilometers or instruments that can measure the cloud optical depth could be helpful in obtaining quality cloud measurements that can help the selection of good analogs.
- To incorporate aerosol measurements into the solar irradiance forecasts. Although the analog ensemble approach is able to correct the bias that the aerosol optical depth is producing on the solar irradiance forecasts, it is important to measure this variable in order to control the effect of this variable on the solar irradiance forecasts. It is expected that the inclusion of this variable can improve the selection of proper analogs, which could lead to more accurate solar irradiance forecasts.

Bibliography

[1] Ministerio de Energia, "Energia 2050," 2015.

[2] R. A. Escobar, C. Cortés, A. Pino, E. B. Pereira, F. R. Martins, and J. M. Cardemil, "Solar energy resource assessment in Chile: Satellite estimation and ground station measurements," *Renewable Energy*, vol. 71, pp. 324–332, nov 2014. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S096014811400264X

[3] R. A. Escobar, C. Cortés, A. Pino, M. Salgado, E. B. Pereira, F. R. Martins, J. Boland, and J. M. Cardemil, "Estimating the potential for solar energy utilization in Chile by satellite-derived data and ground station measurements," *Solar Energy*, vol. 121, pp. 139–151, 2015. [Online]. Available: https://ac.els-cdn.com/S0038092X15004703/1-s2.0-S0038092X15004703-main.pdf?{_}tid=261ecad0-fc80-11e7-aeb1-00000aacb35e{&}acdnat=1516301478{_}8a9fbe675e82c758a1d4f518bbeb37cc

[4] P. Campos, L. Troncoso, P. D. Lund, C. Cuevas, A. Fissore, and R. Garcia, "Potential of distributed photovoltaics in urban Chile," *Solar Energy*, vol. 135, pp. 43–49, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.solener.2016.05.043

[5] R. Rondanelli, A. Molina, and M. Falvey, "The Atacama surface solar maximum," *Bulletin of the American Meteorological Society*, vol. 96, no. 3, pp. 405–418, 2015.

[6] A. Molina, M. Falvey, and R. Rondanelli, "A solar radiation database for Chile," *Scientific Reports*, vol. 7, no. 1, 2017.

[7] J. Haas, R. Palma-Behnke, F. Valencia, P. Araya, G. Díaz-Ferrán, T. Telsnig, L. Eltrop, M. Díaz, S. Püschel, M. Grandel, R. Román, and G. Jiménez-Estévez, "Sunset or sunrise? Understanding the barriers and options for the massive deployment of solar technologies in Chile," *Energy Policy*, vol. 112, pp. 399–414, jan 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301421517306249

[8] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," 2016.

[9] C. Brancucci Martinez-Anido, B. Botor, A. R. Florita, C. Draxl, S. Lu, H. F. Hamann, and B. M. Hodge, "The value of day-ahead solar power forecasting improvement," *Solar Energy*, 2016.

[10] D. Yang, J. Kleissl, C. A. Gueymard, H. T. Pedro, and C. F. Coimbra, "History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining," *Solar Energy*, 2018.

[11] D. K. Sahu, H. Yang, and J. Kleissl, "Assimilating observations to simulate marine layer stratocumulus for solar forecasting," *Solar Energy*, 2018.

[12] G. Huang, Z. Li, X. Li, S. Liang, K. Yang, D. Wang, and Y. Zhang, "Estimating surface solar irradiance from satellites: Past, present, and future perspectives," *Remote Sensing of Environment*, vol. 233, p. 111371, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425719303906

[13] J. Kleissl, C. F. Coimbra, and R. Marquez, *Solar Energy Forecasting and Resource Assessment*, 2013.

[14] J. A. Rutllant, H. Fuenzalida, and P. Aceituno, "Climate dynamics along the arid northern coast of Chile: The 1997–1998 Dinámica del Clima de la Región de Antofagasta (DICLIMA) experiment," *Journal of Geophysical Research*, 2003.

[15] R. D. Garreaud, "The climate of northern Chile: Mean state, variability and trends," in *Revista Mexicana de Astronomia y Astrofisica: Serie de Conferencias*, 2011.

[16] S. J. Abel, D. N. Walters, and G. Allen, "Evaluation of stratocumulus cloud prediction in the Met Office forecast model during VOCALS-REx," *Atmospheric Chemistry and Physics*, 2010.

[17] D. Pozo, J. C. Marín, L. Illanes, M. Curé, and D. Rabanus, "Validation of WRF forecasts for the Chajnantor region," *Monthly Notices of the Royal Astronomical Society*, 2016.

[18] M. Reyers, C. Boehm, L. Knarr, Y. Shao, and S. Crewell, "Synoptic-to-regional-scale analysis of rainfall in the atacama desert (18°-26°S) using a long-term simulation with WRF," *Monthly Weather Review*, 2020.

[19] R. C. Muñoz, J. Quintana, M. J. Falvey, J. A. Rutllant, and R. Garreaud, "Coastal clouds at the Eastern Margin of the Southeast Pacific: Climatology and trends," *Journal of Climate*, 2016.

[20] Y. Gu, K. Liou, S. Ou, and R. Fovell, "Cirrus cloud simulations using wrf with improved radiation parameterization and increased vertical resolution," *Journal of Geophysical Research: Atmospheres*, vol. 116, no. D6, 2011.

[21] R. Stull, *Practical Meteorology: An Algebra-Based Survey of Atmospheric Science*, ser. BC Open Textbook Collection. AVP International, University of British Columbia, 2016. [Online]. Available: https://books.google.cl/books?id=xP2sDAEACAAJ

[22] L. D. Monache, F. Anthony Eckel, D. L. Rife, B. Nagarajan, and K. Searight, "Probabilistic weather prediction with an analog ensemble," *Monthly Weather Review*, 2013.

[23] S. Alessandrini, L. Delle Monache, S. Sperati, and G. Cervone, "An analog ensemble for

short-term probabilistic solar power forecast," *Applied Energy*, 2015.

[24] S. E. Haupt, B. Kosović, T. Jensen, J. K. Lazo, J. A. Lee, P. A. Jiménez, J. Cowie, G. Wiener, T. C. McCandless, M. Rogers, S. Miller, M. Sengupta, Y. Xie, L. Hinkelman, P. Kalb, and J. Heiser, "Building the Sun4Cast System: Improvements in Solar Power Forecasting," *Bulletin of the American Meteorological Society*, 2018.

[25] G. Cervone, L. Clemente-Harding, S. Alessandrini, and L. Delle Monache, "Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble," *Renewable Energy*, 2017.

[26] X. Zhang, Y. Li, S. Lu, H. F. Hamann, B. M. Hodge, and B. Lehman, "A Solar Time Based Analog Ensemble Method for Regional Solar Power Forecasting," *IEEE Transactions on Sustainable Energy*, 2019.

[27] P. A. Jimenez, J. P. Hacker, J. Dudhia, S. E. Haupt, J. A. Ruiz-Arias, C. A. Gueymard, G. Thompson, T. Eidhammer, and A. Deng, "Wrf-solar: Description and clear-sky assessment of an augmented nwp model for solar power prediction," *Bulletin of the American Meteorological Society*, vol. 97, no. 7, pp. 1249–1264, 2016.

[28] National Centers for Environmental Prediction, National Weather Service, NOAA, U.S. Department of Commerce, "NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive," Boulder CO, 2015. [Online]. Available: https://doi.org/10.5065/D65D8PWK

[29] R. C. Muñoz, M. J. Falvey, M. Arancibia, V. I. Astudillo, J. Elgueta, M. Ibarra, C. Santana, and C. Vásquez, "Wind energy exploration over the atacama desert : A numerical model-guided observational program," *Bulletin of the American Meteorological Society*, 2018.

[30] R. A. Martínez, "Modelación numérica de chorros nocturnos de bajo nivel en valles del desierto de Atacama: Evaluación y diagnóstico," 2018. [Online]. Available: http://repositorio.uchile.cl/handle/2250/168478

[31] D. Hoffmeister, "Meteorological and soil measurements of the permanent weather stations in the atacama desert, chile," 2018. [Online]. Available: https://www.crc1211db.uni-koeln.de//DOI/doi.php?doiID=1

[32] W. F. Holmgren, C. W. Hansen, and M. A. Mikofski, "pvlib python: a python package for modeling solar energy systems," *Journal of Open Source Software*, vol. 3, no. 29, p. 884, Sep. 2018. [Online]. Available: https://doi.org/10.21105/joss.00884

[33] D. Yang, S. Alessandrini, J. Antonanzas, F. Antonanzas-Torres, V. Badescu, H. G. Beyer, R. Blaga, J. Boland, J. M. Bright, C. F. Coimbra, M. David, Â. Frimane, C. A. Gueymard, T. Hong, M. J. Kay, S. Killinger, J. Kleissl, P. Lauret, E. Lorenz, D. van der Meer, M. Paulescu, R. Perez, O. Perpiñán-Lamigueiro, I. M. Peters, G. Reikard, D. Renné, Y. M. Saint-Drenan, Y. Shuai, R. Urraca, H. Verbois, F. Vignola, C. Voyant, and J. Zhang, "Verification of deterministic solar forecasts," *Solar Energy*, vol. 210, pp. 20–37, nov 2020.

[34] V. Lara-Fanego, J. Ruiz-Arias, D. Pozo-Vázquez, F. Santos-Alamillos, and J. Tovar-Pescador, "Evaluation of the wrf model solar irradiance forecasts in andalusia (southern spain)," *Solar Energy*, vol. 86, no. 8, pp. 2200–2217, 2012.

[35] P. G. Kosmopoulos, S. Kazadzis, M. Taylor, E. Athanasopoulou, O. Speyer, P. I. Raptis, E. Marinou, E. Proestakis, S. Solomos, E. Gerasopoulos *et al.*, "Dust impact on surface solar irradiance assessed with model simulations, satellite observations and ground-based measurements," *Atmospheric Measurement Techniques*, vol. 10, no. 7, pp. 2435–2453, 2017.

[36] S. Federico, R. C. Torcasio, P. Sanò, D. Casella, M. Campanelli, J. F. Meirink, P. Wang, S. Vergari, H. Diémoz, and S. Dietrich, "Comparison of hourly surface downwelling solar radiation estimated from msg–seviri and forecast by the rams model with pyranometers over italy," *Atmospheric Measurement Techniques*, vol. 10, no. 6, pp. 2337–2352, 2017.

[37] O. Hyvärinen, "A probabilistic derivation of heidke skill score," *Weather and Forecasting*, vol. 29, no. 1, pp. 177–181, feb 2014. [Online]. Available: https://journals.ametsoc.org/view/journals/wefo/29/1/waf-d-13-00103{_}1.xml

[38] R. J. Hogan, E. J. O'Connor, and A. J. Illingworth, "Verification of cloud-fraction forecasts," *Quarterly Journal of the Royal Meteorological Society*, 2009.

[39] I. T. Jolliffe and D. B. Stephenson, *Forecast verification: a practitioner's guide in atmospheric science*. John Wiley & Sons, 2012.

[40] P. Mathiesen, C. Collier, and J. Kleissl, "A high-resolution, cloud-assimilating numerical weather prediction model for solar irradiance forecasting," *Solar Energy*, 2013.

[41] R. B. Stull, *An introduction to boundary layer meteorology*. Springer Science & Business Media, 2012, vol. 13.

[42] W. Hu, G. Cervone, L. Clemente-Harding, and M. Calovi, "Parallel analog ensemble," 2019. [Online]. Available: https://weiming-hu.github.io/AnalogsEnsemble/

[43] A. Heidinger, "NOAA NESDIS CENTER for SATELLITE APPLICATIONS and RESEARCH ALGORITHM THEORETICAL BASIS DOCUMENT ABI Cloud Mask," Tech. Rep., 2012.

[44] W. Hu, G. Cervone, G. Young, and L. D. Monache, "Weather analogs with a machine learning similarity metric for renewable resource forecasting," *arXiv preprint arXiv:2103.04530*, 2021.

Annexes

Annexed A: HSS scores for WRF and WRFDDA forecast for all met stations.

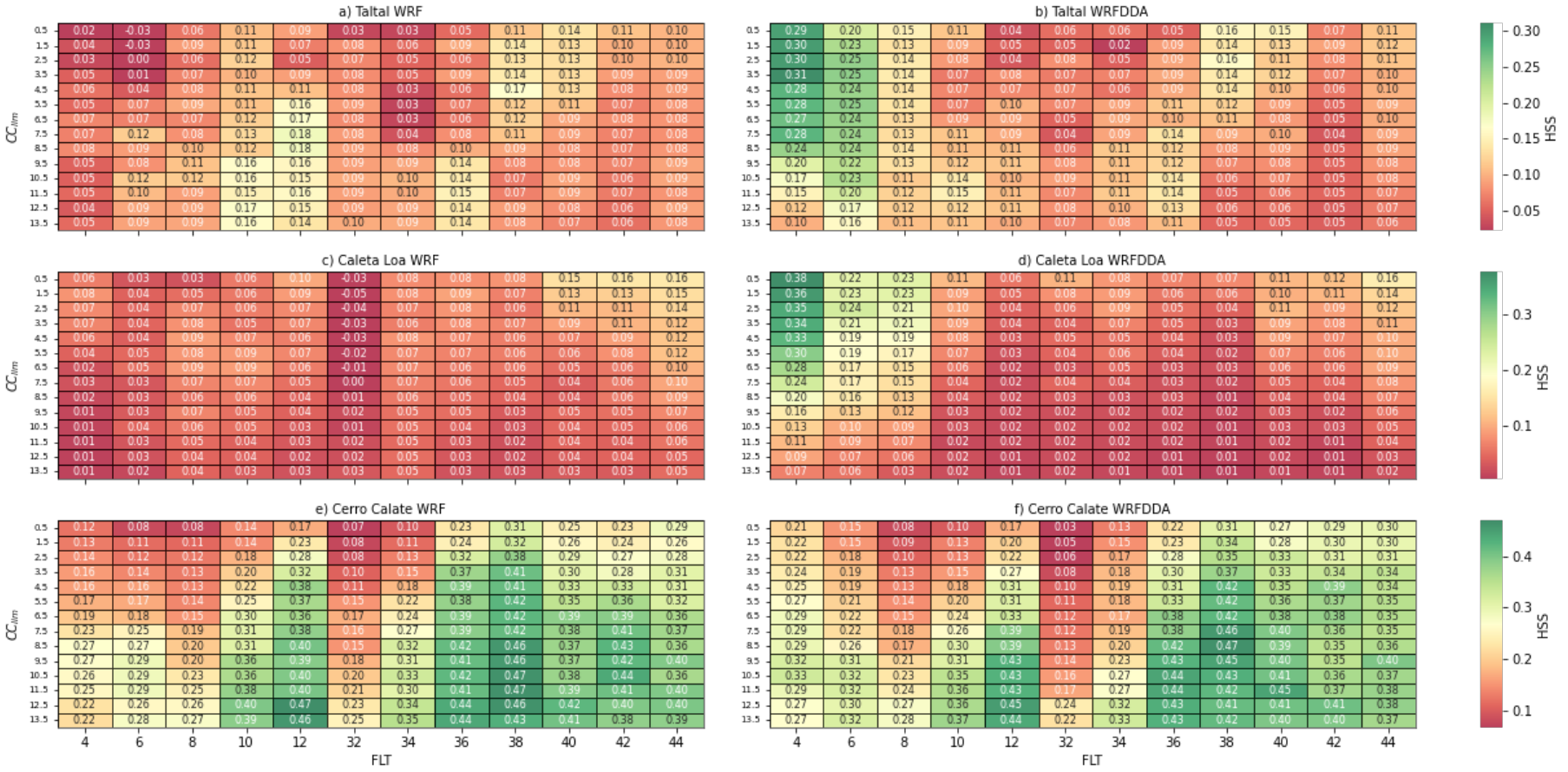


Figure A.1 : HSS skill score for the cloud cover forecasts for Taltal, Caleta el Loa and Cerro Calate using WRF and WRFDDA.

Figure A.2 : HSS skill score for the cloud cover forecasts for Crucero, Quebrada Mani and Salar Llamara using WRF and WRFDDA.

Annexed B: Forecasts comparison against GFS.

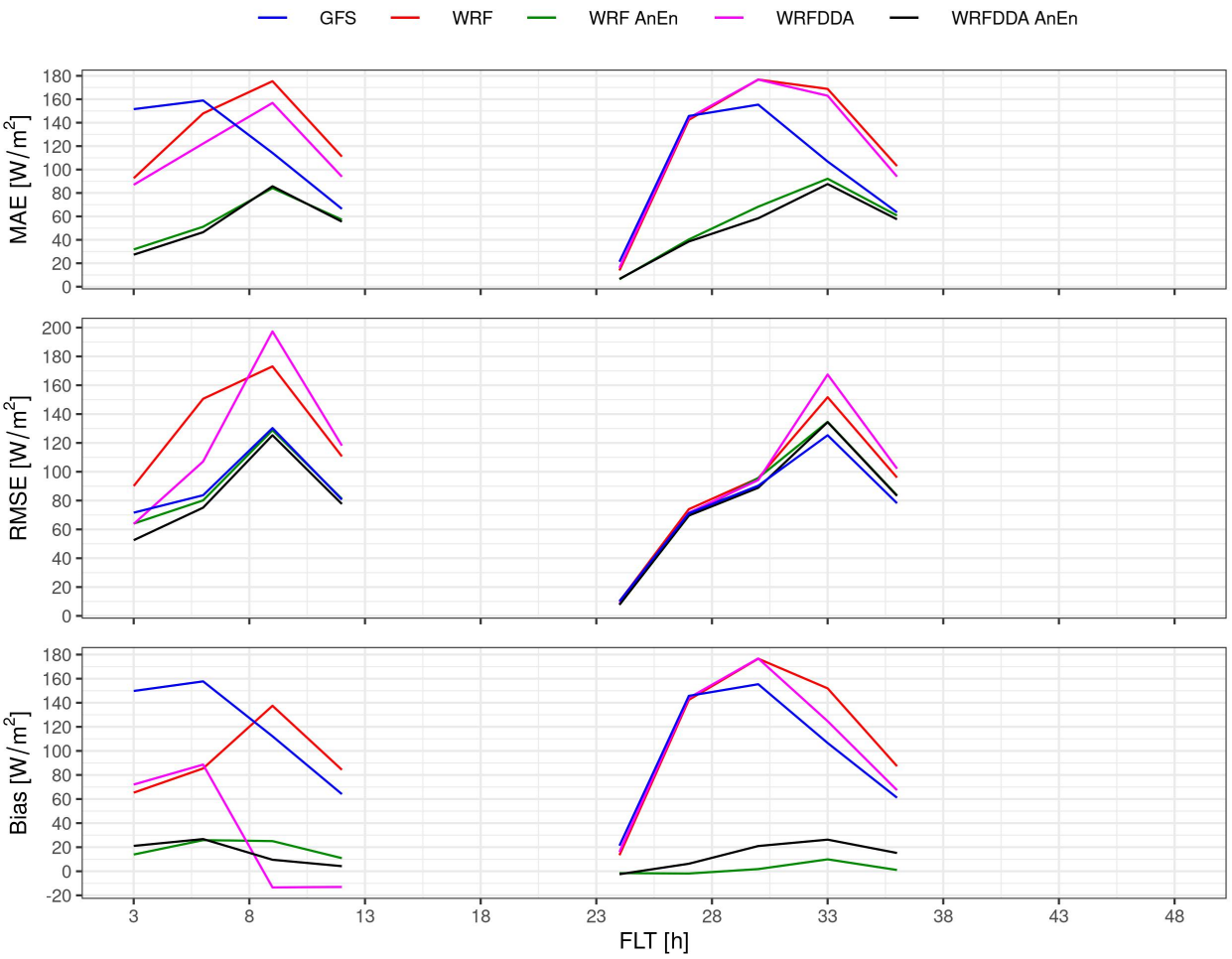


Figure B.1 : MAE, RMSE and Bias between GFS, WRF, WRF-AnEn, WRFDDA and WRFDDA-AnEn for Caleta Loa .

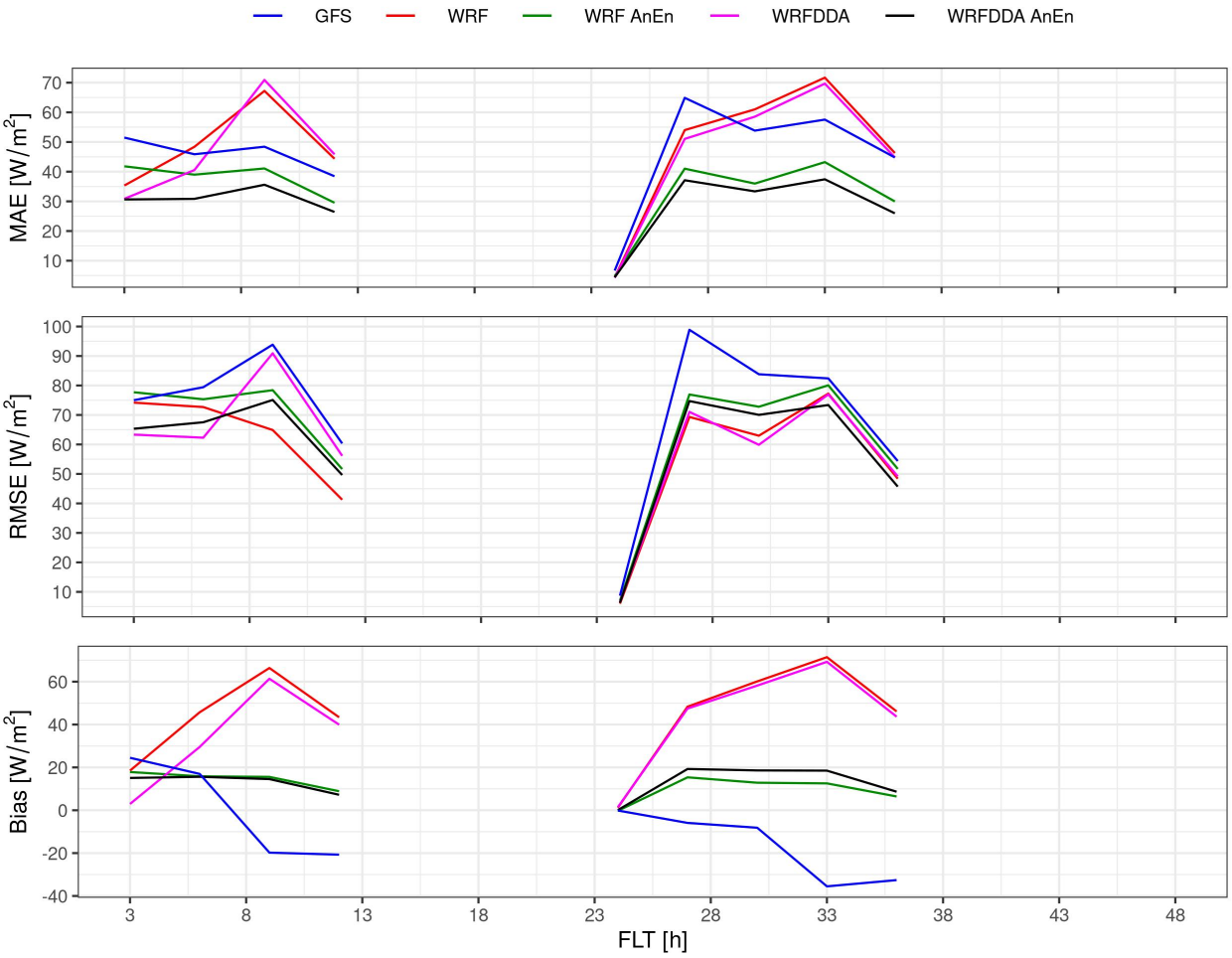


Figure B.2 : MAE, RMSE and Bias between GFS, WRF, WRF-AnEn, WRFDDA and WRFDDA-AnEn for Cerro Calate .

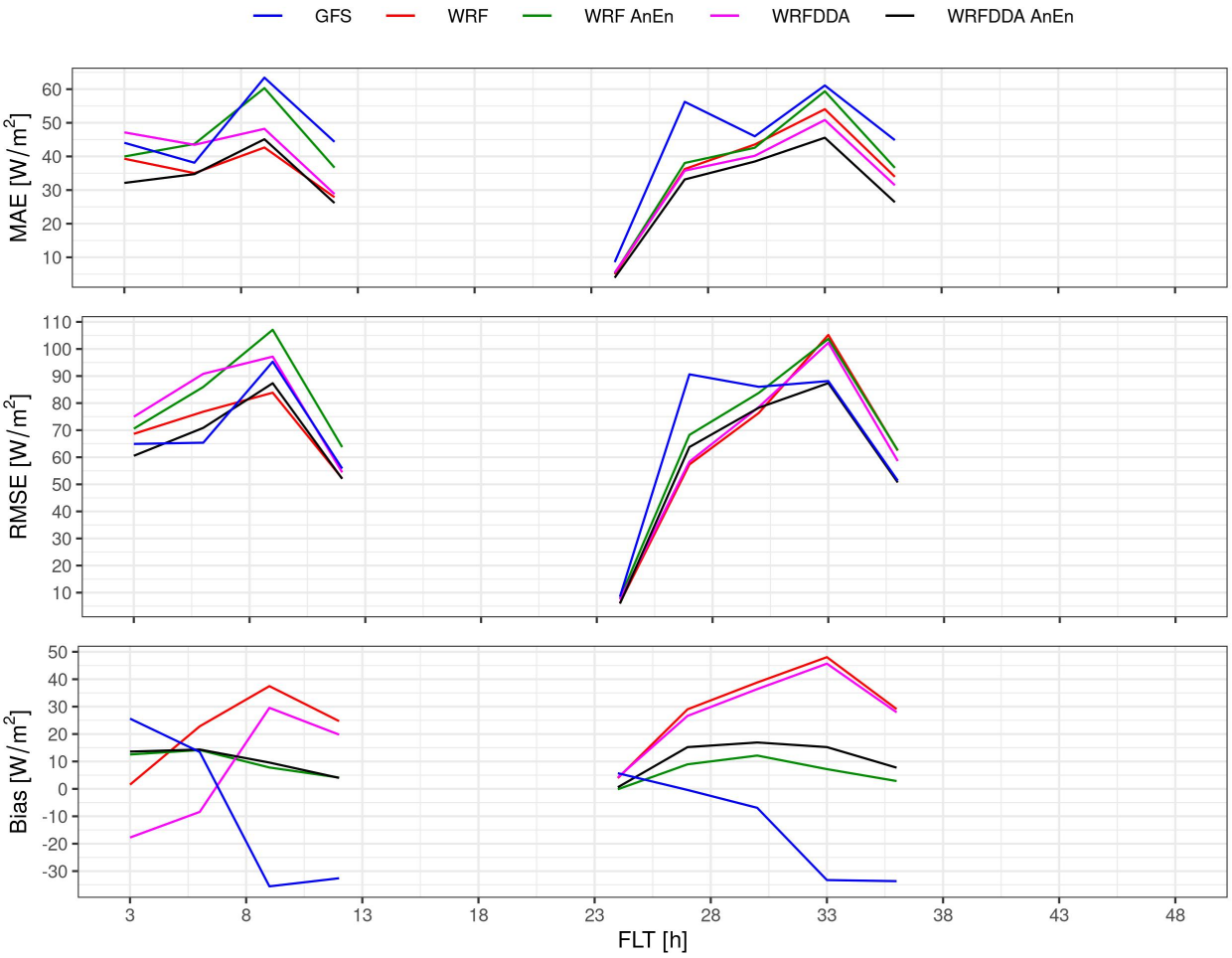


Figure B.3 : MAE, RMSE and Bias between GFS, WRF, WRF-AnEn, WRFDDA and WRFDDA-AnEn for Crucero .

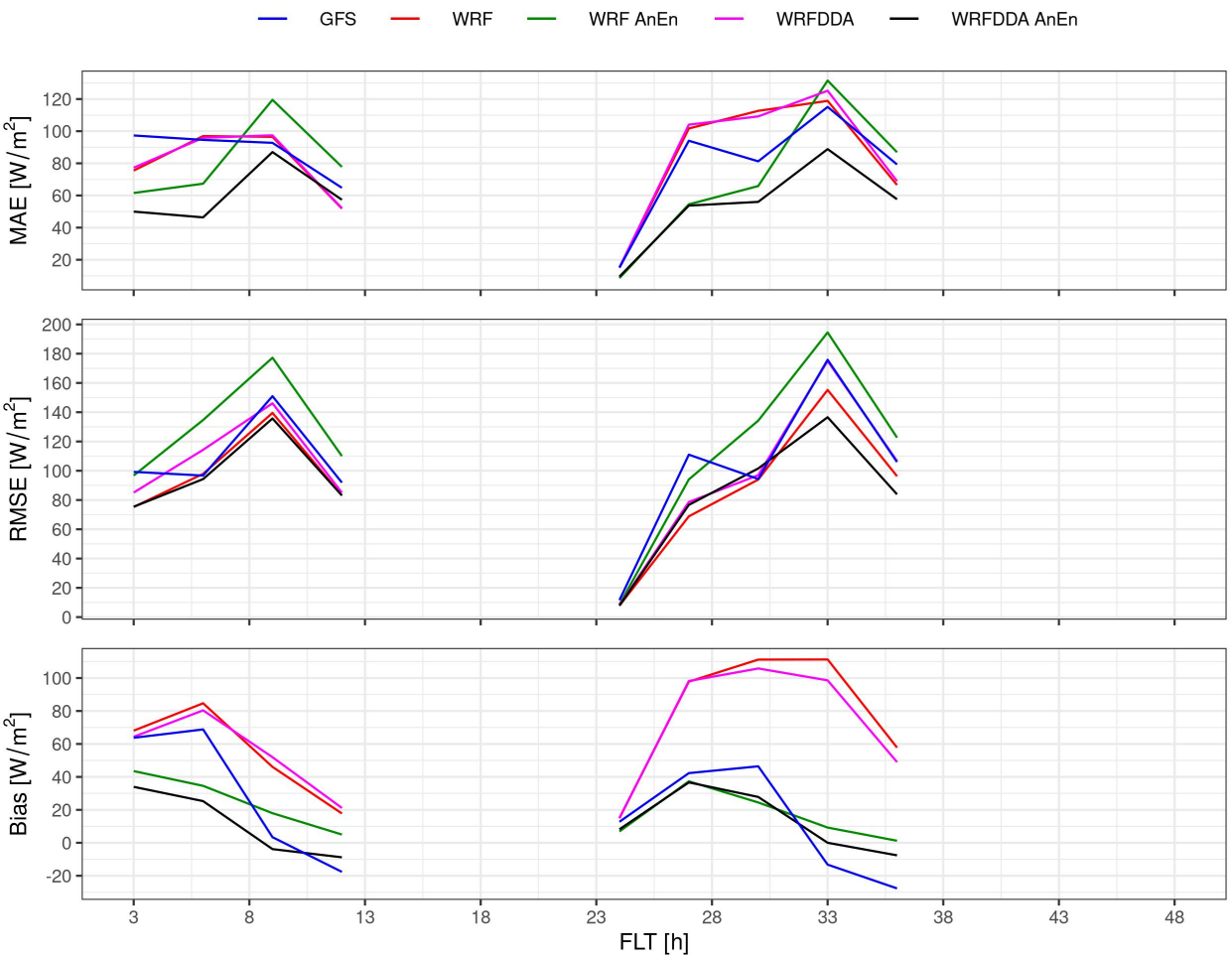


Figure B.4 : MAE, RMSE and Bias between GFS, WRF, WRF-AnEn, WRFDDA and WRFDDA-AnEn for Quebrada Mani .

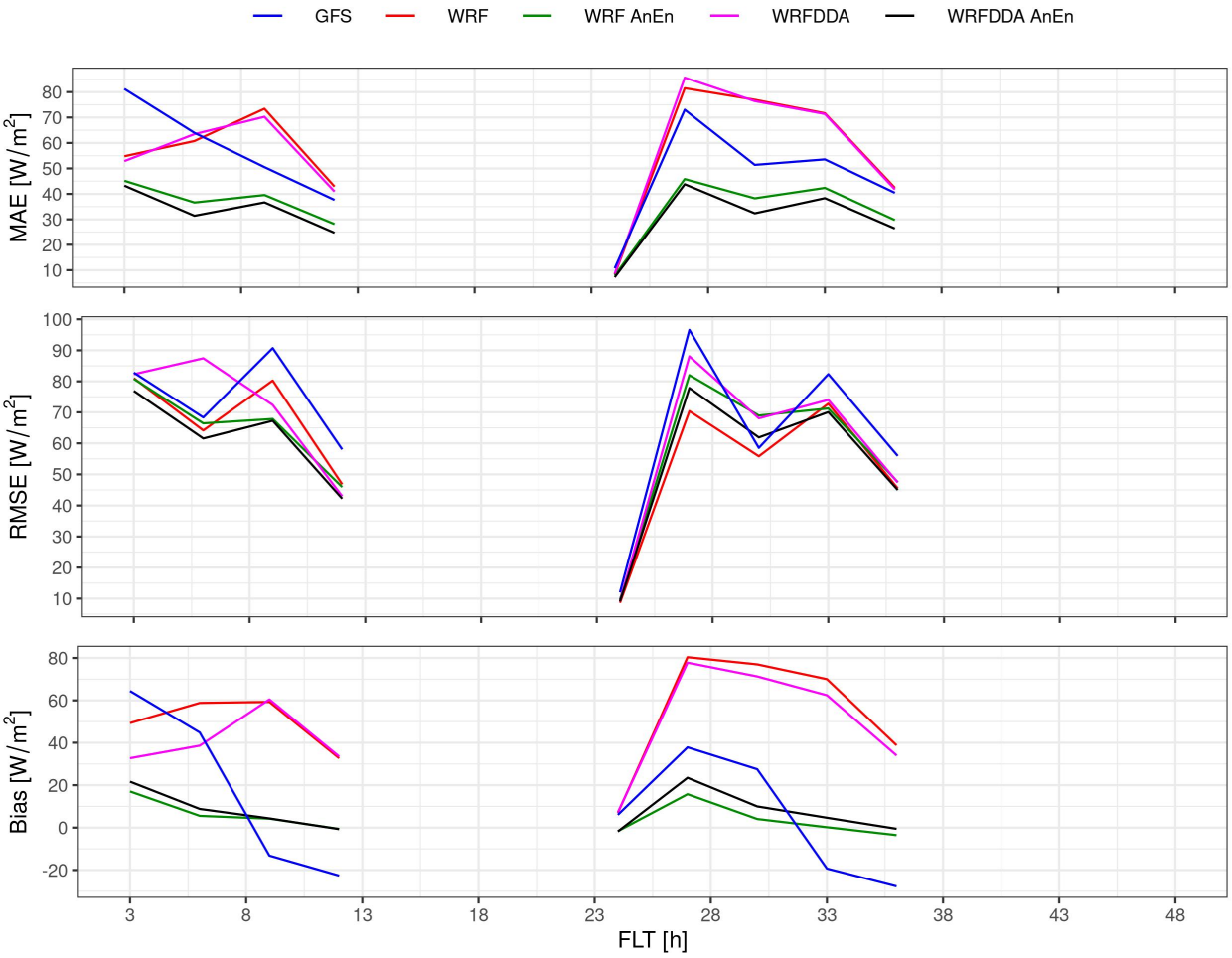


Figure B.5 : MAE, RMSE and Bias between GFS, WRF, WRF-AnEn, WRFDDA and WRFDDA-AnEn for Salar Llamara .

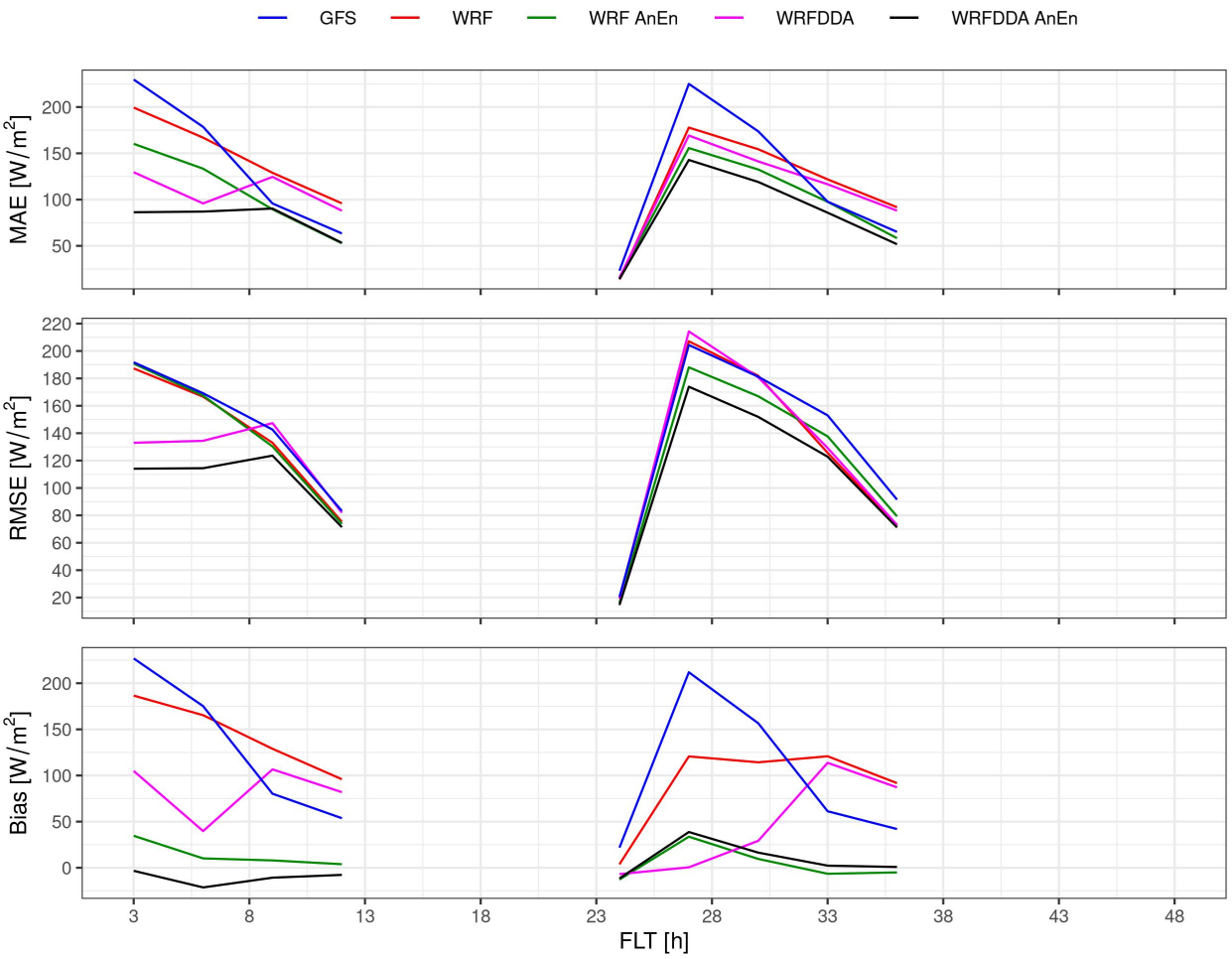


Figure B.6 : MAE, RMSE and Bias between GFS, WRF, WRF-AnEn, WRFDDA and WRFDDA-AnEn for Taltal.

Annexed C: Forecasts comparison for cloudy skies.

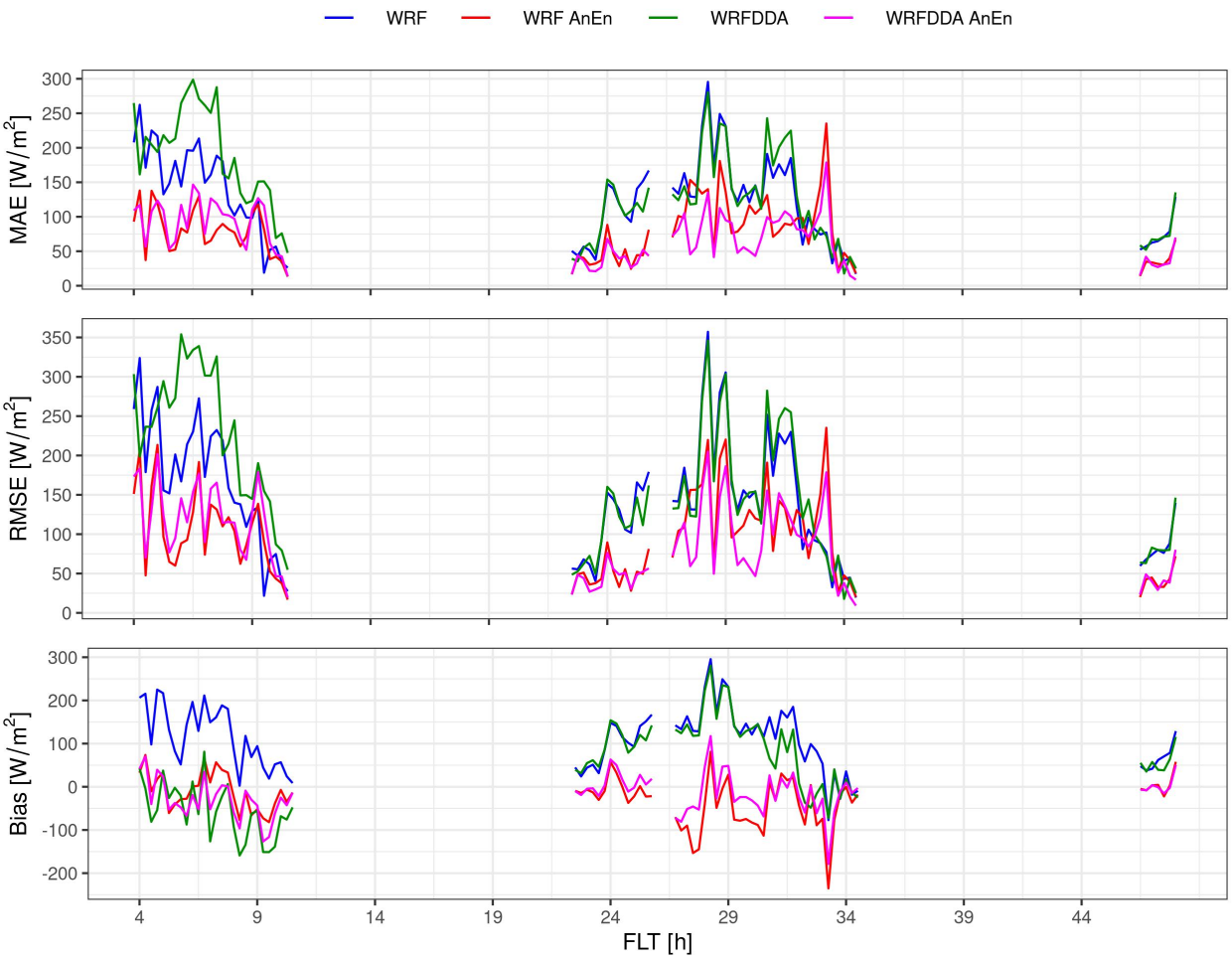


Figure C.1 : MAE, RMSE and Bias between GFS, WRF, WRF-AnEn, WRFDDA and WRFDDA-AnEn for Caleta Loa for cloudy skies.

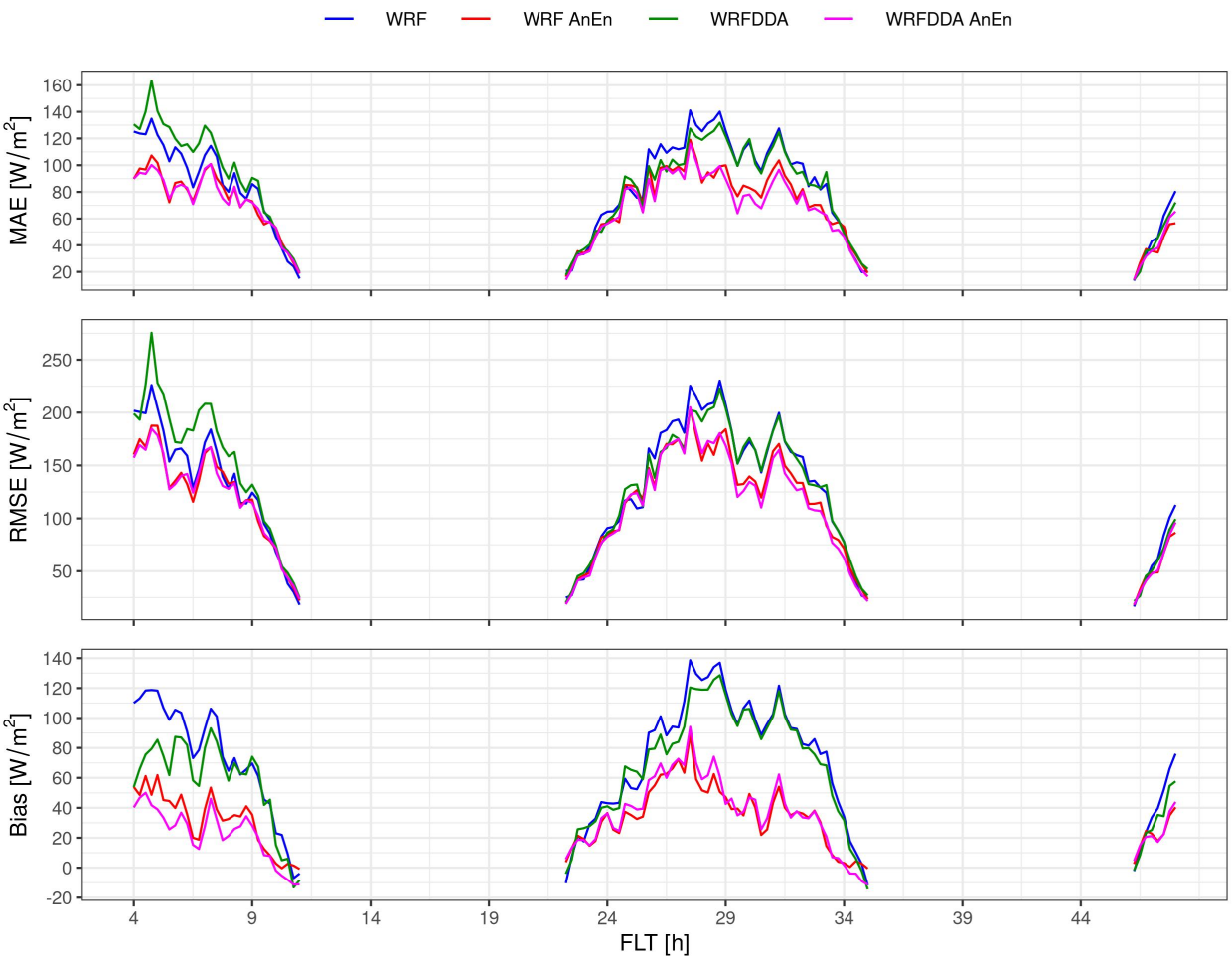


Figure C.2 : MAE, RMSE and Bias between GFS, WRF, WRF-AnEn, WRFDDA and WRFDDA-AnEn for Cerro Calate for cloudy skies.

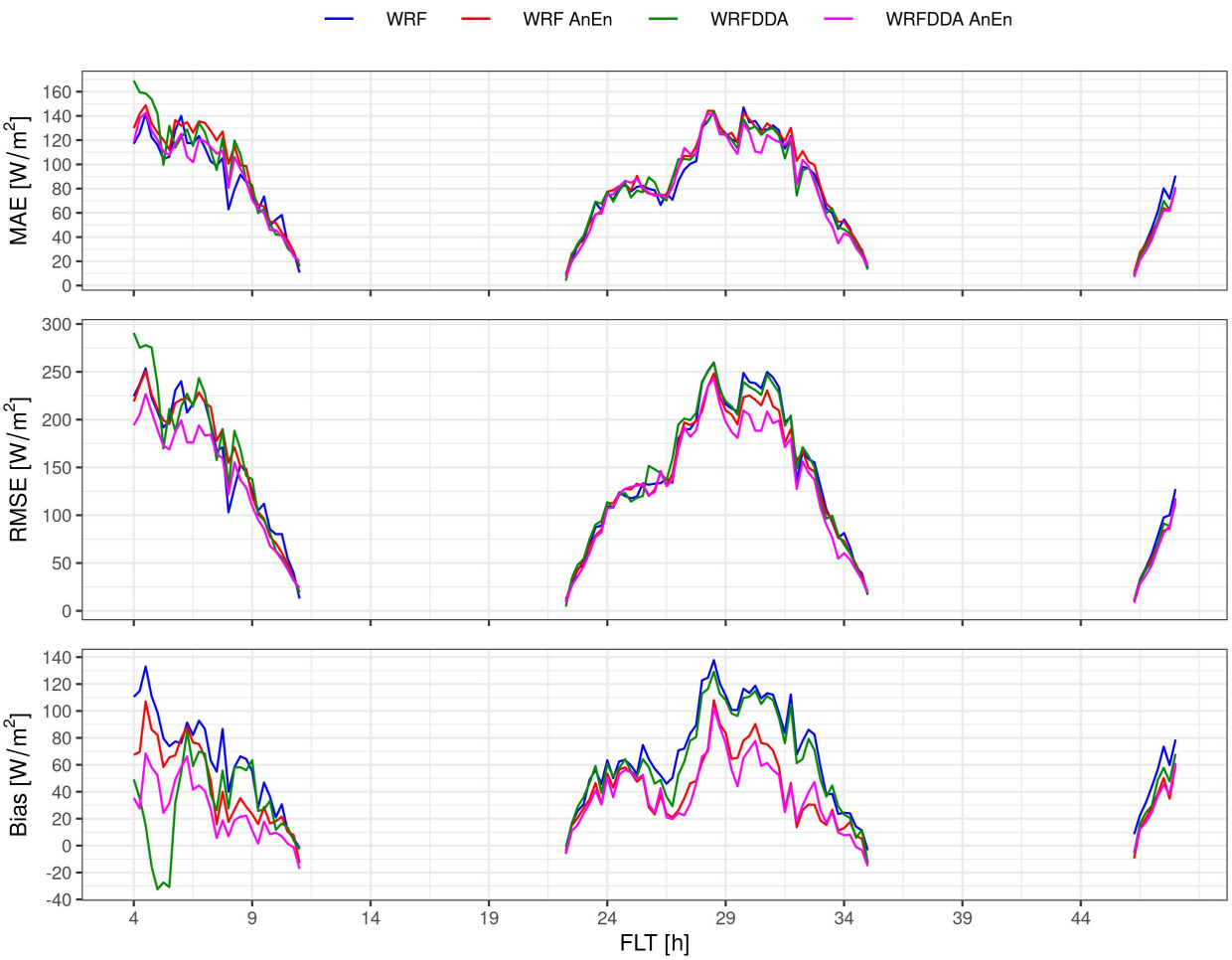


Figure C.3 : MAE, RMSE and Bias between GFS, WRF, WRF-AnEn, WRFDDA and WRFDDA-AnEn for Crucero for cloudy skies.

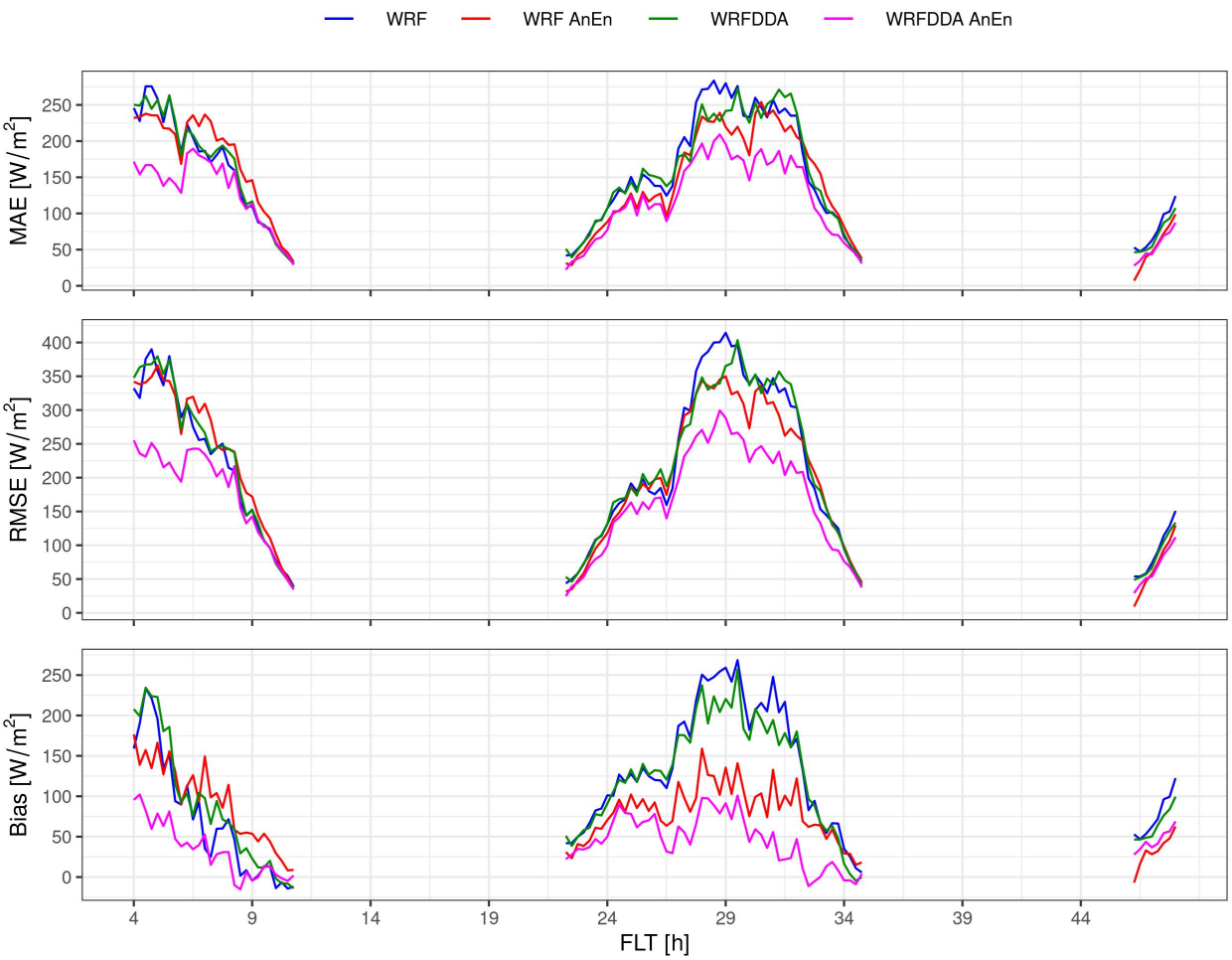


Figure C.4 : MAE, RMSE and Bias between GFS, WRF, WRF-AnEn, WRFDDA and WRFDDA-AnEn for Quebrada Mani for cloudy skies.

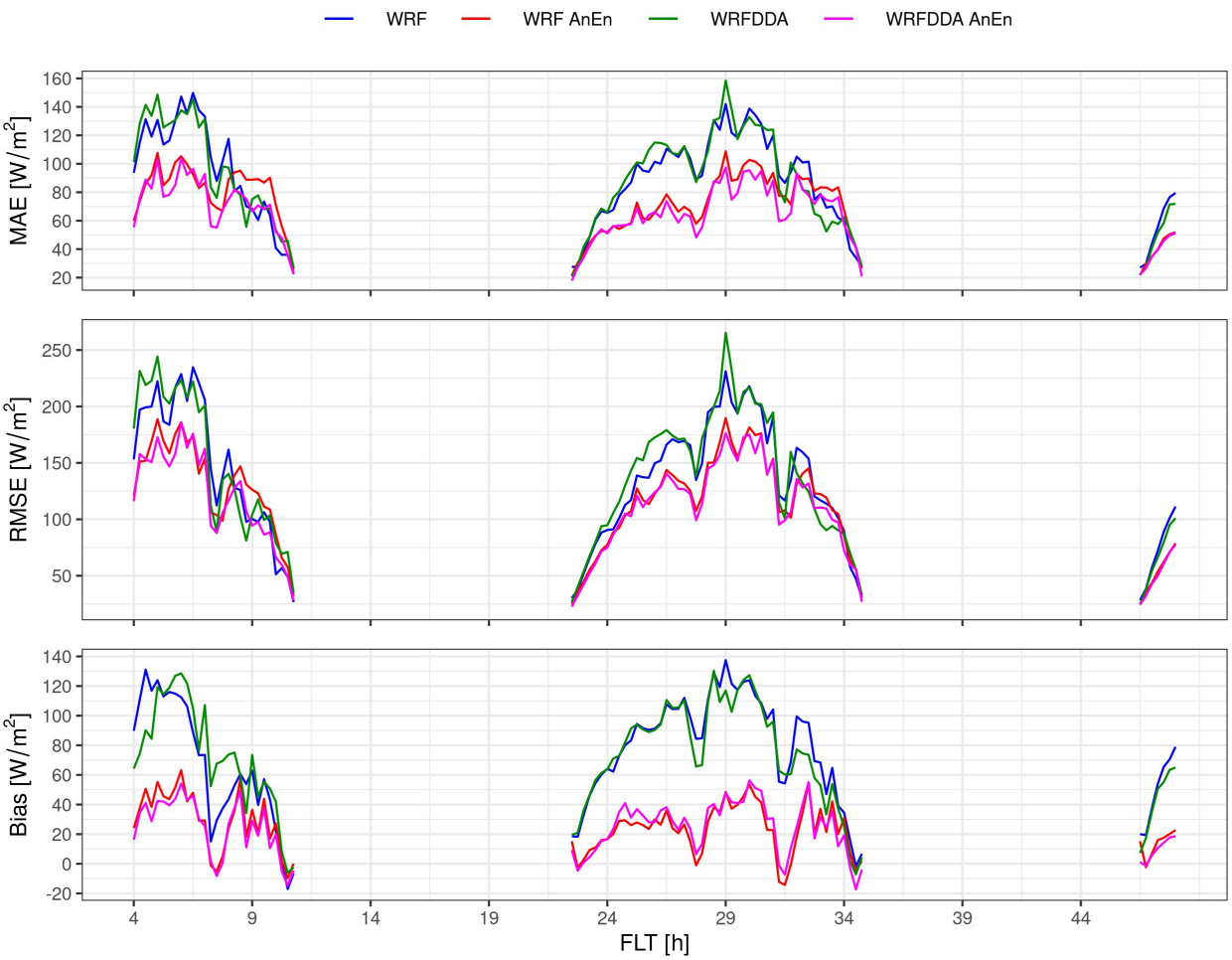


Figure C.5 : MAE, RMSE and Bias between GFS, WRF, WRF-AnEn, WRFDDA and WRFDDA-AnEn for Salar Llamara for cloudy skies.

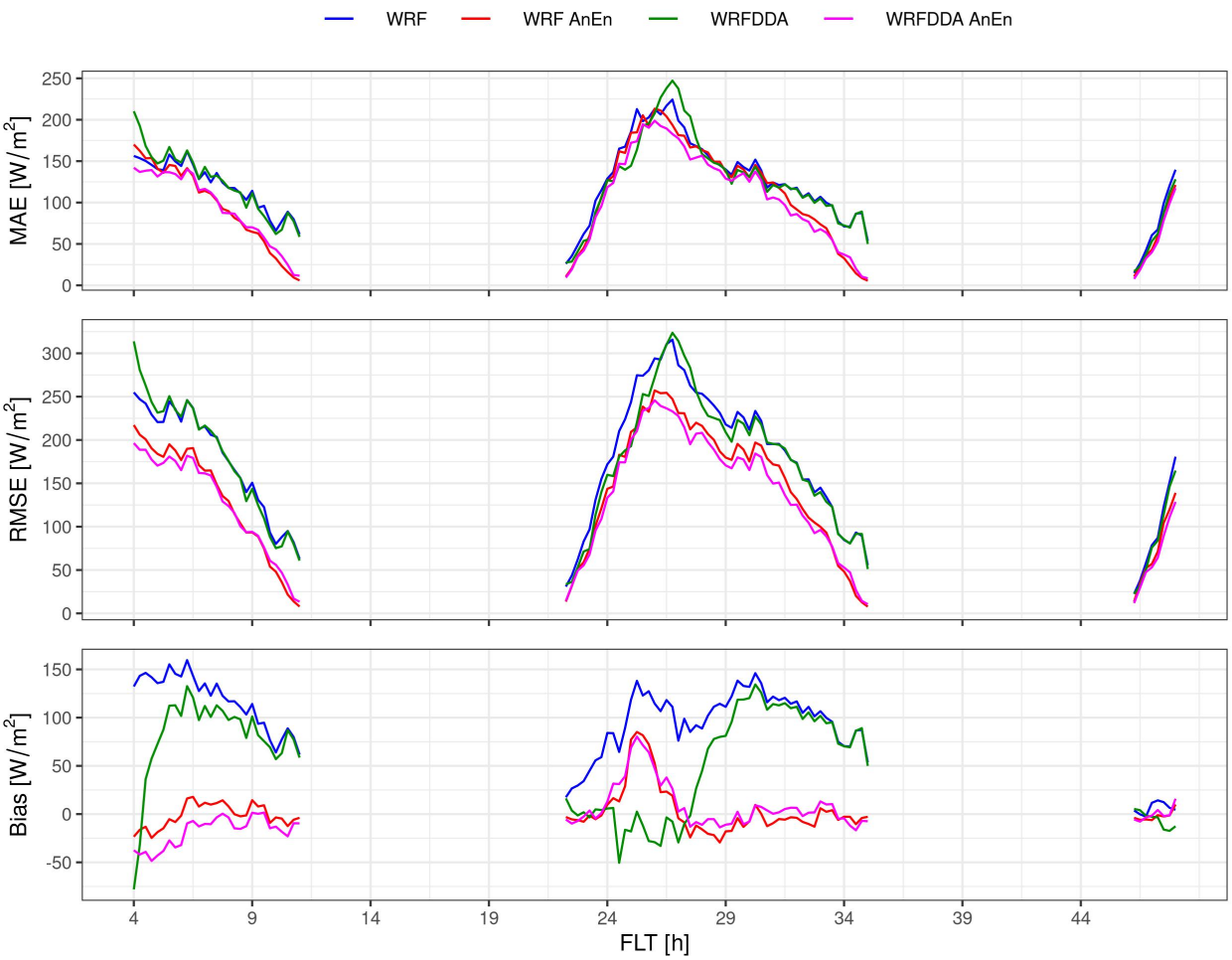


Figure C.6 : MAE, RMSE and Bias between GFS, WRF, WRF-AnEn, WRFDDA and WRFDDA-AnEn for Taltal for cloudy skies.