

Impact of climate change on site characteristics of eight major astronomical observatories using high-resolution global climate projections until 2050

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open access

Haslebacher, C., Demory, M.-E., Demory, B.-O., Sarazin, M. and Vidale, P. L. ORCID: <https://orcid.org/0000-0002-1800-8460> (2022) Impact of climate change on site characteristics of eight major astronomical observatories using high-resolution global climate projections until 2050. *Astronomy & Astrophysics*, 665. A149. ISSN 1432-0746 doi: <https://doi.org/10.1051/0004-6361/202142493> Available at <https://centaur.reading.ac.uk/106833/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1051/0004-6361/202142493>

Publisher: EDP Sciences

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur



CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Impact of climate change on site characteristics of eight major astronomical observatories using high-resolution global climate projections until 2050

Projected increase in temperature and humidity leads to poorer astronomical observing conditions

C. Haslebacher¹, M.-E. Demory^{2,6}, B.-O. Demory³, M. Sarazin⁴, and P. L. Vidale⁵

¹ Center for Space and Habitability and Department of Space Research & Planetary Sciences, University of Bern, 3012 Bern, Switzerland

e-mail: caroline.haslebacher@unibe.ch

² Institute for Atmospheric and Climate Science, ETH Zurich, 8092 Zurich, Switzerland

e-mail: marie-estelle.demory@env.ethz.ch

³ Center for Space and Habitability, University of Bern, 3012, Bern, Switzerland

e-mail: brice.demory@unibe.ch

⁴ European Southern Observatory, Garching, Germany

⁵ NCAS-Climate, Department of Meteorology, University of Reading, Reading, UK

⁶ Wyss Academy for Nature, University of Bern, 3012 Bern, Switzerland

Received 20 October 2021 / Accepted 2 August 2022

ABSTRACT

Context. Sites for next-generation telescopes are chosen decades before the first light of a telescope. Site selection is usually based on recent measurements over a period that is too short to account for long-term changes in observing conditions such as those arising from anthropogenic climate change. For astronomical facilities with a typical lifetime of 30 yr, it is therefore essential to be aware of climate evolution to optimise observing time.

Aims. In this study, we analyse trends in astronomical observing conditions for eight sites. Most sites either already host telescopes that provide in situ measurements of weather parameters or are candidates for hosting next-generation telescopes. For a fine representation of orography, we use the highest resolution global climate model (GCM) ensemble available provided by the high-resolution model intercomparison project and developed as part of the European Union Horizon 2020 PRIMAVERA project.

Methods. We evaluate atmosphere-only and coupled PRIMAVERA GCM historical simulations against in situ measurements and the fifth generation atmospheric reanalysis (ERA5) of the European centre for medium-range weather forecasts for the period 1979–2014. The projections of changes in current site conditions are then analysed for the period 2015–2050 using PRIMAVERA future climate simulations.

Results. Over most sites, we find that PRIMAVERA GCMs show a good agreement in temperature, specific humidity, and precipitable water vapour compared to in situ observations and ERA5. The ability of PRIMAVERA to simulate those variables increases confidence in their projections. For those variables, the model ensemble projects an increasing trend for all sites, which will result in progressively poorer astronomical observing conditions compared to current conditions. On the other hand, no significant trends are projected for relative humidity, cloud cover, or astronomical seeing and PRIMAVERA does not simulate these variables well compared to observations and reanalyses. Therefore, there is little confidence in these projections.

Conclusions. Our results show that climate change will negatively impact the quality of astronomical observations and is likely to increase time lost due to bad site conditions. We stress that it is essential for astronomers to include long-term climate projections in their process for site selection and monitoring. We show that high-resolution GCMs can be used to analyse the effect of climate change on site characteristics of next-generation telescopes.

Key words. site testing – telescopes – Earth – methods: data analysis – astronomical databases: miscellaneous

1. Introduction

Sites for ground-based telescopes are selected considering a wide variety of aspects, such as climatic conditions, proximity to other astronomical facilities, prevalent infrastructure, cultural environment, accessibility to resources, and economic implications (e.g. Schoeck et al. 2009). Since prevailing climate conditions significantly influence the quality of astronomical

observations made by ground-based telescopes (e.g. Falvey & Rojo 2016), site selection studies are designed to find places with the most advantageous climate conditions. The most important atmospheric parameters measured in a site selection process include astronomical seeing, cloud cover, precipitable water vapour (PWV), air temperature, and air humidity (Schoeck et al. 2009; Vernin et al. 2011). The process of building a next-generation telescope, starting with a site selection study

and ending with the first light of a telescope, can take more than a decade. To give two examples, the whole process takes 20 yr in the case of the European extremely large telescope (ELT; ESO 2011) and approximately 27 yr for the thirty meter telescope (TMT; Schoeck et al. 2009)¹. If we approximate the typical lifetime of these telescopes for 30 yr², we implicitly expect the winner of the site selection study to stay the winner for over five decades, a timescale for which climate change might already play a role. Therefore, every site survey brings the risk that site conditions measured during the survey do not reliably represent the long-term climate conditions (Schoeck et al. 2009). Astronomers are already worried about changes due to anthropogenic climate change and see a need to act (Burtscher et al. 2021)³.

The global mean surface temperature has already increased by $1.09_{-0.14}^{+0.11}$ °C from 2011–2020 compared to the period from 1850 to 1900 (IPCC 2021). Already now, anthropogenic climate change has an effect on extreme weather and climate events everywhere on Earth (IPCC 2021). It is a fact that anthropogenic greenhouse gas emissions drive the rapid increase in global mean surface temperature (IPCC 2021).

Studies considering the impact of climate change on site characteristics of ground-based telescopes started to emerge only recently. Hellemeier et al. (2019) analysed the site characteristics' cloud cover, 200-hPa-wind-speed, PWV, vertical wind velocity, and aerosol index of 15 well-established astronomical sites using the FriOWL database (Graham et al. 2008). The FriOWL database combines results from two reanalysis data sets: the ERA-40 reanalysis of the European centre for medium-range weather forecasts (ECMWF) covering the period from 1958–2002 (Källberg et al. 2004), and the national centers for environmental prediction/national center for atmospheric research reanalysis covering the period since 1948 (Kalnay et al. 1996). A reanalysis assimilates archived observational data into a continuous weather simulation and is therefore a valuable tool for studying past climatic changes. To explain trends in long-term site conditions, especially in cloud cover, PWV, and 200-hPa wind speed, Hellemeier et al. (2019) explored the possibility of anthropogenic climate change as a cause. However, the low horizontal resolution of ERA-40 of approximately 278 km limits the reliability of this study.

The first study, to the best of our knowledge, to look explicitly at the influence of climate change on ground-based telescopes was done by Cantalloube et al. (2020). Cantalloube et al. (2020) conducted a study for Cerro Paranal on the impact of climate change on astronomical observations using reanalysis data from the fifth generation atmospheric reanalysis (ERA5; Hersbach et al. 2020) and the first atmospheric reanalysis of the 20th century (ERA-20C; European Centre for Medium-Range Weather Forecasts 2014) of ECMWF, and one global climate model (GCM) projection from the coupled model intercomparison project phase 6 (CMIP6), namely the Beijing climate centre BCC-CSM2-MR CMIP6 GCM (Wu et al. 2019). A global climate model (also denoted as a general circulation model) simulates physical processes in subsystems of Earth such as the atmosphere or the ocean on a three-dimensional grid. They found an increase in near-surface air temperature on Cerro Paranal of 1.5 °C between 1980 and 2020 assimilated by ERA5 and an increase of 4 °C by 2100 projected by the CMIP6 GCM under the shared socioeconomic pathway SSP5–8.5 scenario

(Kriegler et al. 2017; Riahi et al. 2017). The SSP5–8.5 scenario combines the worst-case scenario of five shared socioeconomic pathways (SSPs), which describe different possible developments of societal, demographic, and economic changes, with the worst-case representative concentration pathway (RCP) 8.5 of four RCPs defining the forcing in 2100 from 2.6 to 8.5 W m⁻². Cantalloube et al. (2020) state that such an increase in temperature brings the active temperature control system of the four unit telescopes of the very large telescope (VLT) on Cerro Paranal to its limits more often, which would enhance dome seeing (Tallis et al. 2020; atmospheric turbulence inside the dome due to temperature differences inside and outside the dome). Cantalloube et al. (2020) also analysed seeing measurements from 1986 to 2020 and found a positive trend in seeing due to an increased surface layer turbulence, either caused by climate change or constructional modifications and alterations of measurement devices and measurement locations. Higher temperature gradients, potentially arising from surface temperature increases, do indeed enhance convection and therefore turbulence close to the ground. Furthermore, Cantalloube et al. (2020) found indications that the area around Cerro Paranal may become drier in the future, although this result is uncertain due to the use of one GCM only and the coarse horizontal resolution of 110 km used by the BCC-CSM2-MR GCM (Wu et al. 2019).

Another study using ERA5 was done by Osborn & Sarazin (2018). Osborn & Sarazin (2018) find that ERA5 model data on 137 vertical pressure levels show strong agreement with the high-altitude resolution optical turbulence profiler, and their model is able to predict strong turbulence events, which is crucial because strong turbulence dominates instrument performance limitations of adaptive optics systems. Adaptive optics systems are used to correct the light wavefront distortions induced by the atmospheric turbulence at optical and infrared wavelengths in near-real time, which is an integral part of future 30-m-class telescopes.

Considering the typical timeline of next-generation telescopes from site selection to the end of lifetime, site selection processes as well as current telescopes might need to address the challenges of changing climate conditions. As suggested by Graham et al. (2004) and already shown by Cantalloube et al. (2020), climate models that project scenarios of anthropogenic radiative forcing into the future may prove useful. To address particular topographic textures as they often appear at astronomical sites, GCMs with the highest possible horizontal resolution are advantageous (Moreno-Chamarro et al. 2022). Even if the use of regional climate models (RCMs) would allow for even higher resolution, RCMs do not provide global coverage or global consistency. With RCM ensembles, such as the coordinated regional climate downscaling experiment (CORDEX; Giorgi et al. 2009), only a few sites would be covered by the computational domains. Furthermore, it is indispensable to validate the agreement between model output and measurements for astronomical sites, especially since well-established astronomical sites are often located on highly elevated mountains that are not particularly well represented in climate models.

In this study, we aim to assess the reliability of using global climate models to project future changes in site characteristics for astronomical observations. To do so, we make use of six GCMs from the PRIMAVERA project (Roberts et al. 2018), which is the European contribution to the high-resolution model intercomparison project (HighResMIP; Haarsma et al. 2016). We analyse the PRIMAVERA GCM simulations against in situ measurements and the ERA5 reanalysis product for eight well-established sites of major astronomical observatories around

¹ <https://www.tmt.org/page/timeline>, accessed on 2019 Apr. 17.

² <https://elt.eso.org/about/faq/>, accessed on 2021 Apr. 23.

³ Also see the open letter of Astronomers for Planet Earth: <https://astronomersforplanet.earth/open-letter/>, accessed on 2022 Jun. 14.

the world. Reanalysis products bridge the gap between measurements and global climate models because they provide homogeneous information across the globe. Here, ERA5 is chosen because it is the latest reanalysis product of ECMWF; it provides information over a long period and it has a spatial resolution similar to the PRIMAVERA GCMs. The evaluation of ERA5 against in situ measurements is therefore a necessary step for a careful evaluation of PRIMAVERA GCMs.

The investigated sites are Mauna Kea on the island of Hawaii (USA), San Pedro Mártir in Baja California in Mexico, the three Chilean sites Cerro Paranal, Cerro Tololo, and La Silla, La Palma on the Canary Islands (Spain), Sutherland in South Africa, and Siding Spring in Australia. We investigate air humidity, air temperature, PWV, cloudiness, and astronomical seeing parameters. After evaluating how well ERA5 and PRIMAVERA represent the present climate conditions of the eight sites, we use PRIMAVERA GCMs to project future conditions at these sites up to the year 2050. We focus on trends of mean changes and do not consider changes in climate variability or changes in the extremes, which are also to be expected (IPCC 2021).

In Sect. 2, we introduce the atmospheric variables we analysed. In Sect. 3, we introduce each site in detail. In Sect. 4, we describe in situ measurements, the ERA5 reanalyses, and the PRIMAVERA GCMs. In Sect. 5, we present the general workflow for evaluating ERA5 and PRIMAVERA and the method employed for the trend analysis. In Sect. 6, we discuss the results of the comparison between in situ measurements, ERA5, and PRIMAVERA GCMs and in Sect. 7 the results of future trends projected by PRIMAVERA. The last Sect. 8 concludes the study.

2. Atmospheric variables

2.1. Air temperature

Air temperature influences the seeing (Cavazzani et al. 2012), specific humidity and with it, PWV and clouds (McIlveen 1992). Air temperature is the easiest atmospheric variable to measure (McIlveen 1992). Archived temperature data spanning a long time range are often provided by weather stations and long-time series are available at observatories, often measured by a weather tower 2 m to 10 m above ground and sometimes enclosed in an instrument shelter. It is also the best modelled variable so far, and the most trustworthy for temperature projections (IPCC 2021). GCMs experimental data submitted to the CMIP6 intercomparison enabled the Intergovernmental Panel on Climate Change (IPCC) to state that anthropogenic carbon emissions driving climate change will cause the global mean surface temperature to rise at least until 2050 under all emission scenarios (IPCC 2021). Under the highest greenhouse gas (GHG) emission scenario (SSP5–8.5), the global surface temperature at the end of the century (2081–2100) is projected to be 3.3 °C to 5.7 °C higher compared to the average between 1850 and 1900.

2.2. Air humidity

For astronomical observatories, the relative humidity is an important parameter because whenever the relative humidity is too high, a telescope has to close its dome to protect sensitive parts such as the optical system or electrical wires from moisture, which could lead to oxidation or leaching (Bradley et al. 2006). Therefore, the relative humidity is often measured by observatories. As for the temperature, this is most often done with a weather tower that measures 2 m above the ground (e.g. Schoeck et al. 2009).

The relative humidity RH depicts the ratio of the specific humidity q to the saturation specific humidity q_s in air for a specified temperature and pressure (Peixoto et al. 1992)

$$\text{RH} = 100\% \frac{q}{q_s} = 100\% \frac{e}{e_s}, \quad (1)$$

which is the same as the ratio of the water vapour pressure e to the saturated vapour pressure e_s . The higher the values of relative humidity, the more saturated the air is, and saturation is defined as a relative humidity value of 100%.

The specific humidity q is defined as the ratio of the water vapour mass M_v of a specified volume of air to the total mass M_{tot} of the same volume of air (Peixoto et al. 1992)

$$q = \frac{M_v}{M_{\text{tot}}}. \quad (2)$$

Since at least 1970, there has been an increasing trend of near-surface specific humidity over land and ocean that can be attributed with medium confidence to human activities (IPCC 2021). Since the temperature over land increases faster than the temperature above the ocean, near-surface relative humidity is expected to decrease over land, which has been observed since the year 2000 (IPCC 2021). The result is regional drying due to increased evapotranspiration.

2.3. Precipitable water vapour

The PWV is the amount of water vapour in the atmosphere integrated over a unit column from the surface to the top of the atmosphere (Peixoto et al. 1992). For the PWV, we integrate until the tropopause, because most weather events happen in the troposphere and most of the water is contained inside the troposphere. Put another way, the PWV is the mass of liquid water if all the water vapour in a unit column of the atmosphere were condensed (Peixoto et al. 1992). It is given by

$$W(\theta, \phi, t) = \int_0^{p_0} q \frac{dp}{g}, \quad (3)$$

where q is the specific humidity, g is the gravitational acceleration, p is the pressure and p_0 is the surface pressure. The PWV depends on the longitude θ , the latitude ϕ and the time t at which it is observed.

The PWV pertains to the important site characteristics of telescopes observing in the visible and infrared because water absorbs light over a wide range of wavelengths with an absorption maximum in the infrared. Not only does the light from an observed astronomical object get absorbed, but also thermal emissions of Earth itself can get backscattered into the telescope. Both contributions fluctuate in time due to a turbulent atmosphere. One example of an astronomical observation that depends on low PWV values is direct imaging, where the mid-infrared band is used preferentially because the contrast in flux between planet and star is optimal in the mid-infrared (Pathak et al. 2021; Bessell et al. 2005). The temporal variation of PWV is in general relatively large due to evaporation and precipitation (e.g. Kalinnikov & Khutorova 2017; Otarola et al. 2019). For example at Paranal observatory on the timescale of an hour, the PWV fluctuates between 1% and 5.5% around the mean when measured in 120 s time intervals (Otarola et al. 2019).

The PWV content can be measured, for example with radiosondes, or it can be estimated, for example from measurements with an infrared radiometer that observes at 20 μm

wavelength (e.g. [Chapman et al. 2004](#)), with opacity measurements at 225 GHz (e.g. [Dempsey et al. 2013](#)), or with 183 GHz radiometers (e.g. [Kerber et al. 2012](#); [Wiedner et al. 2001](#)). These measurements can be used by observatories with telescopes operating in the (sub)millimetre range to subtract spectra due to water vapour ([Chapman et al. 2004](#)).

PWV is expected to rise globally with increasing global mean surface temperature in the future ([Held & Soden 2006](#); [Trenberth et al. 2005](#)). This is backed up by an already observed increasing trend in total column water ([IPCC 2021](#)).

2.4. Cloud cover

Astronomical observations in the visible and infrared can only be carried out when the sky is clear, with as little cloudiness as possible. According to the European Southern Observatory (ESO; [Kerber et al. 2014](#)), a photometric night is defined by no visible clouds and transparency variations of the atmosphere below 2%, a clear sky is defined as a sky with less than 10% cloud cover and transparency variations below 10% and thin cirrus is the term for transparency variations above 10%. Methods were developed to assess the cloud coverage with infrared all-sky cameras that are sensitive to very thin clouds but not deceived by variations of PWV ([Kerber et al. 2014](#)).

Even though evaporation and precipitation must be in equilibrium globally, precipitation exceeds evaporation for the equatorial belt, where many clouds form, due to strong convection, and evaporation exceeds precipitation in the subtropical latitudes due to weak upward motion, which leads to a minimum in cloudiness ([Peixoto et al. 1992](#)). A maximum in cloud cover is found at latitudes near 50–60 °C ([Peixoto et al. 1992](#)).

In GCMs, the dynamics of convection and cloud formation usually cannot be simulated explicitly because their horizontal resolution is too low. A horizontal grid spacing of a few kilometres would be necessary, whereas GCMs contributing to CMIP6 have a median horizontal atmospheric resolution of ca. 140 km ([IPCC 2021](#)). GCMs therefore have to rely on the use of semi-empirical parameterisation schemes that generate large uncertainties in the simulation of such processes associated with cloud cover ([Bony et al. 2015](#), and references therein).

Cloud interactions with the global climate will likely further amplify anthropogenic warming ([IPCC 2021](#)). However, clouds are the most uncertain parameter in the assessment of global climate feedback, which gets more complicated due to a lack in the understanding of cloud forming processes and also because clouds need to be parameterised in GCMs due to the too low resolution of GCMs ([IPCC 2021](#)).

2.5. Seeing

The distortion of light wavefronts due to atmospheric turbulence is a well-studied issue in optical astronomy ([Roddier 1981](#)) and is caused by a variation in refractive index of the atmosphere along the line of sight due to temperature and density gradients. The Gladstone-Dale relation,

$$\frac{n-1}{\rho} = \text{const} \quad (4)$$

relates the refractive index n to the density ρ of a medium ([Gladstone & Dale 1863](#)). In the framework of the Kolmogorov model ([Tatarski et al. 1961](#)), the local power spectrum of a well developed turbulence can be characterised by a single index of refraction structure coefficient C_n^2 . The effect of the variations

of C_n^2 along the line of sight is discussed in detail in Sect. 5.2.3. The alterations of a plane wavefront travelling through this turbulence, integrated along the line of sight, can be described by their coherence radius r_0 ([Fried 1966](#)), for a wavelength λ at zenith:

$$r_0 = \left(0.423 \cdot \int C_n^2(z) dz \cdot \left(\frac{2\pi}{\lambda} \right)^2 \right)^{-3/5}. \quad (5)$$

This ‘‘Fried parameter’’ r_0 sets the sharpness of the image of a point source through a telescope limited by the atmosphere, usually described by its full width at half maximum (FWHM) in radian:

$$\epsilon = 0.976 \cdot \frac{\lambda}{r_0}. \quad (6)$$

Astronomers characterise the atmospheric observing conditions by the so-called seeing, defined as the FWHM at a wavelength of $\lambda = 500$ nm expressed in arcseconds.

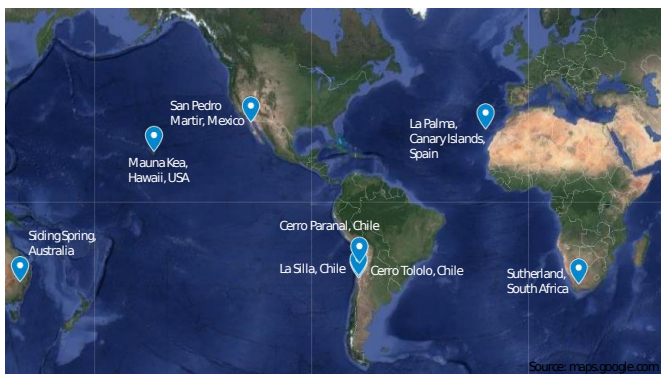
The equivalence between seeing and image quality is only valid within the Kolmogorov turbulence model, which does not set upper limits to the scale of the turbulence in the inertial range. With the increase of the size of modern ground based telescopes, departures from this model were observed in the infrared on images sharper than predicted by the seeing. Obviously due to the deficit of large tilts within the wavefront, the observations could be explained by the introduction of an outer scale L_0 of the turbulence. The general Eq. (6) is however suitable for studying climate impact on optical turbulence precisely because it is independent of the outer scale, which would depend on the telescope diameter. Deformable mirrors that can be adapted in real time to wave distortions due to atmospheric turbulence (adaptive optics), with which it is possible to get close to the diffraction limit of the telescope, are nowadays part of a standard instrumentation for ground-based telescopes ([Davies & Kasper 2012](#)). The performance of adaptive optics correction depends on the instrument attached to the telescope (e.g. spectrograph, coronagraph) and on spatiotemporal properties of the wavefront.

Since the seeing quality can be, depending on the orography, better on a mountain due to a higher chance of unperturbed laminar airflow on the elevated level of the mountain, many observatories are, despite other advantages, situated on mountains. Because impactors such as cars, trees or buildings have a significant effect on the seeing (e.g. [Teare et al. 2000](#)) but are not represented in GCMs, we do not expect GCMs to represent absolute seeing values realistically. However, we verify in this study whether GCMs are able to simulate the free atmosphere seeing. The free atmosphere seeing excludes the contribution from the ground and planetary boundary layer and starts at approximately 1–2 km above observatory level ([Osborn & Sarazin 2018](#)). The free atmosphere seeing therefore includes the tropopause level at 200 hPa, where, dependent on the latitude, a jet stream might prevail. At around 30° latitude is the subtropical jet stream, so all sites in this study are influenced by the subtropical jet stream. For example in La Silla and Mauna Kea, the jet stream at 8–12 km altitude was found to be the dominant contribution to the free atmosphere seeing ([Vernin 1986](#)). The lower the wind speed at 200 hPa, the better the seeing ([Vernin 1986](#)), and seeing below 1 arcsecond requires wind speeds below 20 m s⁻¹ at the tropopause. Trends in the strength of the subtropical jet stream could influence the free atmosphere seeing. In this context, the winter subtropical jet stream has accelerated and shifted polewards over the period 1979–2008 and extended regions in comparison to the period 1958–2008 ([Pena-Ortiz et al. 2013](#)).

Table 1. Overview of sites evaluated in this study including longitude (Lon), latitude (Lat), elevation and pressure.

Observatory site	Lon/Lat (°)	Elevation (m)	Pressure (hPa)
Mauna Kea (USA)	−155.47/19.82	4200	616.28 ± 1.70
Cerro Paranal (Chile)	−70.40/−24.63	2635	743.66 ± 0.50
La Silla (Chile)	−70.73/−29.25	2400	771.06 ± 2.00
Cerro Tololo (Chile)	−70.80/−30.17	2200	781.28 ± 2.66
La Palma (Spain)	−17.89/28.76	2370	771.21 ± 3.10
Siding Spring (Australia)	149.07/−31.28	1165	891.56 ± 4.55
Sutherland (South Africa)	20.81/−32.38	1798	826.49 ± 1.50
San Pedro Mártir (Mexico)	−115.46/31.04	2800	732.70 ± 3.20

Notes. Pressure: monthly mean in situ value ± standard deviation.

**Fig. 1.** Map of sites evaluated in this study (generated with Google maps; maps.google.com, accessed on 2020 Sep 06).

Ensemble means of future simulations project a strengthening of the southern mid-latitude jet stream under SSP5–8.5, while trends for the northern jet stream wind speeds are uncertain due to large natural internal variability, opposing changes in temperature gradients and identified simulation weaknesses (IPCC 2021).

Astronomical seeing is mainly measured with a Differential Image Motion Monitor (DIMM; Sarazin & Roddier 1990) or with a Multi-Aperture Scintillation Sensor (MASS; Kornilov et al. 2003). The DIMM probes the entire atmosphere, starting at the level of the monitor. The MASS instrument returns an atmospheric six-layer turbulence profile excluding the lowest 0.5 km.

3. Description of major astronomical sites

We selected eight sites where one or more telescopes are already in operation (Table 1 and Fig. 1) and in situ measurements archived by the observatories are available. Furthermore, the selected sites have the potential to host next-generation telescopes.

An overview of the climatology of each selected site is presented hereafter. In general, all sites have an above average percentage of nights with clear skies. Also, observatories are keen to minimise light pollution to be able to observe dark skies, for example by choosing isolated sites.

3.1. Mauna Kea

The dormant volcano Mauna Kea in the north-east of the Hawaiian Big Island and north of the world’s largest volcano

Mauna Loa, is the highest elevated site, at 4200 m, that we investigate in this study. Between 1982–2011, measurements at the University of Hawaii 2.24-m (UH2.2) telescope with a weather station 2 m above top of the 23 m high dome gave an average annual temperature of 1.3 °C and an average annual wind speed of 5.8 m s^{−1}, with very low variability (standard deviation of 1.3 m s^{−1}; Da Silva et al. 2012). The climate on top of Mauna Kea is very dry and often free of clouds, because the summit lies above the trade wind inversion at about 2000 m, which is a temperature inversion that traps moist air (Ward & Galewsky 2014). However, there are seasonal and inter-annual variations in the quality of the observations, for instance due to the El Niño southern oscillation (ENSO; Lyman et al. 2020), a quasi-periodic cycle that causes warming or cooling of the sea surface temperature of the eastern Pacific Ocean that impacts the global general circulation. Furthermore, Giambelluca & Schroeder (1986) emphasised that the microclimatic conditions on Hawaii range from dry leeward areas to humid windward flanks.

The site hosts 13 telescopes, nine of which are observing in the optical and infrared, namely the 0.9-m University of Hawaii (UH) Hilo educational telescope, the UH 2.2-m telescope, the 3.0-m NASA infrared telescope facility (IRTF), the 3.6-m Canada-France-Hawaii telescope (CFHT), the 3.8-m United Kingdom infra-red telescope (UKIRT), the two 10-m W. M. Keck observatories Keck I and Keck II, the 8.3-m Subaru telescope and the 8.1-m Gemini northern telescope, three in the submillimetre range, namely the 10.4-m Caltech submillimeter observatory (CSO), the 15-m *James Clerk Maxwell* telescope (JCMT) and the 8 × 6-m submillimeter array, and lastly the 25-m very long baseline array (VLBA) in radio⁴.

3.2. Chilean sites

A VLT site selection study conducted in 1990 (Grenon 1990) chose an interdisciplinary approach to investigate the climate in Northern Chile using biogeography. In Chile, the further east, the more precipitation and clouds. This is why coastal sites in Chile are preferred for astronomical observatories. Grenon (1990) found some evidence that, since 1940, Northern Chile got drier due to climate change. Cerro Tololo, La Silla and Cerro Pachon lie in the transition zone between the Atacama desert and the Chilean Matorral. This zone is characterised by a Mediterranean-type climate with sparse events of southern winter rains and a dry summer and autumn. The climate at Cerro Tololo is more humid than at Cerro Paranal, especially in southern winter. It is also in southern winter that most nights are not photometric (Grenon 1990).

The subtropical jet stream is the dominant driver of the movement of the atmosphere for these Chilean sites, since the jet stream oscillates within a latitude range that includes Cerro Paranal, La Silla and Cerro Tololo. Its temporal variability is quite high: within a few day’s time, the wind velocity can become up to four times weaker or stronger (Sarazin 2001). In southern winter (July–September), jet stream wind speeds are higher than in southern summer (January–August).

3.2.1. Cerro Paranal

Cerro Paranal lies in one of the driest regions of the world, in the Atacama desert 12 km east of the South Pacific Ocean coast, with a typical relative humidity between 5 and 20% according to

⁴ http://legacy.ifa.hawaii.edu/mko-old/telescope_table.shtml, accessed on 2022 Jan. 25.

ESO⁵. On average, the sky is photometric in 78% of the nights and PWV is less than 1 mm in 8.2% of the nights⁶.

The optical very large telescope (VLT) is run by ESO on Cerro Paranal since 1998 with four 8.2-m primary mirror unit telescopes and four auxiliary 1.8-m primary mirror movable telescopes, each of which can be combined to form the VLT interferometer. Only 20 km east of Cerro Paranal lies Cerro Armazones, the mountain where the ELT is under construction, which will be the next biggest optical telescope with a primary mirror of 39-m diameter whose first light is expected in 2026.

3.2.2. La Silla

La Silla observatory is located circa 500 km south of ESO Paranal and is operated by ESO since 1960. It shows a similar climatology than Paranal that is slightly wetter since La Silla is on the edge of the dry area of Chile (Sarazin 1988). Also, the seasonal cycle of humidity in La Silla shows a larger amplitude than for Paranal (Sarazin & VLT Site Selection Working Group 1990). Short-term bad weather associated with low pressure events coming from Antarctica occur in southern winter and hit La Silla harder than Paranal, for example in the form of snow storms. At La Silla observatory, two 3.6-m telescopes, namely the new technology telescope (NTT) and the ESO 3.6-m telescope, are operated by ESO (ESO 2015). Smaller telescopes hosted by La Silla observatory include the 2.2-m Max-Planck telescope, the 1.2-m Swiss telescope, the 1-m Schmidt telescope, the 1.5-m Euler telescope, the 0.6-m REM telescope, the 0.6-m TRAPPIST telescope and the 1.5-m Danish telescope.

3.2.3. Cerro Tololo

Cerro Tololo is another site providing very good astronomical observing conditions in the region of Coquimbo, south to the Atacama desert. The Cerro Tololo inter-American observatory (CTIO) operates the Víctor M. Blanco 4-m telescope and four 1-m-class telescopes⁷. Additionally, the small and moderate aperture research telescope system (SMARTS) consortium on Cerro Tololo operates the SMARTS 1.5-m, 1.3-m, 1.0-m and 0.9-m telescope. Approximately 10 km (beeline) south-east to Cerro Tololo lies Cerro Pachon, where the 4.1-m southern astrophysical research (SOAR) and the 8.1-m Gemini south telescope are located and where the Vera C. Rubin observatory is under construction with the planned 8.4-m Simonyi survey telescope.

3.3. La Palma

The Spanish island ‘Isla de la Palma’ belongs to the Canary Islands. The total area of the island is small, ca. 700 km². On the highest point of the island, at 2370 m, lies the Roque de los Muchachos Observatory (ORM). It houses 16 telescopes, among them the 10.4-m optical and near infrared gran telescopio Canarias, the 4.2-m optical and near infrared William Herschel telescope (WHT), the 3.58-metre telescopio nazionale Galileo (TNG) and the 2.6-m nordic optical telescope (NOT)⁸. Roque de los Muchachos lies well above the tree line and the trade wind

inversion based on ca. 1000 m, which prevents clouds from forming at ORM (McInnes & Walker 1974). The convex profile of the Caldera de Taburiente directs air flow around the peak (McInnes & Walker 1974).

3.4. Siding Spring

Siding Spring observatory (SSO) is located at the brink of a National Park near Coonabarabran in south-eastern Australia. It is the site with the lowest elevation (1165 m) out of the selected sites. The climate at Siding Spring is drier and nights are colder compared to the rest of Australia⁹. One consequence of anthropogenic climate change that also threatens observatories is an increase in the frequency, intensity and duration of wildfires, which is expected with high confidence (IPCC 2021). One example is the Wambelong bush fire in 2013 that destroyed facilities at the Siding Spring Observatory (SSO), but due to measures that were taken beforehand, it caused little damage to the telescopes (AAO 2013). The site hosts 11 telescopes, among which is the 3.9-m optical Anglo-Australian telescope (AAT), the 2.3-m Australian national university (ANU) telescope, and the 1.2-m UK Schmidt telescope (UKST).

3.5. Sutherland

Near the town of Sutherland in the semi-desert of the Great Karoo, 12 optical and infrared telescopes are operated, among which is the 10-m segmented southern African large telescope (SALT). Temperatures are coldest in the town of Sutherland compared to the rest of South Africa. There is little precipitation due to a mountain chain surrounding the Great Karoo that works as a humidity trap¹⁰. Almost the whole area of the Great Karoo has been classified as an astronomy advantage area (AAA) by the South African minister of science and technology, an area where regulations can be introduced to control communication, transport and other technology to support operation of specialised telescopes (Walker et al. 2019).

3.6. San Pedro Mártir

The National Astronomical Observatory (OAN) is located on top of the Sierra de San Pedro Mártir (SPM) in the north of the Mexican Baja California peninsula. It hosts a 2.1-m optical and infrared telescope and seven other 1-m-class telescopes¹¹. Compared to the three selected Chilean sites, the amount of precipitation is higher for the observatory located in SPM. Precipitation events are strong but rare (Schoeck et al. 2009). The site offers low relative humidity and good seeing conditions (López et al. 2003). Between 1984 and 2002, 63% of the nights were photometric (Cruz-Gonzales et al. 2004).

4. In situ, reanalysis, and model data

4.1. In situ measurements

Table 2 provides details on the in situ measurements archived by the observatories. We filtered the in situ data before averaging with the following upper and lower cut-offs. The valid relative

⁵ <http://www.eso.org/sci/facilities/paranal/astroclimate/site.html>, accessed on 2020 Aug. 27.

⁶ See footnote 5.

⁷ <https://noirlab.edu/science/programs/ctio/telescopes>, accessed on 2022 Jan. 25.

⁸ <https://www.iac.es/en/observatorios-de-canarias/roque-de-los-muchachos-observatory/telescopes-and-experiments>, accessed on 2022 Jan. 25.

⁹ <http://www.bonzle.com/c/a?a=p&p=315007&d=w&s=siding%20springs%20observatory&cmd=sp&c=1&x=149%2E065885&y=%2D31%2E274767&w=40000&mpsec=0> accessed on 2020 Aug. 27.

¹⁰ http://www.thegreatkaroo.com/index.php?page=climate_and_vegetation, accessed on 2020 Aug. 27.

¹¹ <https://www.astrossp.unam.mx/en/users/telescopes/2-1m-telescope>, accessed on 2022 Jan. 25.

Table 2. Overview of in situ data.

Site	Data	Observatory or telescope	Time period	Frequency
Mauna Kea	Meteo ^(a)	CFHT	2000–2019	1–10 min (24 h)
	Seeing ^(b)	MKWC	2010–2019	2 min (night-time hours)
	PWV ^(c)	JCMT	2009–2019	2 s (24 h)
Paranal	Weather loss	Gemini South	2014–2017	1 day (night-time)
	Meteo, Seeing	VLT	2000–2019	1 min (24 h)
	PWV	VLT	2016–2019	2 min (24 h)
La Silla	photometric nights ^(d)	VLT	1984–2017	1 month
	Meteo	La Silla observatory	2000–09/2008, 2011–2019	10 min (24 h)
	Seeing	La Silla observatory	2000–2019	1 min (24 h)
Cerro Tololo	PWV ^(e)	La Silla observatory	2000–2007	3 h (24 h)
	photometric nights ^(d)	VLT	1984–2017	1 month
	Meteo	CTIO	2002–2018	5 min (24 hours)
Cerro Pachon	Seeing	CTIO	2004–2019	1 min (night-time)
	Weather loss	Gemini North	2013–2017	1 day (night-time)
La Palma	Meteo ^(f)	NOT	2004–2019	5 min (24 h)
	Seeing ^(g)	IAC (ING)	2004–2019	3 min (night-time)
Siding Spring	Weather loss ^(h)	TNG	2012–2016	1 day (night-time)
	Meteo	AAT	2004–2019	10 min (24 h)
	Seeing ⁽ⁱ⁾	AAT	1993–2019	by semester (night-time)
Sutherland	Weather loss	AAT	1993–2018	1 day (night-time)
	Meteo	SALT	2013–2019	1 min (24 h)
SPM	Meteo ^(j)	OAN	2006–2019	15 min (24 h)
	Seeing ^(k)	TMT site testing campaign	10/2004–08/2008	1 min (night-time)
	Pressure ^(k)	TMT site testing campaign	10/2004–08/2008	2 min (24 h)

Notes. It is indicated from which observatory the data come from, the time period and the frequency of the measurements. If no web page is indicated in the following, we retrieved the in situ data via private communication with the indicated observatory. All web pages were accessed in the time period September 2019 to September 2020. ^(a)<http://mkwc.ifa.hawaii.edu/archive/wx/cfht/>; ^(b)<http://mkwc.ifa.hawaii.edu/current/seeing/mass/>; ^(c)constructed URL with ‘YYYY’ as placeholder for the year, ‘MM’ for the month and ‘DD’ for the day: <https://www.cadc-ccda.hia-ihp.nrc-cnrc.gc.ca/data/pub/JCMT/YYYYMMDD.wvm>; Col. 10; ^(d)<https://www.eso.org/gen-fac/pubs/astclim/paranal/clouds/clouds-LP0.txt>; ^(e)<https://www.eso.org/gen-fac/pubs/astclim/lasilla/h2o/laseranal.dat>; ^(f)<http://www.not.iac.es/weather/archive/met/>; ^(g)<http://catserver.ing.iac.es/robodimm/>; ^(h)(Molinari & Hernandez 2012, 2014); ⁽ⁱ⁾(Brookfield et al. 2020); ^(j)<http://tango.astrosen.unam.mx/weather15/>, (Plauchu-Frayn et al. 2020); ^(k)https://sitedata.tmt.org/Available_data/data_fields.html

humidity as well as the valid cloud cover data lie between 0% and 100%. The valid temperature data lie between -20°C and 50°C . Every pressure value with an error of more than ± 100 hPa with respect to the mean value was filtered. The specific humidity is calculated from values that were filtered as just described. The valid range of seeing ranges from 0 to 10 arcsec. The filters are tolerant because the measurement errors we found appeared to be very extreme, for example a temperature value of 2000°C or a seeing value of -1.2 arcsec.

4.1.1. In situ measurements of meteo data

For all sites, temperature, relative humidity and pressure are included in “Meteo” in Table 2, except for SPM where the pressure came from the TMT site testing campaign (Schoeck et al. 2009). No meteorological data of La Silla are available in 2009 and 2010 due to technical issues during a replacement of the weather tower. Meteo data are most often measured by an automatic weather station at a high frequency of 15 min at minimum (Table 2). In Siding Spring, meteo data are measured 1.5 m above ground, on La Silla, Cerro Paranal, Cerro Tololo and La Palma, 2 m above ground, at Sutherland and Mauna Kea, 3 m above ground and at SPM 6 m above ground.

4.1.2. In situ measurements of PWV

Not all observatories measure and archive the PWV. Measurements of PWV are only available for Mauna Kea, La Silla and Cerro Paranal. At Paranal, PWV is measured with the Low Humidity and Temperature Profiling microwave radiometer (LHATPRO; Kerber et al. 2012). LHATPRO is specifically designed for low humidity environments with low PWV values as it is the case at Paranal. The device uses the strong water vapour emission line at 183 GHz for humidity profiling. The temperature is measured separately. At La Silla, PWV is not measured on a regular basis but was retrieved for the VLT site selection process from sky radiance monitors (Sarazin & VLT Site Selection Working Group 1990) and later investigated with different techniques including archival optical spectra done with various spectrographs and radiosonde launches (Querel et al. 2010). The data used in this study were provided by ESO¹² and are based on satellite measurements (Table 2). On Mauna Kea, PWV is converted from opacity values measured by radiometers at a frequency of 225 GHz and 183GHz¹³.

¹² <https://www.eso.org/gen-fac/pubs/astclim/lasilla/h2o/>, accessed on 2020 Aug. 30.

¹³ <https://www.eaobservatory.org/jcmt/observing/weather-bands/>, accessed on 2020 Aug. 28.

4.1.3. In situ measurements of cloud cover

No observatory could provide cloud cover measurements. Instead, observatories on Mauna Kea, La Palma and Siding Spring record and archive the time lost due to bad weather at night (Table 2). This so-called weather loss could be triggered by clouds that absorb visible and infrared light, but there are also other reasons such as high wind speed or high humidity¹⁴. ESO provides monthly fractions of photometric nights for Paranal and La Silla (Table 2). We substituted weather loss data of Cerro Tololo with weather loss data from Cerro Pachon, since the distance between the two sites is only 10 km.

4.1.4. In situ measurements of seeing

In situ data of seeing come from measurements with a DIMM in Paranal, La Silla, Cerro Tololo and La Palma. For Paranal, data from a combined MASS-DIMM instrument are available only between 2016 and 2019. In situ data from SPM and Mauna Kea are measured with a MASS instrument. For Siding Spring, because seeing is measured by different instruments and occasionally with a DIMM, seeing might be influenced by dome air conditioning¹⁵. However, seeing measurements are only provided per semester (March – August, September – February).

4.2. ERA5 reanalysis

A reanalysis is a continuous weather simulation that assimilates archived observational data. The simulation of observational data via the laws of physics built into the weather model act as a physics-based interpolator and provide 4-dimensional data given by three spatial dimensions and a temporal dimension (Parker 2016).

In this study, we consider a reanalysis for two reasons. First, a reanalysis builds a bridge between in situ observations and GCMs, as it provides homogeneous and continuous data over a long time period that enables the validation of GCMs. Second, a reanalysis can help verify the feasibility of the study. Its spatial resolution is either similar or higher than the resolution of GCMs so any technical issue encountered with the reanalysis may be similar with GCMs. Furthermore, a reanalysis provides data for sites where no observation is available and makes sites more inter-comparable.

We use the latest ECMWF ERA5 reanalysis, which was made publicly available in early 2019 and then updated every 3 months (Hersbach et al. 2018a,b). According to IPCC (2021), ERA5 is the most reliable reanalysis, at least to assess climate trends. The ERA5 reanalysis provides hourly data on single or multiple pressure levels from 1979 onwards, and since 2021 preliminarily also from 1950 (Bell et al. 2020). Data are available for the atmosphere, the land surface and ocean waves of the Earth. The horizontal resolution of ERA5 is $0.25^\circ \times 0.25^\circ$ (latitude \times longitude), where 0.25° correspond to 27.75 km at the equator. The atmosphere is represented on 137 pressure levels, and outputs are available on 37 pressure levels (Fig. D.1). The pressure level variables we use are air temperature, specific humidity, relative humidity, geopotential height and the two horizontal components of wind speed. The single level variables we use are total column water and total cloud cover. The total column water includes all kinds of water in the atmosphere such as water vapour, liquid water, cloud ice, rain and snow. The total column water is given as instantaneous values, which refer to

one point in time and not to a temporal average. For total cloud cover, monthly means of night-time values are considered for the comparison to in situ data as well as the comparison with PRIMAVERA.

In ERA5, data on single levels are computed on the model surface. The model surface elevation, also called orography, does not coincide with Earth's surface elevation at any specific location because of scale mismatch. The ERA5 reanalysis uses an interpolation from different surface elevation data sets to calculate the orography¹⁶. The orography of ERA5 is shown in Figs. C.1 to C.8 for the eight sites and discussed in Sect. 5.2.4. Due to its spatial resolution, the orography of ERA5 is smoother than Earth's surface elevation. As a consequence, elevations of the observatories are lower in ERA5 than observed, although they are generally located on the locally highest elevated grid cells of ERA5. This difference in elevation is an issue when evaluating surface data. To solve this problem, we considered data on pressure levels, which consequently have a pressure offset to the surface pressure unique to each grid cell, which only varies slightly in time.

4.3. PRIMAVERA GCMs

We use GCMs provided by the European Horizon 2020 PRIMAVERA project (PRocess-based climate sIMulation: AdVances in high-resolution modelling and European climate Risk Assessments)¹⁷. The PRIMAVERA project provides the European contribution to HighResMIP (Haarsma et al. 2016), endorsed by CMIP6. The initialisation of CMIP dates back to 1995 (Meehl 1995), a project whose main objective is to compare the latest GCM simulations of the past, present and future, ran under controlled conditions, in order to assess changes in climate (Meehl et al. 2000). The latest CMIP Phase 6 (CMIP6) focusses on 1) Earth's response to radiative forcing, 2) systemic model biases and 3) future climate changes under the light of internal climate variability, predictability and scenario uncertainties (Eyring et al. 2016).

The primary goal of PRIMAVERA is to deliver a new generation of advanced high-resolution GCMs in order to provide detailed climate information at the regional and global scales, including extreme events. The horizontal resolution of PRIMAVERA GCMs is higher than those provided by the standard IPCC CMIP activities (CMIP3, CMIP5, CMIP6). Therefore, PRIMAVERA also enables to evaluate the effects of increasing spatial resolution on climate simulations¹⁸. Never before has a comparable study been conducted with an ensemble of GCMs that ensure such a high level of fidelity. A detailed overview of the PRIMAVERA models is given in Table 3. Each GCM uses a different dynamical core with a grid spacing between 18 km and 50 km. All monthly pressure level data are output on the same 19 vertical pressure levels (Fig. D.1). Almost all PRIMAVERA GCMs provide a historical and a future simulation (Table 3). The historical simulation covers the period from 1950 until 2014 and the future simulation goes from 2015 until 2050. There are two types of simulations: an atmosphere-land only (hereafter atmosphere-only) and a coupled atmosphere-ocean (hereafter coupled) climate simulation. While the coupled models allow for interactions between the ocean and

¹⁶ <https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation>, accessed on 2020 Aug. 30.

¹⁷ <https://www.primavera-h2020.eu/>, accessed on 2020 Jun. 07.

¹⁸ See <https://www.primavera-h2020.eu/output/scientific-papers/> for a complete list of publications using PRIMAVERA data sets.

¹⁴ https://www.eso.org/sci/facilities/paranal/sciops/At_Telescope.html accessed on 2021 Jan. 14.

¹⁵ Priv. comm. with Dr. Daniel Cotton (AAT).

Table 3. Overview of PRIMAVERA models.

Model name	HadGEM3-GC31-HM	EC-Earth3P-HR	CNRM-CM6-1-HR	MPI-ESM1-2-XR	CMCC-CM2-VHR4	ECMWF-IFS-HR
Institute	Met Office	KNMI, SMHI, BSC, CNR	CERFACS	MPI-M	CMCC	ECMWF
Reference	Roberts (2017) Schiemann et al. (2019)	EC-Earth Consortium (2018)	Voltaire (2019)	von Storch et al. (2017)	Scoccimarro et al. (2017)	Roberts et al. (2017)
Short name	HadGEM	EC-Earth	CNRM	MPI	CMCC	ECMWF
Atmosphere horizontal resolution (at 50N°)	25 km	36 km	50 km	34 km	18 km	25 km (Output at 50 km)
Atmos dynamical scheme (grid)	Grid point (SISL, lat-long)	Spectral (linear, reduced Gaussian)	Spectral (linear, reduced Gaussian)	Spectral (triangular, Gaussian)	Grid point (finite volume, lat-long)	Spectral (cubic octohedral, reduced Gaussian)
Ensemble members	r1i1p1f1	r1i1p2f1	r1i1p1f2	r1i1p1f1	r1i1p1f1	r1i1p1f1
Historical simulations	coupled-past atmos-past	coupled-past ^(a)	coupled-past atmos-past	coupled-past atmos-past	coupled-past atmos-past	coupled-past
Future simulations	coupled-future atmos-future	coupled-future ^(a)	coupled-future atmos-future	...	coupled-future atmos-future	coupled-future

Notes. The terms ‘coupled-past’ and ‘coupled-future’ stand for the historical and future simulations of the coupled models and the terms ‘atmos-past’ and ‘atmos-future’ stand for the atmosphere-land models, respectively. The ensemble member identifiers ‘ripf’, e.g. r1i1p2f1, stand for realisation ‘r’, initialisation ‘i’, physics ‘p’ and forcing ‘f’ (Taylor et al. 2018). The realisation defines the initial conditions. The initialisation distinguishes between different initialisation procedures. The physics parameter indicates the physics schemes used. Table adapted and extended from <https://www.primavera-h2020.eu/modelling/our-models/>, accessed on 2020 Aug. 06. ^(a)Not used for cloud cover.

the atmosphere, atmosphere-only simulations are constrained by observed sea surface boundary conditions (sea ice cover and sea surface temperature), which simplifies the representation of the climate system by removing interactions with the ocean and makes their results more comparable to observations and reanalyses. Moreover, atmospheric GCMs do not lack the ability to project anthropogenic forcings for the land and the atmosphere (He & Soden 2016). The PRIMAVERA atmosphere-only simulations are forced by HadISST2.2.0.0 sea surface boundary conditions¹⁹. The PRIMAVERA atmosphere-only simulations provide an opportunity to decide whether performance issues arise from ocean-atmosphere coupling (Moreno-Chamarro et al. 2022).

PRIMAVERA coupled and atmosphere-only simulations use the same external radiative forcings, as described in the High-ResMIP protocol (Haarsma et al. 2016). Those are based on observations for the historical period and on the shared socio-economic pathway 5–8.5 (SSP5–8.5; Krieglner et al. 2017) for the future period, which is built upon the CMIP5 representative concentration pathway (RCP) 8.5. The RCP 8.5 (Riahi et al. 2011) exhibits highest greenhouse gas emissions compared with the other existing RCPs. It demonstrates the worst case scenario and does not include any mitigation measures. If we follow RCP 8.5 until the end of the century, we would end up with a radiative forcing of 8.5 W m⁻². The two atmosphere-only simulations are called highresSST-present (hereafter: atmos-past) and highresSST-future (hereafter: atmos-future), since they make use

¹⁹ For details, refer to <https://hrm.ceda.ac.uk/research/cmip6-highresmip/highresmip-protocol/> and Kennedy et al. (2017).

of SST measurements. The coupled climate model simulations are called hist-1950 (hereafter: coupled-past) and highres-future (hereafter: coupled-future).

Climate model simulations approximate deterministic chaotic but non-linear and non-periodic processes by discrete systems, which makes them sensitive to initial conditions. Furthermore, coupled climate models have to overcome the difficulty of coupling two subsystems with very different timescales, defined as the time needed for a new equilibrium after a small perturbation, which are much shorter for the atmosphere than for the ocean (Peixoto et al. 1992). Therefore, PRIMAVERA coupled GCMs have a spin-up period of 50 yr to ensure physical equilibrium (Haarsma et al. 2016). A major role for climate variability play quasi-periodic cycles such as ENSO or the quasi-biennial oscillation (QBO; e.g. Baldwin et al. 2001), which are simulated by GCMs. In general and concluding from this and the previous section, we expect ERA5 to agree better with in situ observations than PRIMAVERA GCMs agrees with in situ observations and ERA5.

5. Model evaluation and trend methods

5.1. Data download and code

We downloaded hourly ERA5 data (Hersbach et al. 2018b,a) through the Copernicus climate change service climate data store (Copernicus Climate Change Service (C3S) 2020), and PRIMAVERA data through the Earth system grid federation node²⁰. The

²⁰ <http://esgf-index1.ceda.ac.uk>, accessed on 2020 Sep. 20.

Table 4. ERA5 and PRIMAVERA pressure levels used for temperature (T), relative humidity (RH) and specific humidity (SH) and vertical integration lower limits for PWV and seeing model (Eq. (13)).

Site	T, RH, SH	T, RH, SH	PWV	PWV	Seeing model	Seeing model
	ERA5 (hPa)	PRIMAVERA (hPa)	ERA5 (hPa)	PRIMAVERA (hPa)	ERA5 (hPa)	PRIMAVERA ^(a) (hPa)
Mauna Kea	600	600	600	600	800	850
Cerro Paranal	750	700	750	700	900	850
La Silla	775	850	775	700	825	850
Cerro Tololo	800	850	775	850	825	887.5
La Palma	800	850	775	700	1000	925
Siding Spring	900	925	900	925	950	925
Sutherland	850	850	825	850	850	850
SPM	750	700	750	700	850	925

Notes. ^(a)Median of individual model surface pressure values.

complete code we wrote for our analysis is available online on Github²¹.

5.2. Evaluation of ERA5 and PRIMAVERA

For evaluating ERA5 and PRIMAVERA against in situ data, we considered the grid point that is closest to the location of the astronomical site for each individual climate model. There are several reasons for considering the nearest grid point instead of a possible weighted grid point: (1) for sites on islands or on the coast, we can select the nearest point on land, while the weighted nearest point method may give weights to points that are in the ocean; (2) for sites near mountain tops, the circulation may induce different climate whether the sites are on the windward or the leeward side of the mountain. This may not be so relevant at spatial resolution of CMIP6 models, but it starts to be more and more relevant as the resolution of the model increases, such as for PRIMAVERA. The nearest grid point therefore seems more appropriate. More information about the orography is provided in Sect. 5.2.4 and in the Appendix C. For the analysis of the six PRIMAVERA GCMs (Table 3), we calculated a multi-model ensemble mean for each simulation (atmos-past, atmos-future, coupled-past, coupled-future) and we expect these ensemble means to outperform individual GCM simulations, based on Gleckler et al. (2008).

For temperature, relative humidity and specific humidity, we considered ERA5 and PRIMAVERA pressure levels that show a compromise between best skill score, best representation of in situ orography (closest pressure level) and minimal absolute bias between models and observations (Table 4). While testing different pressure levels, we noticed that future trends were not sensitive to the choice of the pressure level, for example between 700 and 850 hPa for PRIMAVERA, the values of future trends are very similar (not shown).

We considered the “total cloud cover” model variable to assess the cloudiness. This is a single level variable that includes clouds over the entire atmospheric column of the model.

5.2.1. General workflow

To best represent night-time astronomical observations, it would be preferable to validate night-time ERA5 and PRIMAVERA

against night-time in situ observations. However, preliminary analyses on the diurnal cycle using hourly data show that hourly ERA5 data are not able to represent the in situ diurnal cycle (Appendix B). This could be caused by the use of convective parameterisation and by the still too coarse horizontal resolution and discrepancies in orography of ERA5 compared to the real topography. Therefore and because we cannot expect GCMs to show a better match for the diurnal cycle than ERA5, we decided to evaluate monthly-mean data.

Monthly ERA5 data are validated against in situ data by using a skill score classification (see Sect. 5.2.5 and Table 5). That classification is also used to validate monthly historical PRIMAVERA GCM data against in situ and ERA5 data. We investigated trends in future and historical PRIMAVERA simulations and in ERA5. Historical trend analysis allowed to set PRIMAVERA future projections into context. More details about the methods of trend analysis can be found in Sect. 5.3.

5.2.2. Calculation of specific humidity and PWV

Specific humidity is computed following the equation for the saturated vapour pressure e_s (Buck 1981)

$$e_s = 6.1121 \cdot \exp\left(\left(18.678 - \frac{T}{234.5}\right)\left(\frac{T}{257.14 + T}\right)\right), \quad (7)$$

where the temperature T is in °C and e_s is in hPa. With the formula for the saturation specific humidity (Peixoto et al. 1992)

$$q_s = 0.622 \cdot \frac{e_s}{P} \quad (8)$$

and Eq. (1), we arrive at a formula for specific humidity computed with relative humidity RH (in decimal notation), temperature T (in °C) and pressure P (in hPa)

$$q = 3.802 \cdot RH \cdot \frac{1}{P} \exp\left(\left(18.678 - \frac{T}{234.5}\right)\left(\frac{T}{257.14 + T}\right)\right). \quad (9)$$

We use this formula to calculate specific humidity using relative humidity, pressure and temperature in situ data measured by observatories.

The PWV is computed directly from Eq. (3) by replacing the integral with the sum over all discrete pressure levels

$$PWV = \frac{1}{g} \sum_{i=1}^N q_i \cdot \Delta p_i, \quad (10)$$

²¹ <https://github.com/CarolineHaslebacher/Astroclimate-future-project>, accessed on 2021 Sep. 22.

where p_i is the pressure (in hPa) and q_i is the specific humidity (in kg kg^{-1}) on pressure level i . We use the Euler forward method to compute Δp_i .

Specific humidity values are integrated from the surface pressure (Table 4) of the observatory to the top of the atmosphere (250 hPa), following Eq. (10). This integral is called the “integrated PWV” hereafter. Moreover, we verified that ERA5 integrated PWV is equivalent to ERA5 total column water (as described in Sect. 4.2), which confirms that total column water is mostly composed of water vapour in long term means (Peixoto et al. 1992).

5.2.3. Calculation of astronomical seeing

The contribution of the C_n^2 profile that varies the most is the ground layer contribution, which consists of the boundary layer (0 km to 1–2 km altitude) and the surface layer (0 m to 50 m altitude), whereas the free atmosphere contribution (1–2 km to 20 km altitude) appears to be similar independent of the geographical site location (Abahamid et al. 2004). Racine (2005) shows that surface layer turbulence is the dominant cause of seeing at low elevations above ground, decreasing with a scale height of about 3.5 m. He has proposed a model vertical distribution of the turbulence which can fit the median seeing values of major telescopes over the world only based upon altitude and elevation. The GCMs used in our study do not have the required resolution to model surface layer turbulence, so we must assume that the top of large telescopes enclosures is naturally high enough above ground (15 m or more) so that the surface turbulence layer does not influence the seeing at the telescope level. Moreover, even in open enclosures such as the VLT that uses passive flushing, the outside local turbulence cannot easily enter at the level of the observing floor (10 m above ground) partly because of the wind screen on the front and of the smaller size ventilation slots on the back.

We introduce two different approaches to estimate seeing with ERA5 and PRIMAVERA output: The first approach, identified as the ‘200-hPa-wind-speed seeing’ hereafter, uses the correlation between wind speed at 200 hPa, where strong winds prevail, and the free atmosphere seeing (Vernin 1986; Racine 2005; Sarazin et al. 2013). This correlation, which is based on physical shear, is expected to exist at all sites, modulated locally by the intensity and frequency of appearance of the jet stream. The turbulence integral can be split into a lower and an upper contribution:

$$\int_{0 \text{ km}}^{20 \text{ km}} C_n^2(z) dz = \int_{0 \text{ km}}^{2 \text{ km}} C_n^2(z) dz + \int_{2 \text{ km}}^{20 \text{ km}} C_n^2(z) dz. \quad (11)$$

The upper contribution set between 2 km and 20 km is essentially the free atmosphere seeing, which is dominated by wind speeds at 200 hPa, whereas the lower contribution set between 0 km and 2 km represents the surface layer and planetary boundary layer seeing. For wind speeds greater than 1 m s^{-1} , wind shear produces wind speeds that are approximately proportional to turbulent velocities (Mahrt et al. 2013). Empirically, Vernin (1986) find:

$$\int_{2 \text{ km}}^{20 \text{ km}} C_n^2(z) dz \approx A \cdot (u_{200 \text{ hPa}}^2 + v_{200 \text{ hPa}}^2), \quad (12)$$

where u and v are the horizontal components of the wind velocity and A is a calibration constant.

The second approach, hereafter referred to as the ‘seeing model’, is based on Osborn & Sarazin (2018). Their formula for

the C_n^2 is based on temperature, wind speed and pressure of an atmospheric vertical column and is a combination of formulas introduced by Tatarskii (1971), Masciadri et al. (1999, 2017), and the Gladstone relation (Gladstone & Dale 1863). An estimation of the optical turbulence refractive index structure constant C_n^2 is given by

$$C_n^2(z) = k \cdot \left(\frac{80 \times 10^{-6} \cdot P(z)}{T(z)\Theta(z)} \right)^2 \cdot L(z)^{4/3} \cdot \left(\frac{d\Theta}{dz} \right)^2, \quad (13)$$

where $P(z)$ is the pressure, $\Theta(z)$ is the potential temperature, $T(z)$ is the temperature at the altitude z and k is a calibration factor, which depends on the atmospheric stability (Masciadri et al. 1999). The calibration factor was determined with the mean of in situ seeing data. For Sutherland, we used a mean of 1.32 arcsec (Catala et al. 2013). The scale of the largest energy input $L(z)$, also called outer scale (Masciadri et al. 1999), is defined by

$$L(z) = \sqrt{\frac{2S^2}{\frac{g}{\Theta(z_i)} \left| \frac{d\Theta}{dz} \right|}}, \quad (14)$$

where the vertical wind shear S

$$S = \left(\left(\frac{du}{dz} \right)^2 + \left(\frac{dv}{dz} \right)^2 \right)^{1/2} \quad (15)$$

is the square root of the turbulent kinetic energy. The potential temperature Θ is given by

$$\Theta = T \left(\frac{P_0}{P} \right)^{R/c_p}. \quad (16)$$

The variable R is the gas constant and c_p is the specific heat at constant pressure of air ($R/c_p = 0.286$). The variable P_0 is the reference pressure at sea level for which we choose $P_0 = 1000 \text{ hPa}$.

The advantage of this numerical turbulence model is that GCM output can be used for the calculation. However, Osborn & Sarazin (2018) only validated the seeing model for Cerro Paranal. For integrating C_n^2 following Eq. (4), we use the Euler forward numerical scheme and integrate to the pressure level defined in Table 4 to include the turbulence ground layer of the model. To calculate the seeing, we used Eqs. (5)–(6) for both approaches. We expect to gain complementary insights from these two different approaches.

5.2.4. Model orography

The model surface elevation (also called orography) is different for all models because they all use different numerical grids (Table 3). Figures C.1–C.8 show the orography of each PRIMAVERA model as well as the orography of the ERA5 reanalysis for all eight sites. Differences in mountain heights between ERA5 and PRIMAVERA GCMs are attributed to different numerical schemes, different spatial resolutions and to different reference satellite data sets used before being interpolated into the model grids.

The MPI, CNRM, ECMWF and EC-Earth PRIMAVERA GCMs use spectral methods that make use of orthogonal spherical harmonics (Table 3). Despite valuable advantages, spectral models are not able to resolve sharp discontinuities because they use a finite amount of wave numbers. This leads to the creation of

unnatural gravity waves (over- and undershooting), called Gibbs oscillations or Gibbs ripples (Navarra et al. 1994). To avoid Gibbs oscillations, the orography is commonly smoothed out (e.g. Vanni re et al. 2020), but more aggressive methods have been found recently, moving ripples from the far ocean closer to the coast (Wedi 2014).

All PRIMAVERA orographies as well as the ERA5 orography exhibit a lower elevation than the elevation of the observatory on the nearest grid point for all sites. This is reasonable since the observatories are located on highly elevated terrain.

Observatories located near coastlines or on an island also present a challenge for our analyses. The land-sea mask (contour lines in Figs. C.1–C.8) defines the frontier between land and sea. For ERA5, HadGEM and CNRM, the land-sea mask has grid points with fractional values, while for EC-Earth, MPI, ECMWF and CMCC only integer values of 1 for land points and 0 for sea points are used. The ERA5 and PRIMAVERA land-sea masks coincide well with the coastlines (Figs. C.1–C.8) in all sites but La Palma.

The island of La Palma is a special case due to its small size. It is treated very differently in all models (Fig. C.5). MPI does not show La Palma as an island (no surface elevation and no land point). ECMWF and EC-Earth show elevated grid points classified as ocean where the island of La Palma lies, so effectively, at the location of La Palma, there is a mountain of water. These are examples of overshooting Gibbs ripples. ERA5, CNRM and the two CMCC and HadGEM grid point GCMs show a surface elevation classified as land. However, the island is smoothed: its elevation is much lower (by 2174 m at minimum), while its size is much larger (by several grid points).

In the case of Mauna Kea, which is located on the Hawaiian Big Island that is larger than La Palma, all models show a surface elevation classified as land. Moreover, all nearest grid points are within the coastline boundaries of Big Island, which is not the case for La Palma. However, as for La Palma, all models represent a much smoother orography than the real topography around Mauna Kea.

Despite the horizontal grid spacings of 18 to 50 km, it is challenging for high-resolution PRIMAVERA GCMs to represent the sites accurately. Based on this assessment, we decided not to include any CMIP5 and CMIP6 GCMs, which are resolved on much coarser horizontal resolutions of 70–400 km and for which the Hawaiian Big Island and La Palma are not represented.

5.2.5. Model skill score

To evaluate the skills of the GCMs, we use Taylor (2001) diagrams, which combine correlation coefficient, centred root mean square error (RMSE) differences and standard deviation of GCMs and in situ observations. Figure 2 shows a fictitious diurnal cycle with its corresponding Taylor diagram. The observational data (black solid line in Fig. 2a) constitute the reference dataset used to evaluate the test data (coloured lines). In the example plot, we have four test data sets, the first is simply offset from the ground truth by a constant, the second overestimates the amplitude of the diurnal cycle, the third demonstrates a delay in phase and the fourth includes both biases of the second and third case. First of all, we note that the Taylor diagram (Fig. 2b) does not account for the bias of the test dataset compared to the reference dataset, so in the first case where the test data set is shifted on the y -axis in Fig. 2a, the point in the Taylor diagram is the same. In the Taylor diagram (Fig. 2b), the overestimation of the amplitude is seen as an increase in standard deviation, but reflects perfect correlation: it lies on the x -axis of the Taylor

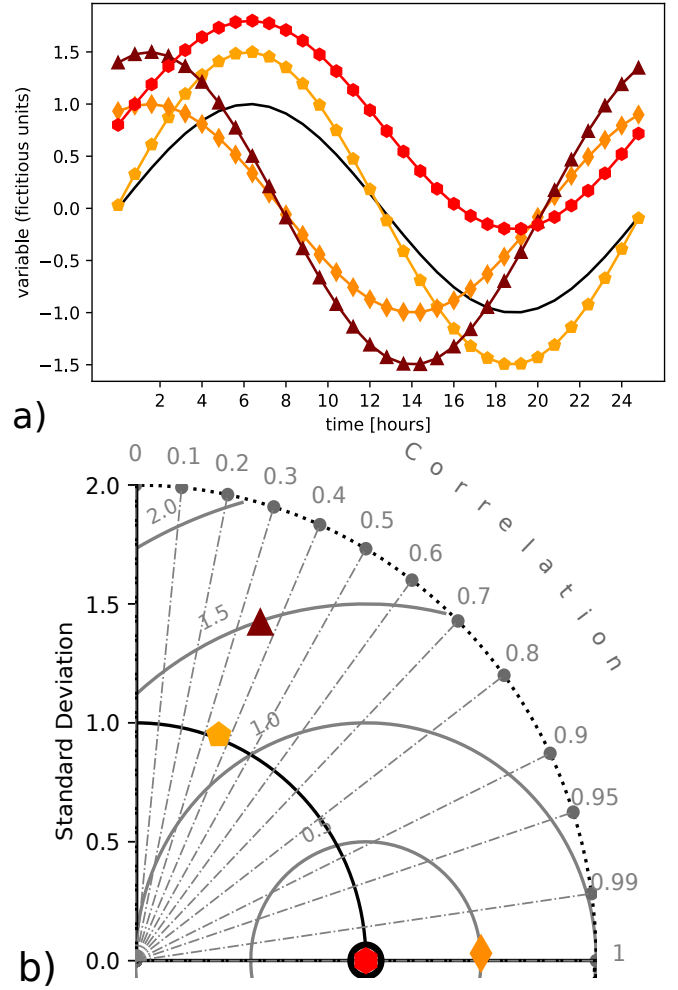


Fig. 2. Example of model performance assessment. (a) Fictitious diurnal cycle with (b) corresponding Taylor diagram. The solid black line in (a) represents the ground truth observation, whereas the solid coloured lines in (a) represent the test data. One test dataset shows an offset (or bias) from the ground truth (filled red hexagons), one overestimates the amplitude of the diurnal cycle (filled orange pentagons), one demonstrates a delay in phase (filled dark-orange diamonds) and one that includes both biases (filled maroon triangles). In the Taylor diagram in (b), the black circle on the x -axis represents the ground truth observation. The displayed sine-functions are all made up of $y = a \cdot \sin(x + b) + c$ with parameters a , b and c as follows: black solid: $a = 0$, $b = 0$, $c = 0$ hexagon (red): $a = 0$, $b = 0$, $c = 0.8$ pentagon (orange): $a = 1.5$, $b = 0$, $c = 0$ diamond (dark-orange): $a = 0$, $b = 1.2$, $c = 0$ up-triangle (maroon): $a = 1.5$, $b = 1.2$, $c = 0$.

diagram. The third dataset showcases a decreased correlation due to the phase delay, but exhibits the same standard deviation as the observation. Finally, the fourth dataset shows the same correlation as the third dataset and the same standard deviation as the second dataset.

In general, the shorter the distance to the reference point in the Taylor diagram, the better the model skill. There are different approaches to quantify the skill of a model. Taylor (2001) suggests two formulas for the calculation of a skill score, one emphasising the correlation and the other highlighting the pattern variation. In this study, we use the formula that puts emphasis on the correlation

$$S = \frac{4(1 + R)^4}{(\hat{\sigma}_f + 1/\hat{\sigma}_f)^2(1 + R_0)^4}, \quad (17)$$

Table 5. Classification of the skill score based on requirements of the correlation (corr) between climate model data and observational in situ data and the standard deviation (std dev) of the model.

Classification	Corr	Std dev	Skill score
excellent	>0.95	< \pm 0.05	>0.91
good	0.80–0.94	\pm (0.04–0.20)	0.63–0.90
mediocre	0.60–0.79	\pm (0.21–0.40)	0.32–0.62
poor	<0.60	> \pm 0.40	<0.32

Notes. The requirements for the model standard deviation are displayed as deviations from the normalised reference standard deviation, which is always 1.

with R being the correlation coefficient and σ_f the standard deviation of the GCM. The hat on top of the standard deviation denotes the normalisation. The parameter R_0 accounts for the fact that no model is able to match observations exactly. It therefore describes the highest possible correlation. Here, R_0 is set to 0.995 arbitrarily. The closer to 1 the skill score S , the better the skill of the model.

For this study, we define a guideline for classifying the skill score between model data and observational data (Table 5). The guideline is based on requirements for the correlation and the model standard deviation out of which the skill score is calculated using Eq. (17).

5.3. Methods for trend analysis

We analyse past and future trends of monthly time series of ERA5 and PRIMAVERA ensemble mean with Bayesian inference. The Bayesian approach allows to approximate the trend per decade while receiving posterior distributions and credible intervals. This way, we make use of all the data available. The credible intervals allow us to decide whether the trends are strong and certain (small credible intervals) or whether the trends are weak and uncertain and not larger than internal climate variabilities (large credible intervals).

We approximate the trends in a linear way which smooths out climate variability since we focus on the mean trends between 2015 and 2050. We chose to display trends in unit per decade to have a natural timescale on which climate change might already be visible, compared to trends expressed in unit per year.

We use a Gaussian probability density function as the Bayesian likelihood function for our data points

$$h_i \sim \mathcal{N}(\mu_i, \sigma), \quad (18)$$

with the standard deviation σ chosen as a weak prior with a uniform distribution with non-zero values between 0 and 10. For the mean μ_i , we chose to infer a sine-shaped seasonal cycle, which could increase, decrease or stagnate linearly with the monthly time series. This progression with time, namely the historical and future yearly trends, is implemented in the model by an added linear function with slope b and intersect a . With this, we arrive at

$$\mu_i = a + b \cdot x_i + C \cdot \sin\left(\frac{2\pi}{12} \cdot (x_i + \phi)\right), \quad (19)$$

where the slope b , the amplitude C and the phase ϕ were chosen to have weak Gaussian prior distributions, which work with every dataset, centred around 0 with standard deviations of 3, 6 and 6, respectively. The offset a has a weak Gaussian prior

distribution centred around the mean of the monthly time series and a standard deviation of 5. The parameter x_i , the predictor variable, is the index of the month, starting at zero. As an uncertainty interval, we take an 89% percentile credible interval of predicted values sampled from the posterior distribution. The described Bayesian approach is visualised in Fig. 3. In simple words, the Bayesian algorithm searches for a set of parameters a , b , C , ϕ and σ that, given that Eq. (19) needs to be used, fit the data the best.

For the computation, we apply a Hamiltonian Monte Carlo (HMC) algorithm with 1000 samples (Betancourt 2017). We monitor the convergence of the chain by checking if the Rubin convergence diagnostic \hat{R} (Gelman & Rubin 1992) is within 2% deviation around 1. If this is not the case, we increase the number of samples to 4000 and run 2 chains.

We applied Bayesian inference using the rethinking package in R²². To allow for reproducibility of the results, we uploaded samples from the posterior distribution to the Github repository as well as a list of simulations for which the Hamiltonian Monte Carlo sampling output chains did not converge optimally.

6. Evaluation of PRIMAVERA GCMs

6.1. Current site conditions seen by ERA5 and PRIMAVERA GCMs

In this section, we focus on the analysis of the PRIMAVERA multi-model ensemble mean for atmos-past and coupled-past simulations (Table 3), using the skill score classification (Table 5). This evaluation aims to provide a basis for assessing the reliability of PRIMAVERA future projections for every site and variable.

The analysis of Taylor diagrams introduced in Sect. 5.2.5 shows that, in most cases, the PRIMAVERA multi-model ensemble mean tends to outperform the PRIMAVERA models taken individually. This is in line with previous studies and might be due to a smoothing of individual features (e.g. Gleckler et al. 2008; Reichler & Kim 2008).

6.1.1. Temperature

Seasonal cycle and time series. The seasonal cycle of temperature (Fig. 4) is well represented by ERA5 and PRIMAVERA for all sites. For Mauna Kea (Fig. 4a), for which the seasonal cycle shows the biggest discrepancies among all sites, all individual PRIMAVERA models as well as ERA5 underestimate the seasonal temperature amplitude. Possible reasons are summarised in Sect. 6.2.3 after we got an overview of all variables. Extra-annual oscillations such as ENSO are not represented in the seasonal cycle, but we see peaks in temperature in the time series on Mauna Kea (Fig. 4a), the Chilean sites (Figs. 4b–d), Siding Spring (Fig. 4f) and SPM (Fig. 4h) that are possibly linked to ENSO, since ERA5, in situ data and most of the time also PRIMAVERA atmosphere-only simulations coincidentally show a peak. Such a peak can be observed for example in the year 2003, 2009 or 2015, which are indeed El Niño years.

ERA5 versus in situ. Averaged over all sites, ERA5 shows an excellent skill score of 0.95 (Fig. 11a and Table A.1). The site with the highest skill score is Sutherland with an excellent skill score of 1.00 (Fig. 11a and Table A.1). This means that (1) in situ data are of high quality, (2) observations that were fed into

²² <https://github.com/rmcelreath/rethinking>, accessed on 2020 Aug. 25.

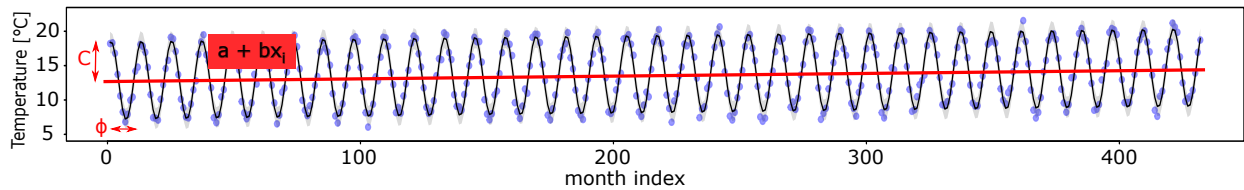


Fig. 3. Conceptual visualisation of Bayesian trend analysis with the example of a temperature time series. The monthly data points (blue filled circles) are approximated by a sine with amplitude C and phase ϕ , combined with a linear function with intercept a and slope b , so that a linear increase, decrease or stagnation of an inferred seasonal cycle with progressing month index x_i can be modelled (black line). With this Bayesian approach, sampling from the posterior distribution gives back an estimate of uncertainty, for which we chose an 89% percentile credible interval (shaded region).

ERA5 have a good temporal and horizontal resolution for the grid point nearest to Sutherland and (3) the nearest grid point is a valid choice, at least for the temperature. This is also true for the other sites that scored excellent, namely Paranal, La Silla, La Palma, Siding Spring and SPM. The site with the worst skill score is Cerro Tololo with a good skill score of 0.84 (Fig. 11a and Table A.1), which could be attributed to the downward trend in in situ data that is not captured by ERA5 (Fig. 4d), on which we elaborate further in Sect. 6.1.2. Despite the proximity of Cerro Tololo to La Silla, we do not see a similar decreasing temperature trend in in situ data of La Silla (Fig. 4c). Just a slightly bit higher scores Mauna Kea with a skill score of 0.86, which might indicate that either (1) in situ data are not of high quality, (2) observations that were fed into ERA5 do not have a sufficient temporal and horizontal resolution for the grid point nearest to Mauna Kea, (3) data assimilation on the ERA5 grid is difficult for Mauna Kea or (4) the nearest grid point does not represent the conditions at the observatory well, at least for the temperature. Regarding the fourth point, we tested other grid points close to the observatory and did not get an improvement. We conclude that ERA5 is not as successful in representing Mauna Kea than it is for most other sites.

PRIMAVERA versus ERA5. Averaged over all sites, PRIMAVERA shows a good skill score of 0.79 (Fig. 11a and Table A.1). This is the highest average skill score over all variables, which is in good agreement with the fact that temperature is the best modelled variable (IPCC 2021). The site that scores highest is Siding Spring with an excellent skill score of 0.93 (Fig. 11a and Table A.1). This is the highest skill score reached by PRIMAVERA in this study. As discussed previously in Sect. 4.3, we expect that PRIMAVERA GCMs show lower skill scores than the reanalysis ERA5 and we see this confirmed for temperature (Fig. 11a and Table A.1). The site with the lowest skill score is Mauna Kea with a mediocre skill score of 0.60 (Fig. 11a and Table A.1). Possible causes are summarised in Sect. 6.2.3. Cerro Paranal shows a similar skill score of 0.61 (Fig. 11a and Table A.1). In the Taylor diagram (Fig. 4b, Taylor diagram), we see that the standard deviation closely matches with ERA5 and in situ observations, but the correlation of 0.77 is lower compared to other sites. The correlation does not improve when the analyses are performed over eastward grid point neighbours that show higher elevation (not shown; see Fig. C.2 grid points).

6.1.2. Relative humidity

Seasonal cycle and time series. The seasonal cycle of relative humidity (Fig. 5) reveals a more complex behaviour than the seasonal cycle of temperature (Fig. 4). For example in the seasonal cycle of SPM, the relative humidity rapidly rises in July,

which is also captured by PRIMAVERA and ERA5 (Fig. 5h). In the seasonal cycle of Siding Spring, a decrease in relative humidity after 2014 is not fully captured by ERA5 (Fig. 5f). This bias is not accounted for in the Taylor diagram, which only considers monthly data until 2014 (Fig. 5f).

The in situ time series for Cerro Tololo (Fig. 5d) shows an increase in relative humidity, whereas ERA5 shows a slight decrease. The increasing trend in in situ data is confirmed by an equivalent trend observed by the weather station of the Gemini South Observatory on Cerro Pachon²³ (not shown). While a drying would be expected according to (Grenon 1990), an increase in relative and specific humidity could be linked to the construction of the ‘Embalse Puclaro’ storage lake²⁴, which was built in the proximity of Cerro Tololo and filled in 2002 (Ribeiro et al. 2015). This hypothesis is supported by the scarcity of ERA5 observational data in the region of Cerro Tololo and the Puclaro storage lake. The closest land weather station is situated at a longitude of -71.20° and latitude of -29.92° , which is approximately 45 km to the west of the storage lake²⁵. Nested climate models by Bischoff-Gauß et al. (2006) confirm changes in temperature and relative humidity 4 km downwind of the storage lake. Even though Cerro Tololo is situated approximately 20 km southward, it might still be influenced. The decreasing temperature (Fig. 4d) on Tololo would be in agreement with the enhancement of latent and sensible heat fluxes in the east due to the storage lake (Bischoff-Gauß et al. 2006). However, since in situ data measurements on Tololo start in 2002, we cannot compare to a period without storage lake. Nevertheless, these differences between ERA5 and in situ data on Cerro Tololo stem from a regional effect. On the other hand, PRIMAVERA does not show any obvious trend and simulates a double-peaking seasonal cycle that is not seen in the seasonal cycles of in situ or ERA5 data (Fig. 5d) that unlikely stems from an elevation mismatch between ground truth elevation and PRIMAVERA orography, since all individual models show the same structure in the seasonal cycle (not shown).

In contrast to other sites, the in situ seasonal cycle of Mauna Kea does not show a clear seasonal variability (Fig. 5a), which was also found with data from another observatory on Mauna Kea (Da Silva et al. 2012), but the coupled PRIMAVERA ensemble mean shows a peak in September, which explains the negative correlation in the Taylor diagram (Fig. 5a). The atmosphere-only PRIMAVERA simulation, which is constrained by SST measurements, exhibits a smaller amplitude

²³ Analysis is based on data spanning the years 2008–2020 requested via priv. comm. with the Gemini South Observatory.

²⁴ lon: -70.85° , lat: -30.00° .

²⁵ Analysis (not shown) is based on data request through the Copernicus Helpdesk at ECMWF <https://confluence.ecmwf.int/site/support>, accessed on 2021 Aug. 02.

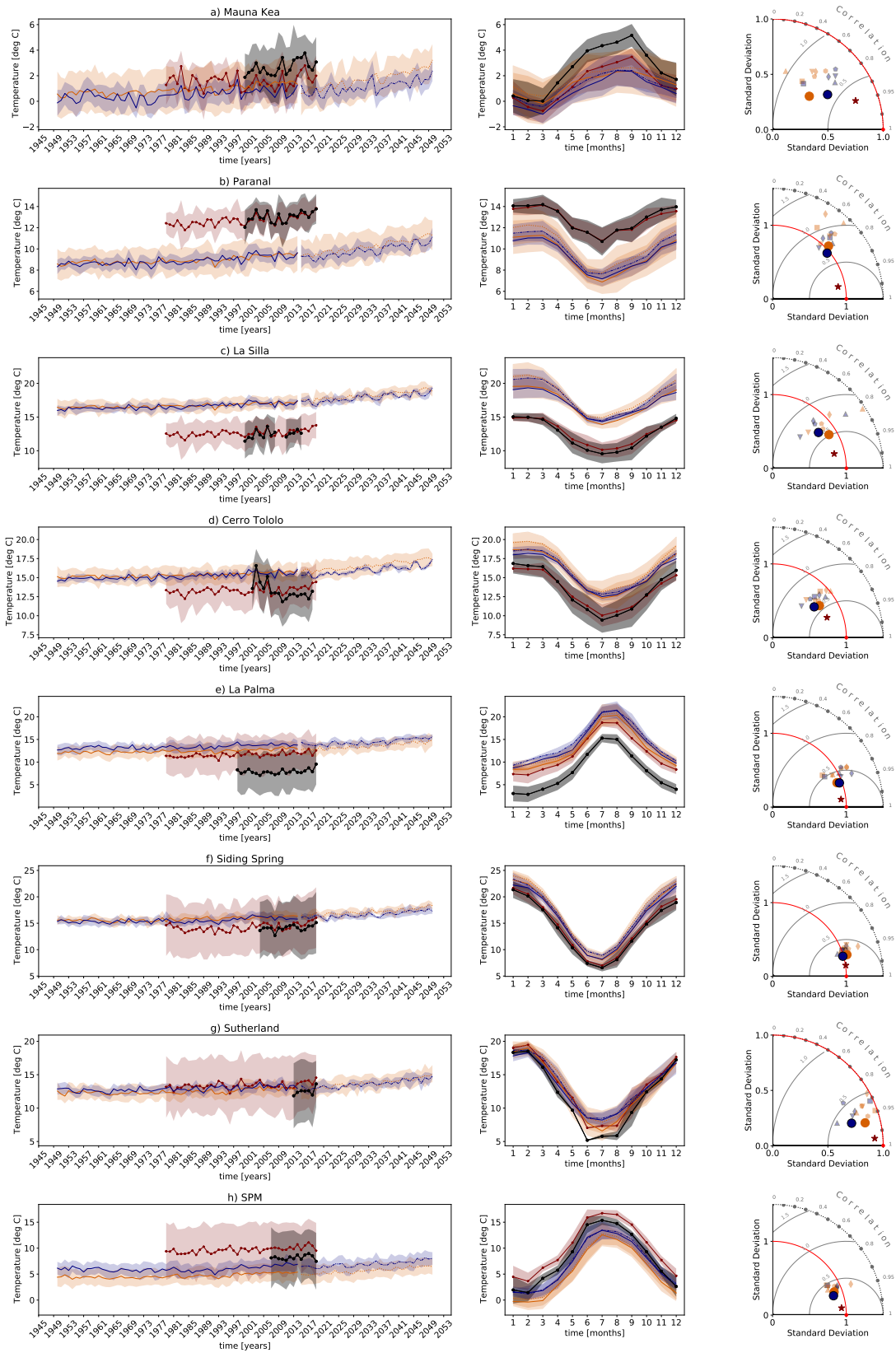


Fig. 4. Time series (*left column*), seasonal cycle (*middle column*) and Taylor diagram (*right column*) for temperature (T). The shaded uncertainty regions are standard deviations. The time series, seasonal cycle and Taylor diagram contain in situ data (black filled circles in time series and seasonal cycle; red filled circle on x -axis of Taylor diagram), ERA5 data (maroon filled stars) and PRIMAVERA data for coupled-past (orange solid), coupled-future (orange dotted), atmos-past (blue solid) and atmos-future (blue dotted). In the Taylor diagram, apart from the ensemble mean (filled circles), we show the individual PRIMAVERA GCMs, namely HadGEM3-GC31-HM (pentagon), EC-Earth3P-HR (hexagon), CNRM-CM6-1-HR (square), MPI-ESM1-2-XR (diamond), CMCC-CM2-VHR4 (up-triangle) and ECMWF-IFS-HR (down-triangle). The selected pressure levels are given in Table 4. The seasonal cycle and the Taylor diagram use monthly time-intersecting data (in situ measurements, ERA5 and historical PRIMAVERA). The future seasonal cycle is plotted for the time period 2015–2050.

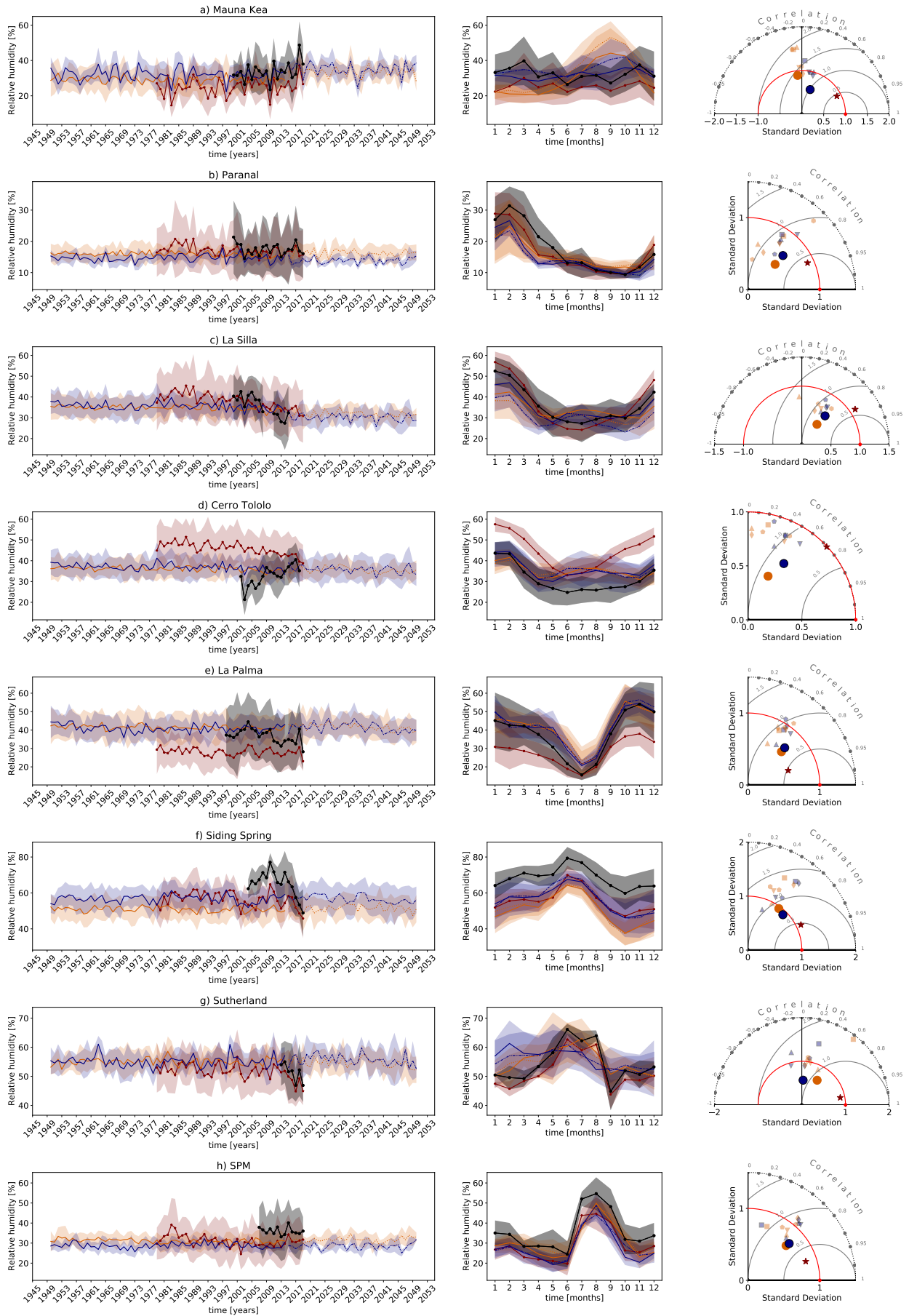


Fig. 5. Same as Fig. 4, but for relative humidity (RH).

of the seasonal cycle and a higher skill score than the coupled simulation (Fig. 5a), which indicates that the coupling of ocean to atmosphere poses additional problems for Mauna Kea. More reasons why PRIMAVERA does not exhibit the seasonal and interdecadal features seen in observations are discussed in Sect. 6.2.3.

ERA5 versus in situ. Overall, the skill scores of ERA5 show an average value of 0.78, which classifies as good (Fig. 11c and Table A.2). For Sutherland, which scores highest, the skill score is excellent with 0.96 (Fig. 11c and Table A.2). Again, we can argue for the same three points as in the above Sect. 6.1.1. The site with the lowest skill score of 0.57 is Cerro Tololo (Fig. 11c and Table A.2). It is the only site with a mediocre skill score, since all the sites except Sutherland score good. Reasons discussed in the previous paragraph might explain this.

PRIMAVERA versus ERA5. On average, over all sites, PRIMAVERA shows a mediocre skill score of 0.39 (Fig. 11c and Table A.2). The site with the highest skill score is SPM with a mediocre skill score of 0.51 (Fig. 11c and Table A.2). The site that scores lowest is Sutherland with a poor skill score of 0.21 (Fig. 11c and Table A.2). However, differences between the sites are small (Fig. 11c and Table A.2). Overall, the underperformance of PRIMAVERA with respect to the performance of ERA5 for relative humidity lets us conclude that PRIMAVERA is not able to represent monthly relative humidity mean values satisfactorily.

6.1.3. Specific humidity

Seasonal cycle, time series and vertical profile. Since the in situ specific humidity was calculated with Eq. (9), the time series and seasonal cycles show influences of both in situ temperature and in situ relative humidity (Figs. 5 and 6).

The comparison of the median vertical profile of specific humidity between ERA5 and PRIMAVERA shows a good agreement for all sites (Fig. F.1). Therefore, we expect a similarly good agreement between ERA5 and PRIMAVERA for PWV.

ERA5 versus in situ. On average, ERA5 has a good skill score of 0.89 (Fig. 11e and Table A.3). The site with the highest skill score is Sutherland with an excellent skill score of 0.99 (Fig. 11e and Table A.3). The lowest skill score is still good with a value of 0.82 and belongs to Mauna Kea. We notice that the spread of skill scores for ERA5 is as small as for the temperature.

PRIMAVERA versus ERA5. On average, PRIMAVERA has a good skill score of 0.66 (Fig. 11e and Table A.3). The site with the highest skill score is Siding Spring with 0.82 (Fig. 11e and Table A.3). The spread of PRIMAVERA skill scores is larger than for temperature due to Mauna Kea, which shows the worst skill score of 0.23 (Fig. 11e and Table A.3). Possible explanations are summarised in Sect. 6.2.3.

These projected local increases for the eight selected sites are in good agreement with globally increasing near-surface humidity.

6.1.4. PWV

Seasonal cycle and time series. As expected due to the close link given by Eq. (3), the PWV time series and seasonal cycle show similarities with specific humidity such as maxima during the same months (Figs. 6 and 7). Less pronounced

similarities exist between temperature and PWV and specific humidity due to enhanced evaporation and higher saturation specific humidity with higher temperature.

For Paranal, the in situ data are compared to future simulations of PRIMAVERA because PWV measurements are only measured since 2016 (Fig. 7b). Therefore, the adequate representation of the seasonal cycle (Fig. 7b) and the mediocre skill score of PRIMAVERA (Fig. 11g and Table A.4) support the credibility of PRIMAVERA future simulations.

ERA5 versus in situ. On average, ERA5 shows a good skill score of 0.70 (Fig. 11g and Table A.4). The skill score classification of ERA5 integrated PWV compared with in situ data ranges from mediocre for Mauna Kea and La Silla to excellent for Paranal (Fig. 11g and Table A.4). For the other sites, no in situ data are available. The excellent result for ERA5 at Paranal observatory strengthens confidence in the measurements performed by LHATPRO.

PRIMAVERA versus ERA5. On average, PRIMAVERA shows a mediocre skill score of 0.61 (Fig. 11g and Table A.4). The site with the best skill score is SPM with a good skill score of 0.86 (Fig. 11g and Table A.4). Sutherland and Siding Spring score good as well. The three Chilean sites and the two sites situated on islands La Palma and Mauna Kea all score mediocre. The worst site is Mauna Kea with a mediocre skill score of 0.34 (Fig. 11g and Table A.4). Possible reasons are summarised in Sect. 6.2.3.

Over sites where no PWV measurements are available (i.e. Cerro Tololo, La Palma, Siding Spring, Sutherland and SPM), we propose that PRIMAVERA is able to reliably simulate PWV because: (1) the differences in the specific humidity skill scores of PRIMAVERA versus in situ and PRIMAVERA versus ERA5 are small (Fig. 11e, blue versus red bars), (2) ERA5 matches good or excellent with in situ specific humidity (Fig. 11e) and can therefore be considered as a surrogate of observations and (3) the PWV skill scores of PRIMAVERA versus ERA5 are mediocre or good over sites without in situ PWV, which are comparable to specific humidity skill scores (Fig. 11e and g).

6.1.5. Cloud cover

Seasonal cycle and time series. We find that the seasonal cycle of in situ data corresponds with the seasonal cycle of ERA5 and PRIMAVERA for all sites except Mauna Kea, where ERA5 agrees to some degree with PRIMAVERA, and La Palma, where PRIMAVERA does not show a minimum in July in contrast to ERA5 and in situ data (Fig. 8). In the Taylor diagram of La Palma (Fig. 8e), the coupled CNRM and atmosphere-only HadGEM simulations lie close to ERA5, and all individual GCMs are scattered widely, which likely reflects the very different representations of orography. The CNRM and HadGEM models allow for fractional values in the land-sea mask (Sect. 4.3) and make up for two out of three models that classify the island of La Palma as land (Figs. C.5a, e). When we analyse the seasonal cycle of these two models, they indeed show a minimum cloud cover in July (not shown).

ERA5 versus in situ. On average, ERA5 nighttime cloudiness shows a mediocre skill score of 0.50 when compared to in situ data (Fig. 11i and Table A.5), but no in situ data are available for Sutherland and SPM. This average is a satisfying result given that in situ data include reasons for down-time other than clouds (Sect. 4.1.3). The site with the highest skill score is Siding Spring with a good skill score of 0.72 (Fig. 11i and Table A.5).

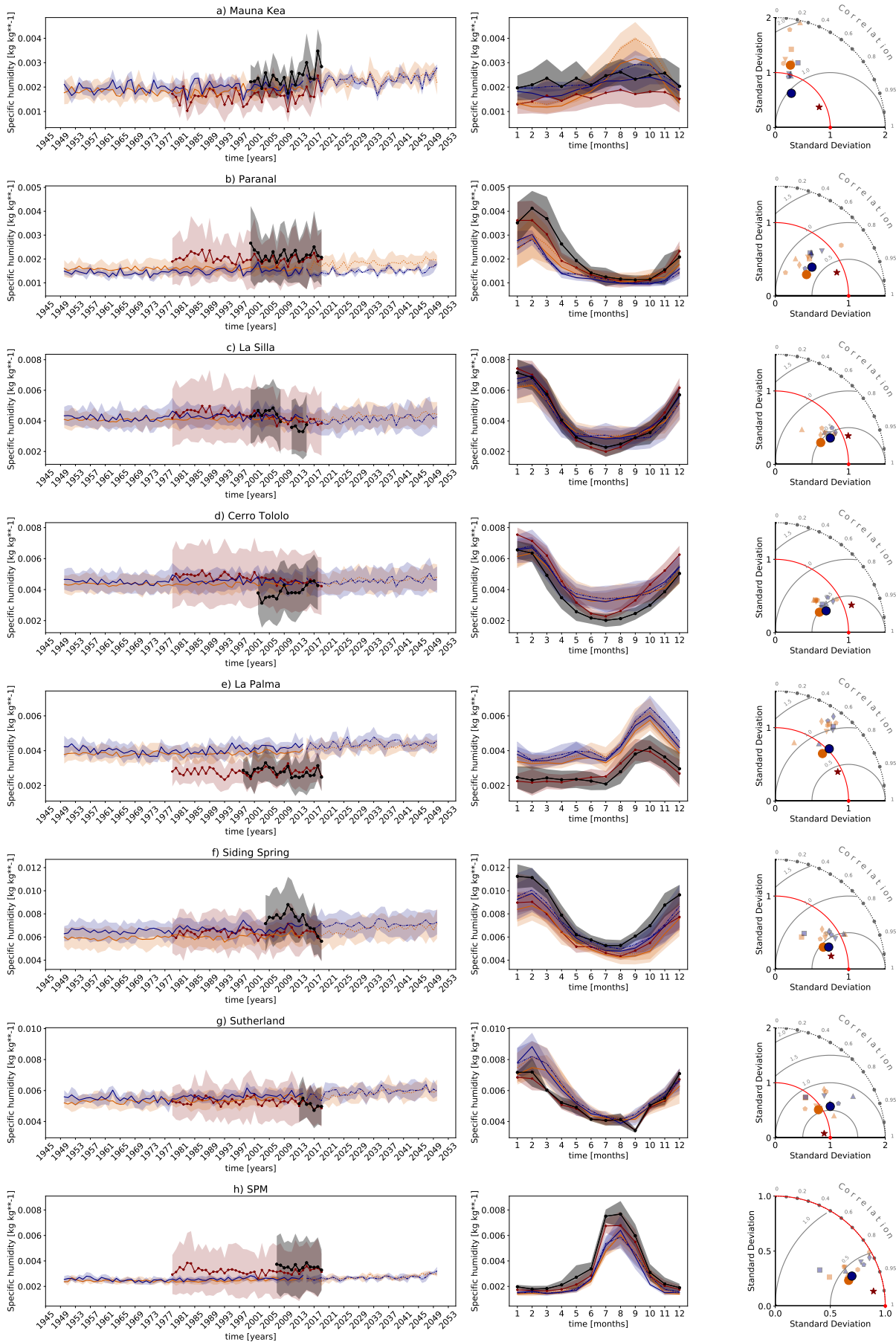


Fig. 6. Same as Fig. 4, but for specific humidity (SH).

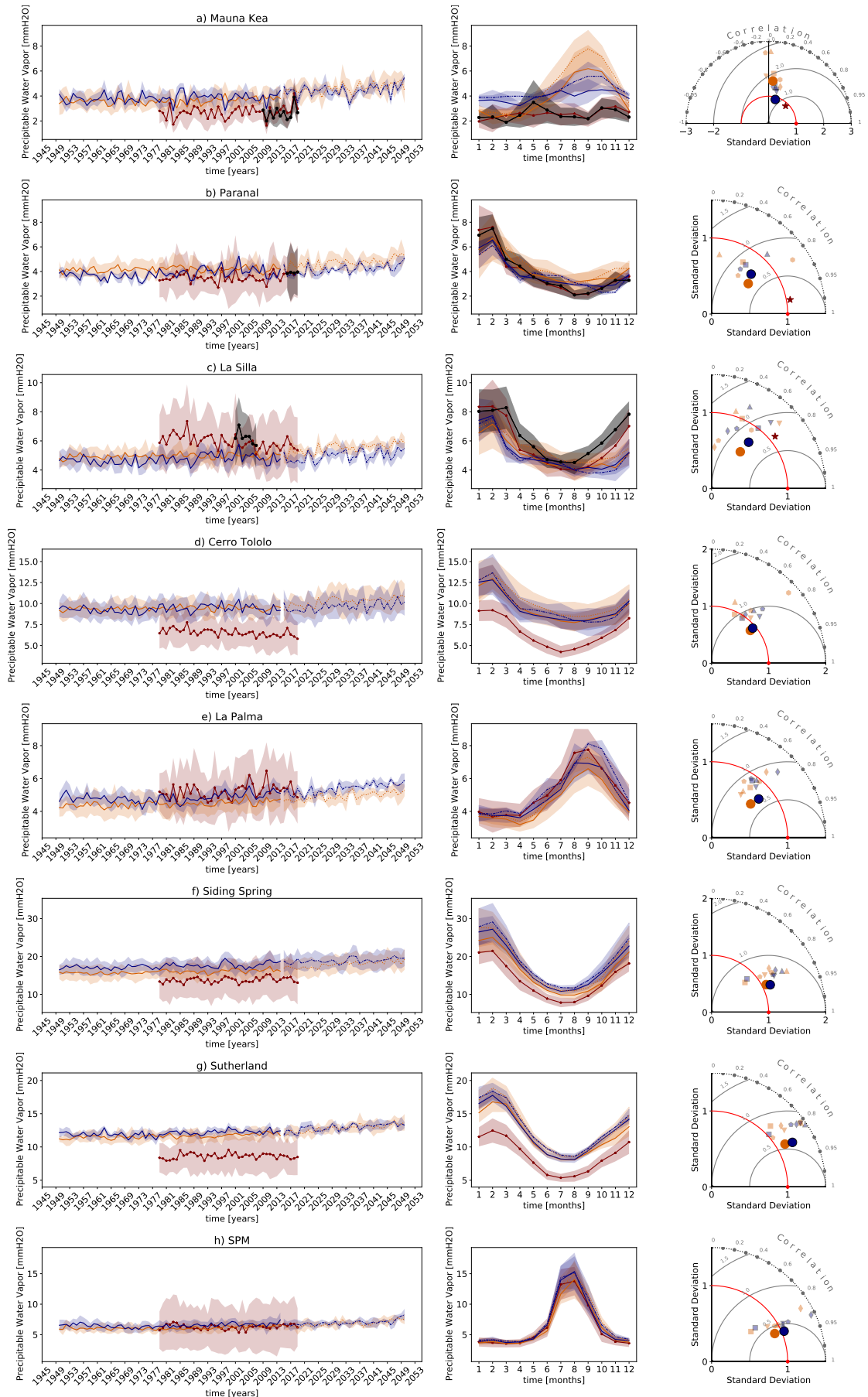


Fig. 7. Same as Fig. 4, but for PWV. For Paranal, the future simulations are taken for the seasonal cycle and the Taylor diagram since in situ data starts only in 2016.

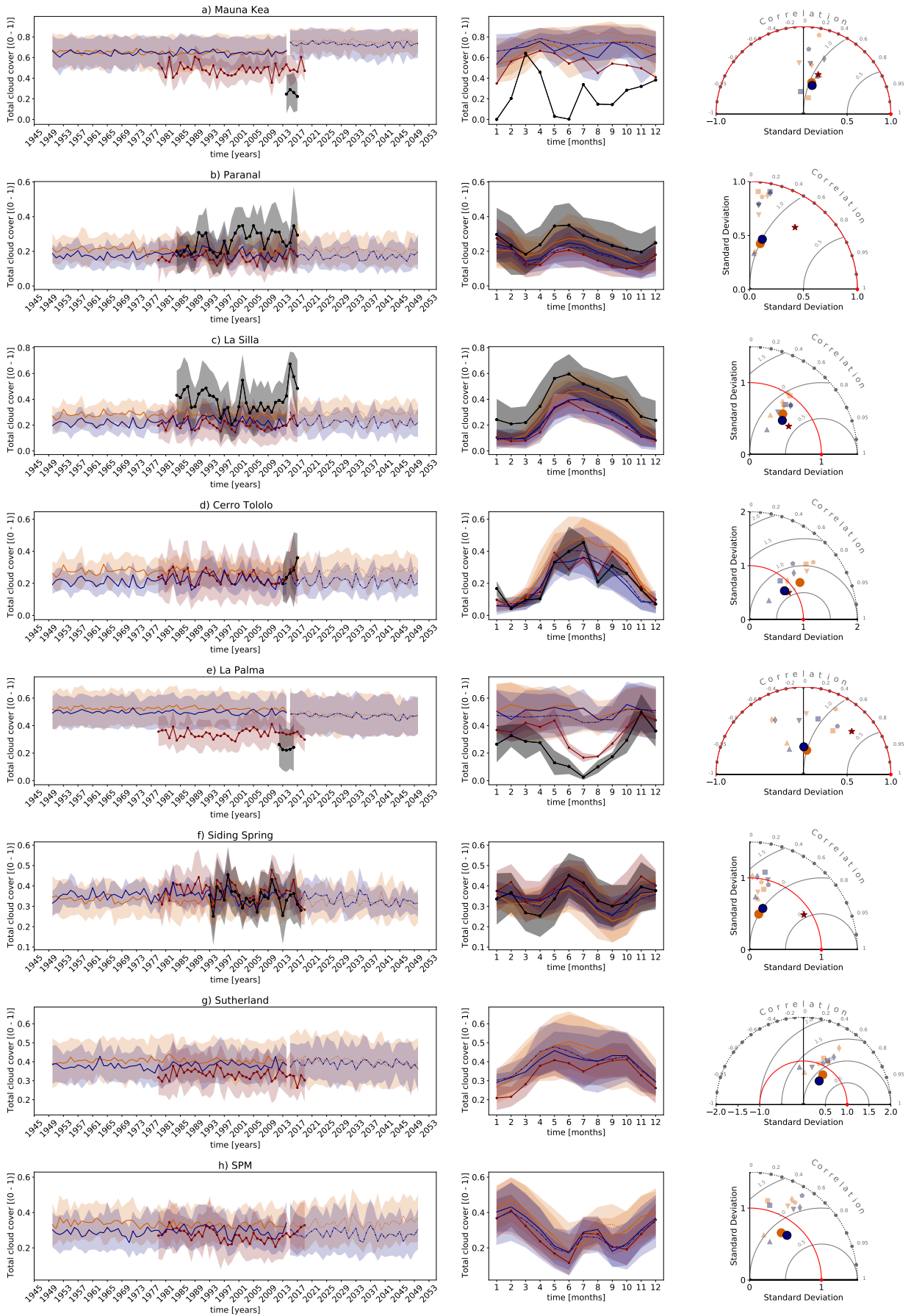


Fig. 8. Same as Fig. 4, but for total cloud cover.

Mauna Kea is the site that exhibits the lowest skill score with a poor value of 0.13 (Fig. 11i and Table A.5). We only considered the year 2014 for comparison for Mauna Kea (Fig. 8a). However, for Cerro Tololo, where in situ data also comes from the Gemini observatory (from Gemini South on Cerro Pachon), we compare two years (2013, 2014), but we get a good match between ERA5 and in situ data (Fig. 11i and Table A.5). Therefore, the poor skill score is likely not due to unreliable in situ data, but must be caused by an inadequate representation of cloud cover in ERA5 for the grid point nearest to the observatory, the great distance to the continent and microclimatic conditions. Tests (not shown) by shifting the selected ERA5 grid point to the highest elevated grid point on Hawaii (Fig. C.1g) only increased the skill score from 0.13 to 0.16, while it increases from 0.13 to 0.21 if only the “high cloud cover” (above 6 km) variable is considered. These results still classify as poor (Table 5).

PRIMAVERA versus ERA5. On average, PRIMAVERA shows a mediocre skill score of 0.33 when compared to 24-hour ERA5 data (Fig. 11i and Table A.5). This makes cloud cover the variable with the lowest average skill score and highlights the difficulty in simulating cloud cover, which is essentially due to the coarse grid resolution that does not enable the explicit representation of cloud processes (Bony et al. 2015). The site with the highest skill score is Cerro Tololo with a good skill score of 0.69, which is as high as the skill score of ERA5 (Fig. 11i and Table A.5). The skill scores of PRIMAVERA are lowest for La Palma (Fig. 11i and Table A.5). We already discussed possible reasons in the first paragraph of this section.

6.1.6. Seeing

Seasonal cycle and time series. The Taylor diagrams (Figs. 9 and 10) show a separation between the 200-hPa-wind-speed seeing and the seeing model, caused by a generally lower standard deviation of the seeing model. The extremely poor match of ERA5 seeing model versus in situ data of La Palma (Fig. 11m) stems from a misrepresentation of the seasonal cycle (Fig. 9e): ERA5 shows a maximum in July and August, whereas in situ data shows a maximum in February. This might be explained with the in situ variability of the temperature inversion layer that leads to best seeing values measured in July and August due to a stronger and lower than usual temperature inversion layer caused by strong trade winds (Muñoz-Tuñón et al. 1997). This enhances the suppression of local convection which leads to clouds below the temperature inversion. Now, since the seeing model integrates well below the inversion layer, it does not take this seasonal phenomenon into account. The opposite is the case: we see an enhanced contribution from the surface winds compared to other sites (Fig. E.1). Also, no individual PRIMAVERA model shows a better match than the ensemble means (Fig. 9e, Taylor diagram). Based on the slightly better matching 200-hPa-wind-speed seeing data sets for La Palma (Figs. 9e and 10e, Taylor diagram, and Figs. 11k and m), we conclude that discrepancies for the seeing model arise from an inadequate representation of the island in the models (Sect. 5.2.4 and Fig. C.5) and a dominant local in situ turbulence ground layer that is poorly linked to global climate. In fact, seeing on La Palma is mostly defined by surface winds which are highly perturbed by the nearby caldera (Vernin & Muñoz-Tuñón 1994).

To explain differences between the 200-hPa-wind-speed and the seeing model approaches, we investigate the vertical profile of the refractive index structure constant C_n^2 (Sect. 2.5) computed for ERA5 and PRIMAVERA following Eq. (13) (Fig. E.1). We would expect the vertical profile of C_n^2 to show a dominant

contribution from the ground layer and varying smaller contributions from the free atmosphere (Osborn & Sarazin 2018). However, Osborn & Sarazin (2018) highlight the importance of considering more than 100 vertical levels for the computation of C_n^2 , while 19 vertical pressure levels are provided for PRIMAVERA and 37 for ERA5. Especially the ground layer turbulence cannot be represented adequately with such a small number of vertical levels (Osborn & Sarazin 2018), and even fine details in topography can have a significant influence on ground layer turbulence (Osborn & Sarazin 2018; Teare et al. 2000). Therefore, it is unsurprising that the C_n^2 profiles of PRIMAVERA and ERA5 do not show a dominant contribution of the median ground layer turbulence (Fig. E.1). To gain deeper insights, we investigate the vertical profiles of northward and eastward wind components of PRIMAVERA and ERA5 (Figs. G.1–G.8). In both ERA5 and PRIMAVERA, the eastward wind shows a global maximum at 200 hPa reaching 20 m s⁻¹ for all sites, which is caused by the subtropical jet stream (Figs. G.1–G.8). This similarity among sites is expected from Abahamid et al. (2004, Sect. 5.2.3). The northward wind shows a different profile for each site. Absolute northward winds do not exceed 10 ms⁻¹, which indicates that the eastward wind component (jet stream) is the dominant contribution at 200 hPa (Figs. G.1–G.8). We conclude that the spatial variability in the vertical C_n^2 profile is dominated by local variabilities in the northward wind speed and the ground layer contribution and temporal variations of the eastward wind speed.

Furthermore, there might be a bias in in situ measurements due to challenges in measuring the seeing. For example at Paranal observatory, it might be that the positive trend in seeing seen in Fig. 9b is influenced by the construction of various telescopes between 1994 and 2003, which could have led to increased surface turbulence (Cantalloube et al. 2020; Sarazin et al. 2008). Moreover, if we compare the more sophisticated DIMM-MASS seeing values measured at Paranal observatory between 2016 and 2019 to ERA5, the skill score of the seeing model would increase from 0.26 (poor) to 0.49 (mediocre; not shown). However, the better skill is likely not only due to the different measurement method with the combination of DIMM and MASS, but also increased by the new location of the DIMM-MASS instrument, which was mounted at a location undisturbed by telescopes (Cantalloube et al. 2020).

ERA5 versus in situ. On average, ERA5 has a poor skill score of 0.20 for the 200-hPa-wind-speed seeing and a poor skill score of 0.28 for the seeing model (Figs. 11k and m and Tables A.6 and A.7). Likely due to local effects, ERA5 is not representing in situ seeing measurements well. The site with the highest skill score for the 200-hPa-wind-speed seeing is Mauna Kea with a mediocre skill score of 0.39 (Fig. 11k and Table A.6). Mauna Kea is also the site with the lowest seeing values and the highest altitude. The C_n^2 profile (Fig. E.1) reveals that ERA5 shows higher C_n^2 values at the jet stream level for Paranal and La Silla compared to Mauna Kea. This goes well together with the fact that the jet stream is less prominent at Mauna Kea compared to Paranal and La Silla, which are closer to -30° latitude. Therefore, we conclude that very low seeing values are better fit with ERA5. The site with the lowest skill score for the 200-hPa-wind-speed seeing is Paranal with a poor skill score of 0.06 (Fig. 11k and Table A.6). If we compare ERA5 to MASS-DIMM data from Paranal, we get an equally poor skill score of 0.07 (not shown). We conclude that the ERA5 200-hPa-wind-speed seeing is not suitable to represent in situ seeing data. As we elaborated in the previous paragraph, the vertical profile justifies the 200-hPa-wind-speed seeing approach. However, since the 200-hPa-wind-speed seeing does not account for the ground layer but

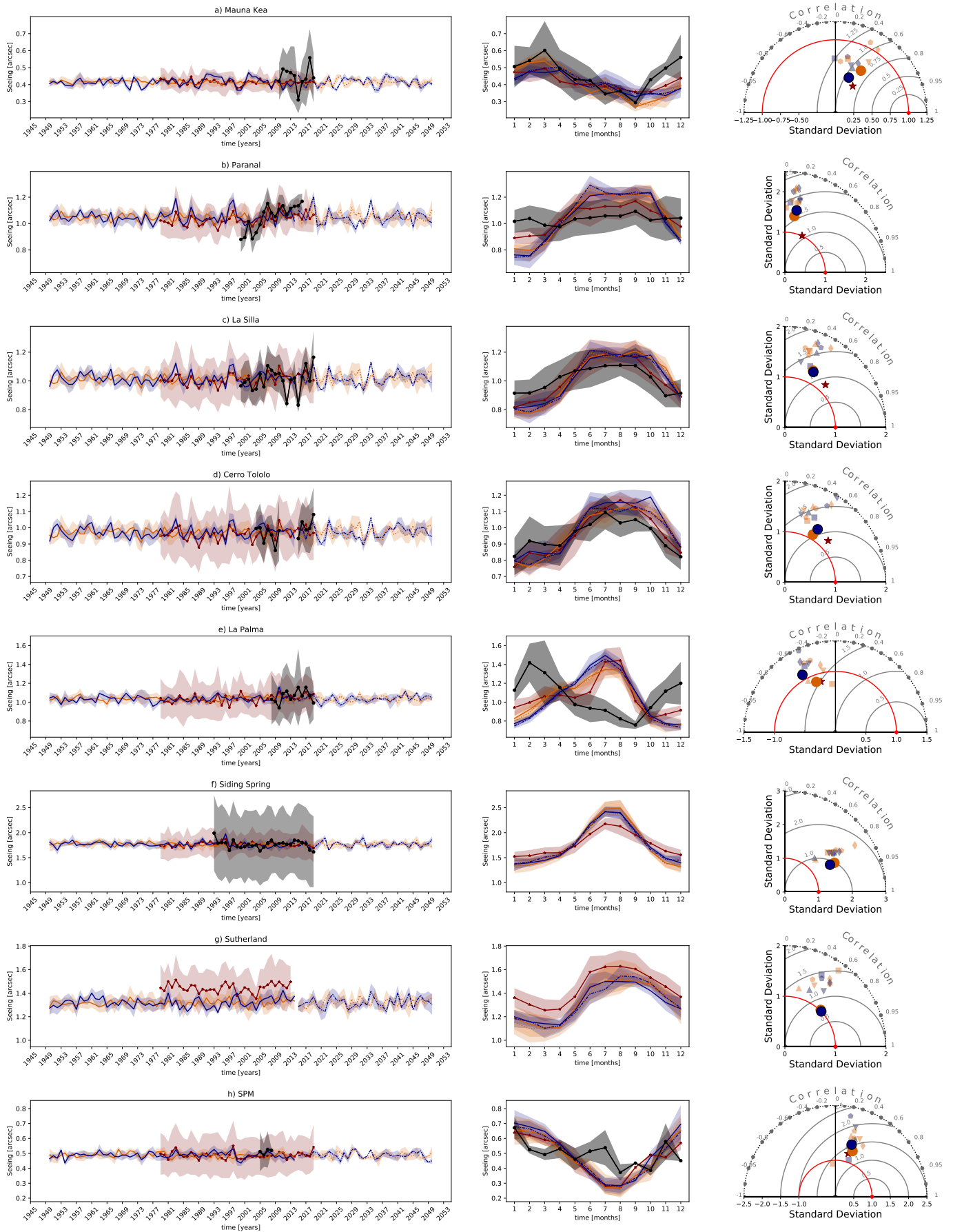


Fig. 9. Same as Fig. 4, but for the seeing model. For Siding Spring (*f*), only yearly in situ data are available, which is therefore neither included in the seasonal cycle nor in the Taylor diagram.

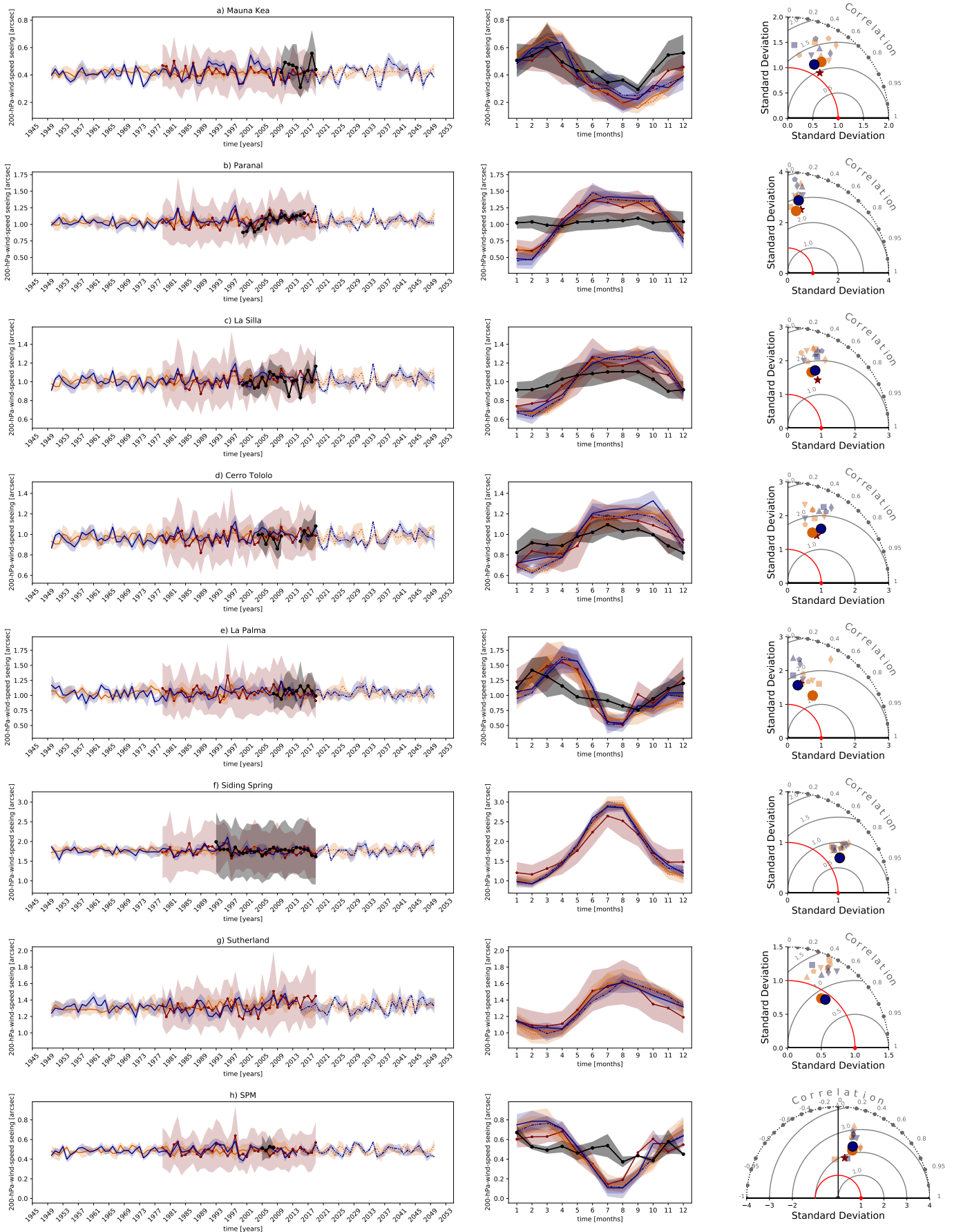


Fig. 10. Same as Fig. 4, but for the 200-hPa-wind-speed seeing. For Siding Spring (*f*), only yearly in situ data are available, which is therefore neither included in the seasonal cycle nor in the Taylor diagram.

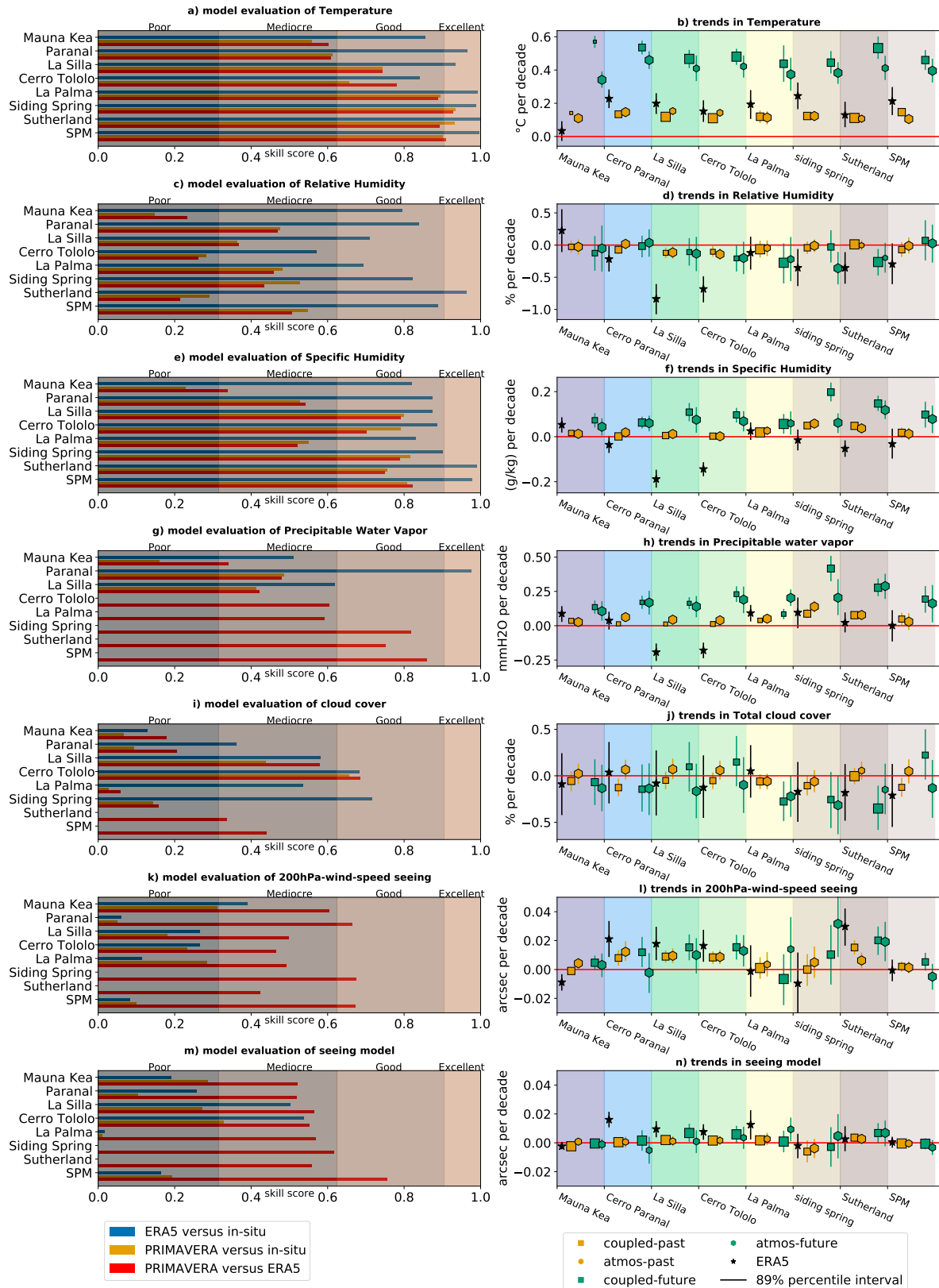


Fig. 11. Skill scores (*left*) and trends (*right*) for all sites and all variables: temperature (*a, b*), relative humidity (*c, d*), specific humidity (*e, f*), PWV (*g, h*), total cloud cover (*i, j*), 200 hPa wind-speed seeing (*k, l*) and seeing model (*m, n*). Skill scores (Sect. 5.2.5) are shown with their classification (Table 5) for the time-intersecting ERA5 versus in situ (blue), time-intersecting PRIMAVERA versus in situ (dark-yellow) and PRIMAVERA versus ERA5 (1979–2014, red). Trends (in units per decade) are shown for ERA5 (1979–2019, black stars), PRIMAVERA historical (1950–2014, dark-yellow markers) and PRIMAVERA future simulations (2015–2050, green markers). Coupled and atmosphere-only simulations are shown with squares and diamonds, respectively. The larger the markers, the higher the skill scores. The error bars show the 89th percentile credible interval of the Bayesian analysis (Sect. 5.3). The data used to produce this figure can be found in Tables A.1–A.14.

is an approximation for the free atmosphere seeing, we would need to compare it to in situ measurements of the free atmosphere. The closest we have is the MASS data of Mauna Kea and SPM. Together with the high elevation of the Mauna Kea observatory, the exclusion of the surface layer by MASS might explain why Mauna Kea is the only site that exhibits a mediocre skill score. On the other hand, it does not explain the poor skill score of SPM. The site with the highest skill score for the seeing model is Cerro Tololo with a mediocre skill score of 0.54 (Fig. 11m and Table A.7). The site with the lowest skill score for the seeing model is La Palma with a poor skill score of 0.02 (Fig. 11m and Table A.7). We already discussed possible explanations for this poor skill score in the previous paragraph.

PRIMAVERA versus ERA5. On average, PRIMAVERA shows a mediocre skill score of 0.56 for the 200-hPa-wind-speed seeing and a mediocre skill score of 0.58 for the seeing model (Fig. 11k and m and Tables A.6 and A.7). The site with the highest skill score for the 200-hPa-wind-speed seeing is Siding Spring with a good skill score of 0.68 (Fig. 11k and Table A.6). The site with the lowest skill score for the 200-hPa-wind-speed seeing is Sutherland with a mediocre skill score of 0.42 (Fig. 11k and Table A.6). The site with the highest skill score for the seeing model is SPM with a good skill score of 0.75 (Fig. 11m and Table A.7). The sites with the lowest skill score for the seeing model are Mauna Kea and Paranal both with a mediocre skill score of 0.52 (Fig. 11m and Table A.7). For both the 200-hPa-wind-speed seeing and the seeing model approach and in contrast to the other studied variables, ERA5 shows lower skill scores on average than PRIMAVERA. This means that PRIMAVERA agrees better with ERA5 than ERA5 represents in situ data. On the one hand, this is due to the low vertical resolution of ERA5, on the other hand, it indicates the challenge of measuring seeing reliably.

6.2. Discussion of adequacy of methods

6.2.1. In situ data

As shown in this study, in situ observations notably allow a thorough evaluation of reanalyses and climate models, which is an important step to assess the trustworthiness of future climate projections. We want to emphasise that it is advantageous for astronomical sites to archive continuous and high-quality in situ observations, with which it is possible to monitor ongoing changes and analyse past trends in site conditions. We found that too few observatories maintain such an archive that is easily accessible.

6.2.2. ERA5

The ERA5 reanalysis performs well over most sites. The skill scores of ERA5 versus in situ are always higher than the PRIMAVERA versus ERA5 skill scores, except for seeing. The good agreement of ERA5 with in situ data shows that ERA5 is a suitable data set for analysing past climate and validating GCMs. However, missing or incongruous in situ data, for example down-time data instead of cloud cover measurements or missing monthly seeing data for Siding Spring, complicate an exhaustive evaluation of ERA5.

6.2.3. PRIMAVERA

It would be possible to improve the reliability of PRIMAVERA projections by considering a weighted multi-model mean based on the skill scores of each model (Knutti et al. 2010). This is

not done in our approach, in which all GCMs are given the same weight. More tests and assessments whether a weighted approach performs better together with a more robust skill metric would be highly recommended (Knutti et al. 2010).

The PRIMAVERA multi-model ensemble mean performs well for most sites and variables. The PRIMAVERA skill scores averaged over all sites classify as good for temperature and specific humidity and as mediocre for relative humidity, PWV, cloud cover and the two seeing approaches. The atmosphere-only simulations show higher skill scores than the coupled simulations in 71% of all cases. This result is expected since atmosphere-only simulations are constrained by SST measurements. The differences are, however, only small.

The site that shows the best agreement between PRIMAVERA and ERA5 throughout all variables is SPM with an average skill score of 0.71. The site that shows the worst agreement of PRIMAVERA versus ERA5 is Mauna Kea with an average of 0.40. Mauna Kea is located on the Big Island of Hawaii with steep orography that is not well represented in most PRIMAVERA GCMs because of their still too low resolution. The model orography is rather smooth and exhibits differences in elevation greater than 2000 m compared to observations. Moreover, Big Island lies in the middle of the North Pacific Ocean with a distance of approximately 4000 km to America and more than 7000 km to Australia. The difficulty of GCMs to simulate climate on the island of Hawaii could also be due to the challenging simulation of the boundary trade wind layer regime (Wyant et al. 2010). To address this question, we have compared PRIMAVERA data to the 3 km horizontal grid spacing Hawaii Regional Climate Model (HRCM), based on the Weather Research and Forecasting (WRF) model V3.3 (Zhang et al. 2012). We find that the HRCM is able to represent the diurnal cycle of temperature adequately compared to in situ data (not shown). If we average monthly, the skill score of HRCM 2-m temperature against in situ data over Mauna Kea is not improved compared to PRIMAVERA (not shown). However, since PRIMAVERA has already a good skill score for temperature, monthly means of other variables could still improve a lot with a regional model such as the HRCM. Furthermore, the microclimates of Hawaii (Giambelluca & Schroeder 1986) highlight the need for a high-resolution model that is able to represent such fine geographical variations. The PRIMAVERA GCMs, with their horizontal grid spacing of 25–50 km, are still too coarse to represent all aspects of the climate on Mauna Kea realistically.

It becomes clear from our results that GCMs coming from the CMIP intercomparisons are not suitable to project changes in all studied sites due to their too coarse grids (i.e. 80–200 km for CMIP6 GCMs; IPCC 2021) and the associated inadequate representation of orography, since we found that PRIMAVERA high-resolution GCMs already show weaknesses in some sites. This is especially true for islands such as Hawaii or La Palma that are not represented in CMIP models (not shown).

7. PRIMAVERA GCMs climate projections

It is important to keep in mind that PRIMAVERA future simulations are run under the highest GHG emission scenario SSP5–8.5 (Kriegler et al. 2017). In the following sections, we look at trends in unit per decade. All results are presented visually in Fig. 11.

7.1. Trends in temperature

All ERA5 and PRIMAVERA historical and future simulations show positive trends in temperature and only Mauna Kea includes the zero-trend line within the 89% credible interval

(Fig. 11b). For all sites except Mauna Kea, ERA5 shows slightly higher trends of the order of 0.1 °C per decade for the period 1979–2019 than historical PRIMAVERA ensemble means for the period 1950–2014 (Fig. 11b). This result is attributed to the different time periods evaluated since anthropogenic climate change likely accelerates the rate of temperature increase with time (Haustein et al. 2017). The PRIMAVERA coupled and atmosphere-only historical simulations show an average trend of $0.13 \pm 0.01\text{ °C}$ per decade, which is similar to the average trend of ERA5 of $0.17 \pm 0.07\text{ °C}$ per decade (Table A.8). This agreement increases the reliability of these results.

One explanation of the weak positive trend in ERA5 data for Mauna Kea could be that in free air at the level of Mauna Kea, a decreasing trend in temperature has been observed in the period 1987–2017 (Kagawa-Viviani & Giambelluca 2020). In situ data shows, however, that the surface temperature at the elevation of Mauna Kea has risen faster than the surface temperature at sea level, resulting in a more stable atmosphere (Giambelluca et al. 2008), which goes well together with a more frequent trade wind inversion (Cao et al. 2007). Since PRIMAVERA data are showing a higher trend than ERA5, the ERA5 trend must be impacted by local atmospheric effects and topography not seen by PRIMAVERA.

PRIMAVERA projects an average increase in future temperature of $0.49 \pm 0.05\text{ °C}$ per decade until 2050 for coupled simulations and of $0.40 \pm 0.03\text{ °C}$ per decade for atmosphere-only simulations for the selected sites (Table A.8). These certain and positive trends in temperature are consistent with global mean surface temperature trends (IPCC 2021).

7.2. Trends in relative humidity

In relative humidity, ERA5 shows a negative average trend with a large uncertainty of $-0.33 \pm 0.33\%$ per decade (Fig. 11d and Table A.9). The PRIMAVERA historical simulations do not show any significant trend, with an average of $-0.05 \pm 0.05\%$ per decade (Table A.9). Largest differences between PRIMAVERA and ERA5 are found for La Silla and Cerro Tololo (Fig. 11d and Table A.9). However, this would not change if other close grid points were considered, since in PRIMAVERA and ERA5, between -70° and -71.5° longitude and between -31.0° and -29.5° latitude, trends are similar to the trend in the nearest grid point (not shown). For ERA5, the more we go away from the coast, the faster relative humidity decreases.

Similarly, PRIMAVERA projects either no change or a slight decrease in future relative humidity with an average of $-0.13 \pm 0.12\%$ per decade until 2050 for coupled simulations and $-0.11 \pm 0.10\%$ per decade for atmosphere-only simulations for the selected sites (Fig. 11d and Table A.9). Since PRIMAVERA shows poor to mediocre agreement with ERA5 and in situ data for the historical period (Fig. 11c), PRIMAVERA projections of changes in relative humidity over the sites may not be reliable. Nevertheless, the trends found are in agreement with the literature. Byrne & O’Gorman (2018) found a slight absolute decrease of $-0.22 \pm 0.20\%$ per decade from 1979 to 2016 within 40°N and 40°S , using the HadISDH dataset (Willett et al. 2014). Moreover, the IPCC Fifth Assessment Report (Collins et al. 2013) states with medium confidence that small reductions in relative humidity over land are likely and linked to faster rising temperatures over land than over the ocean. More specifically, as saturated oceanic air moves and warms over the warmer land, further moistening of the air over land (i.e. through evapotranspiration) is insufficient to maintain constant relative humidity

which therefore decreases. The larger the land-sea temperature gradient, the lower the relative humidity over land.

7.3. Trends in specific humidity

Trends in ERA5 specific humidity are negative for all sites except Mauna Kea and La Palma (Fig. 11f and Table A.10). The positive trend for La Palma might be linked to enhanced evaporation over the ocean caused by ocean warming, which can lead to moister air in ERA5, which mixes values above land and above ocean due to the size of the island of La Palma. The positive trend on Mauna Kea is more delicate and turns negative for one grid-point to the east (not shown). Furthermore, the trend gets slightly negative for the highest gridpoint in ERA5 (one gridpoint to the south of the nearest gridpoint in Fig. C.1). This reflects again the difficulty of ERA5 in representing Mauna Kea adequately. Over all sites, ERA5 finds an average trend in specific humidity of $-0.05 \pm 0.08\text{ g kg}^{-1}$ per decade (Fig. 11f and Table A.10). Historical PRIMAVERA coupled and atmosphere-only simulations project positive trends for all sites with an average of $0.02 \pm 0.02\text{ g kg}^{-1}$ per decade (Fig. 11f and Table A.10). As it was the case for relative humidity, we again find the largest differences in trends between ERA5 and PRIMAVERA historical simulations for Cerro Tololo and La Silla (Fig. 11f and Table A.10).

Specific humidity is projected to increase slightly over all sites by $0.10 \pm 0.04\text{ g kg}^{-1}$ per decade on average until 2050 in the PRIMAVERA coupled simulations and by $0.08 \pm 0.03\text{ g kg}^{-1}$ in the atmosphere-only simulations (Fig. 11f and Table A.10). The median vertical profile of specific humidity shows an increase of specific humidity by future PRIMAVERA simulations over the entire atmospheric column (Fig. F.1).

7.4. Trends in PWV

As shown in Eq. (3), PWV depends on specific humidity that is integrated over the entire atmospheric column. Trends in PWV are therefore similar to trends in specific humidity (Fig. 11h). ERA5 shows an average trend in PWV of $0.00 \pm 0.12\text{ mmH}_2\text{O}$ per decade, while historical PRIMAVERA simulates an average trend of $0.05 \pm 0.03\text{ mmH}_2\text{O}$ per decade (Fig. 11h and Table A.11). As for specific and relative humidity, the largest disagreements between ERA5 and historical PRIMAVERA trends are for La Silla and Cerro Tololo (Fig. 11h and Table A.11). For Paranal, Siding Spring, Sutherland and SPM, ERA5 and historical PRIMAVERA trends agree well (Fig. 11h and Table A.11). These historical trends of increasing PWV are in line with Trenberth et al. (2005) who found positive trends over land over the period 1988–2003 in several reanalyses. Hellemeier et al. (2019), however, found no significant trends between 1967 and 2001 in Cerro Paranal, La Palma, Mauna Kea and Siding Spring. These results are based on the lower resolution ERA-40 reanalysis ($2.5^\circ \times 2.5^\circ$ horizontal grid; Källberg et al. 2004) and consider an older time period than that used in our study.

Future trends in PWV projected by PRIMAVERA are positive for all sites. The PRIMAVERA simulations project an average increase in PWV by $0.21 \pm 0.10\text{ mmH}_2\text{O}$ until 2050 for coupled and by $0.18 \pm 0.06\text{ mmH}_2\text{O}$ for atmosphere-only simulations (Fig. 11h and Table A.11). An increasing PWV is in line with globally increasing total column water (IPCC 2021). In opposition to our results, Cantalloube et al. (2020) suspect that the area around Cerro Paranal will become drier, however, these hypotheses are based on low resolution GCMs.

7.5. Trends in cloud cover

ERA5 shows an average trend in total cloud cover of $-0.10 \pm 0.10\%$ per decade in all sites (Fig. 11j and Table A.12). However, the credible intervals (black error bars in Fig. 11j) are large and these trends are therefore not reliable, which is also the case for PRIMAVERA. On average, historical coupled PRIMAVERA simulates a negative trend of $-0.07 \pm 0.04\%$ per decade while atmosphere-only PRIMAVERA simulates a trend of $0.03 \pm 0.05\%$ per decade (Fig. 11j and Table A.12). These opposite results between coupled and atmosphere-only simulations and the large uncertainty range highlight the difficulty in simulating cloud cover. In contrast, Hellemeier et al. (2019) find a significant positive trend of $+1.6\%$ per decade in cloud cover for the period between 1967 and 2001 with ERA-40 for Cerro Pachon located near Cerro Tololo. This result is not in line with the negative trend of -0.12% per decade in ERA5 for Cerro Tololo, and could again be attributed to the coarser resolution of ERA-40 in comparison to ERA5 as well as to the different time period.

Future trends in cloud cover projected by PRIMAVERA amount on average to $-0.08 \pm 0.21\%$ per decade until 2050 in the coupled simulations and $-0.17 \pm 0.07\%$ in the atmosphere-only simulations, while almost all uncertainty intervals include the possibility of no trends in cloud cover (Fig. 11j and Table A.12). These results need to be considered with caution for several reasons. First, PRIMAVERA has poor and mediocre skill scores compared with ERA5 (Fig. 11i), so PRIMAVERA projections may not be trustworthy. Second, most GCMs, including PRIMAVERA, do not use grids that are fine enough to represent small-scale cloud processes. They instead rely on sub-grid scale parameterisations that are believed to cause most of the model discrepancies (Sect. 2.4). Still, PRIMAVERA uses resolutions that are high enough to represent the dynamical processes associated with cloud formation, such as the location and magnitude of upward and downward motions associated with frontal systems and orography (Haarsma et al. 2016). Their projections may therefore be more trustworthy than conventional IPCC GCMs and the projections of no trend are in line with the results of no significant trends in ERA5 and historical PRIMAVERA.

7.6. Trends in seeing

ERA5 shows an average trend in seeing of 0.01 ± 0.02 arcsec per decade for the 200-hPa-wind-speed seeing and 0.01 ± 0.01 arcsec per decade for the seeing model (Figs. 11l and n; Tables A.14 and A.13). Similarly, historical PRIMAVERA simulates an average trend of 0.01 ± 0.01 arcsec per decade by coupled and 0.01 ± 0.00 arcsec per decade by atmosphere-only simulations for the 200-hPa-wind-speed seeing, and an average trend of 0.00 ± 0.00 arcsec per decade for both the coupled and the atmosphere-only simulations for the seeing model (Figs. 11l and n; Tables A.14 and A.13). Over most sites, the 200-hPa-wind-speed seeing shows higher trends than the seeing model for PRIMAVERA and ERA5, but the uncertainty is also larger, which is attributed to the higher variability of the modelled 200-hPa-wind-speed seeing (Figs. 9 and 10). For La Palma, Hellemeier et al. (2019) find a significant decrease in 200-hPa-wind-speed of -0.6 m s^{-1} per decade for the period between 1967 and 2001 with ERA-40. This trend would lead to a decrease in the seeing value, which we also find with ERA5 for La Palma, although what we find is not significant.

PRIMAVERA projects an average future trend of 0.01 ± 0.01 arcsec per decade until 2050 for coupled and atmosphere-only simulations for the 200-hPa-wind-speed seeing (Fig. 11l and Table A.14). For the seeing model, PRIMAVERA projects an average trend of 0.00 ± 0.00 arcsec per decade for coupled simulations and 0.00 ± 0.01 arcsec per decade for atmosphere-only simulations (Fig. 11n and Table A.13). These projected trends represent a very small change for astronomers. However, as mentioned above, the ground layer contribution to the seeing is not represented adequately in PRIMAVERA GCMs, which is locally highly variable among sites. It is therefore not possible to project changes of the ground layer turbulence into the future with PRIMAVERA. Still, the agreement of both approaches on the magnitude of future trends in astronomical seeing establishes confidence that the free atmosphere seeing is not impacted severely by climate change. These results need to be considered with caution because of the poor or mediocre skill scores between PRIMAVERA and in situ data. However, slight positive future trends in 200-hPa-wind-speed seeing could be linked to a strengthening of the subtropical northern and southern winter jet stream, which is already observable (Pena-Ortiz et al. 2013; IPCC 2021).

7.7. Discussion of PRIMAVERA future trends and possible impacts

Our analysis of future trends projected by PRIMAVERA shows that astronomers observing at major astronomical observatories at any of the selected sites will likely experience an increase in temperature, specific humidity and PWV by 2050. These trends will likely increase the loss in observing time, but each telescope site must evaluate the severity of these impacts for itself.

Rising temperatures must be considered for planning and building next-generation telescopes. For example, at the Paranal Observatory, the dome cooling system, which prevents air turbulences inside the dome (dome seeing; Tallis et al. 2020), is built for a maximum surface air temperature of $16 \text{ }^\circ\text{C}$ (Cantalloube et al. 2020).

Our results for specific humidity are coherent with those of temperature and relative humidity. According to the Clausius–Clapeyron relationship, a warmer atmosphere is associated with an exponential increase in atmospheric saturated vapour pressure of about 7% for each 1 K warming (Allen & Ingram 2002; Held & Soden 2006). In addition, a rise in temperature enhances potential evaporation, which can lead to an increase in specific humidity but would keep relative humidity almost constant. This is consistent with the increase in temperature and specific humidity and the slight decrease in relative humidity projected by PRIMAVERA. Consequences for telescopes include a higher risk of condensation due to an increased dew point. For example, the William Herschel telescope at the ORM on La Palma cannot operate when the mirror temperature is only $2 \text{ }^\circ\text{C}$ or less above dew point²⁶.

Positive trends in PWV are likely leading to a significant decrease in the fraction of the night with excellent PWV conditions in the next 30 yr. For example, Paranal currently exhibits PWV values below 1 mm in 8.2% of the night (Sect. 3.2.1). By 2050, PWV is projected by PRIMAVERA to rise by $0.51^{+0.24}_{-0.27}$ mm. If we assume a homogeneous shift over the distribution of values between 0.0 mm and 1.5 mm, then PWV values will be below $1.51^{+0.24}_{-0.27}$ mm in 8.2% of the night by 2050, which

²⁶ https://www.ing.iac.es//Astronomy/tonotes/misc/weather_safe.html, accessed on 2021 Oct. 14.

will decrease the time with excellent observing conditions. If we further assume linearity between 0 mm and 1.5 mm, then PWV values will be below 1 mm in only $2.0^{+2.4}_{-1.5}\%$ of the night by 2050. A similar picture is shown for La Palma, for which a rise in PWV of 0.20 ± 0.06 mm per decade is projected by the atmosphere-only simulations of PRIMAVERA. Given the measurement of [Castro-Almazán et al. \(2016, Sect. 3.3\)](#), this would result in PWV values below 2.30 ± 0.18 mm in 20% of the time, or assuming linearity, this would lead to PWV values below 1.70 mm in $8.5^{+3.0}_{-2.6}\%$ of the night-time by 2050. However, this needs to be further investigated by assuming a more realistic distribution instead of a linear approximation. For this conclusion, hourly values of GCM projections would be beneficial. Especially for high contrast imaging, a field for which the construction of next-generation telescopes is crucial for emerging ([Meyer et al. 2018](#)), an increase in PWV will likely worsen observing conditions and climate change has to be considered urgently for site selection.

No major changes are projected for seeing, relative humidity, or cloud cover. These results need to be considered with caution due to the low skill score of PRIMAVERA and ERA5 compared to in situ data. Furthermore, trends in seeing are hard to project since the ground layer turbulence contributing to the seeing is extremely sensitive to topography in the proximity of the telescope measuring the seeing ([Osborn & Sarazin 2018](#); [Teare et al. 2000](#)). Still, the rise in surface temperature will likely increase the ground layer contribution of the seeing ([Cantalloube et al. 2020](#)) and the strengthening of the jet stream will likely increase the upper contribution to the seeing. The projected poleward shift would weaken or increase the turbulence contribution of the jet stream for each site individually, dependent on their latitude.

Hypothetically, prolonging the PRIMAVERA atmosphere-only projections to the year 2100 by multiplying the trend per decade by 8 (since there are 8 decades between 2020 and 2100), we would end up with an average increase in temperature of 3.94 ± 0.40 °C, an average increase in specific humidity of 0.80 ± 0.31 g kg⁻¹ and an average increase in PWV of 1.67 ± 0.81 mmH₂O. Certainly, these changes cannot be neglected in the planning and operation of next-generation telescopes. Especially for telescopes observing in the infrared, an average increase of 1.67 ± 0.81 mmH₂O would unmistakably reduce the amount of observing time with excellent conditions, which is defined by PWV values below or equal to 3 mmH₂O ([Kidger et al. 1998](#)). Indeed, the SSP5–8.5 scenario ([Kriegler et al. 2017](#)) assumes the worst case with ever rising CO₂-equivalent emissions to three times the amount of emissions compared with emissions in 2000 ([Riahi et al. 2011](#)), which justifies this prolongation. Nevertheless, high-resolution GCMs running until 2100 would lead to refined estimates.

8. Discussion and conclusions

8.1. Significance of this study

This study shows that the site selection process for next-generation telescopes and the construction and maintenance of astronomical facilities with a typical lifetime of 30 yr requires the consideration of anthropogenic climate change as a potential driver for changes in site characteristics. Nowadays, astronomical observatories are designed to work under the current site conditions and only have a few possibilities for adaptation. Already today, climate change has an influence on every place on Earth ([IPCC 2021](#)).

8.2. ERA5 and PRIMAVERA historical trends

Historical trends obtained with the latest ERA5 reanalysis generally do not agree with the trends found by [Hellemeier et al. \(2019\)](#), using the earlier ERA-40 reanalysis. This result emphasises the influence of model horizontal resolution on long-term trends for a specific site, in addition to other improvements, such as satellite observations and data assimilation techniques ([Hersbach et al. 2020](#)).

On average, ERA5 historical trends are positive within their 89% credible interval for temperature. For relative and specific humidity, PWV, cloud cover, 200-hPa-wind-speed seeing, and the seeing model, the 89% credible interval of ERA5 trends includes the possibility of a zero-trend. However, this gives only a rough overview and we shall conclude that climate change has already had an impact on the temperature for most of the selected eight sites. Trends of other variables have a bigger range over the eight sites and no concurrent conclusion can be drawn.

Agreeing well with ERA5 in general, PRIMAVERA historical trends show a clear increase in temperature. Also, for specific humidity, cloud cover, 200-hPa-wind-speed seeing, and the seeing model, the 89% credible interval of PRIMAVERA historical trends includes the possibility of a zero-trend. For relative humidity, the historical trend analysis shows a slight decrease and for PWV, a slight increase within the 89% credible interval.

8.3. Strengths and weaknesses of PRIMAVERA GCMs

The strength of our approach lies in the use of an ensemble of GCMs that provide a resolution (18–50 km horizontal grid spacing) that is much finer than conventional GCMs used in CMIP5 and CMIP6 for IPCC assessment reports ([IPCC 2013, 2021](#)). This is the first time that such a study has been possible with an ensemble of high-resolution GCMs. We find that high-resolution GCMs such as those developed within the PRIMAVERA project can be used to analyse the impact of climate change on site characteristics of next-generation telescopes. Despite PRIMAVERA showing errors and uncertainties, we can rely on an ensemble mean from six different physics-based and globally consistent climate models that project trends and climate change responses at most of the selected sites. Furthermore, since PRIMAVERA GCMs represent the large-scale circulation well, errors at upper pressure levels are likely less severe than errors at the model surface, which are connected to missing scales of motion for turbulence, a chaotic phenomenon, and misrepresentation of islands and elevation in general.

8.4. Outlook

With these results, we are confident that PRIMAVERA data provide the grounds to assess current and future climate conditions in sites for astronomical observations. Still, the weaknesses of PRIMAVERA to simulate some local aspects of the selected sites, such as those related to topography, show that it would be better to rely on climate models with an even higher resolution to study local aspects of climate change in astronomical sites. One way would be to use climate models with a horizontal grid spacing of a few kilometres that are able to simulate convection explicitly to improve the simulated diurnal cycle of precipitation and precipitable water. These models could be regional climate models, such as those developed within CORDEX ([Giorgi 2019](#)) that could be run over domains that include as many astronomical sites as possible, or global climate models such as those

developed within the Dynamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains (DYAMOND) project (Stevens et al. 2019), but that are – at the moment – still too expensive to run for more than a season. Another very promising effort will be provided through the digital twin of Earth (Bauer et al. 2021), a GCM with a spatial resolution of 3 km that uses data assimilation and the laws of physics to replicate Earth’s climate system and project changes into the future as a contribution to the European Union’s green deal under the project destination Earth²⁷. A promising project under the European Union’s Horizon 2020 research and innovation funding programme is NextGEMS, which aims to provide several global climate models with a spatial horizontal resolution of 2.5 km for the analysis of storms. Such GCMs would probably prove useful for the pre-selection of some of the best future spots for astronomical observations.

We have shown in this study that high-resolution GCMs, such as HighResMIP, are suitable for assessing changes in observing conditions of major telescope sites. Since we have focussed our effort on mean changes and trends, we propose further studies to be investigated: (1) Increasing horizontal resolution in coupled GCMs has also been shown to improve the ocean mean state and variability, such as ENSO (Shaffrey et al. 2009; Roberts et al. 2016). The impact of such modes of variability on telescope sites, as well as their changes in global warming conditions, could therefore be investigated using these models. (2) Synoptic-scale systems are better resolved in higher resolution GCMs, which improves the simulation of cyclones, atmospheric rivers, and extreme precipitation (e.g. Haarsma et al. 2013; Baker et al. 2019; Payne et al. 2020). The variability and intensity of tropical cyclones are also better represented (e.g. Roberts et al. 2015). HighResMIP GCMs would therefore be appropriate to assess the impact of such intense events on astronomy observing conditions. (3) Increasing the temporal resolution to hourly data would allow for an investigation of the diurnal cycle. For example, hourly differences between day- and night-time temperature could be assessed, which might be important because bigger day-to-night differences demand more cooling power to bring down the temperature of the telescope mirror to the ambient night-time temperature.

Acknowledgements. This work has been carried out within the framework of the National Centre of Competence in Research PlanetS supported by the Swiss National Science Foundation under grants 51NF40_182901 and 51NF40_205606. The authors acknowledge the financial support of the SNSF. The PRIMAVERA project is funded by the European Union’s Horizon 2020 programme, grant agreement no. 641727. We thank all the PRIMAVERA climate modelling groups (Met Office, CERFACS, MPI-M, CMCC, ECMWF, KNMI, SMHI, BSC, CNR) for producing and making their model output available. We also acknowledge the Earth system grid federation infrastructure (ESGF), an international effort led by the U.S. Department of Energy’s program for climate model diagnosis and intercomparison, the European network for Earth system modelling, and other partners in the global organisation for Earth system science portals (GO-ESSP). Further, we acknowledge the Copernicus climate change service and thank the ECMWF for producing and publishing the improved reanalysis ERA5. The results contain modified Copernicus climate change service information 2020. Neither the European commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains. We thank the anonymous reviewer for a careful and detailed review that lead to substantial improvements of this publication. Without all the collaborating staff from the observatories who sent us their in situ data, this undertaking would never have come to fruition. We thank Dr. Andrei Tokovinin, Astronomer at the Cerro Tololo inter-American observatory (CTIO), Edison Bustos, Software Engineer at NOIRLab, and Dr. John Blakeslee, head of science staff for observatory support at NOIRLab. From the Canada France Hawaii telescope (CFHT), we thank Mary Beth Laychak, director of strategic

communications, Cameron Wipper, astronomy technical specialist, and Grant Matsushige, senior instrumentation specialist. We thank Dr. Harriet Parsons, support astronomer at the East Asian observatory (EAO). From the University of Hawaii, we thank Dr. Mark Chun, associate director of the University of Hawaii’s institute for astronomy and specialist in instrumentation and data analysis, Dr. Steven Businger, Professor and chair of atmospheric sciences Department, Dr. Luke McKay, telescope engineer, and Robert Calder, senior engineer. We thank Ilse Plauchu-Frayn, astronomer and academic technician in observational support at the Observatorio Astronómico Nacional on the Sierra San Pedro Mártir (OAN-SPM), Marcello Lodi from the telescopio nazionale Galileo (TNG), Dr. Ennio Poretti, director of the TNG, and Dr. Andrew Adamson, associate director of Hawaii site at the Gemini observatory, Dr. Daniel Cotton, astronomer and operations technician at the Anglo-Australian telescope (AAT), and Dr. Steve Lee, senior technician based at the AAT, Michael Sharrott, project manager at Siding Spring observatory, Dr. Daniel Cunnama, science engagement astronomer at the southern African large telescope (SALT), and Dr. Myha Vuong De Breuck, astronomical database specialist at the European Southern Observatory (ESO). M.-E. Demory would like to thank Professor O. Romppainen-Martius and Professor K. Heng for hosting her at the University of Bern. We would like to thank Dr. Alexander Kashev from the faculty science IT support of the University of Bern for offering storage and computing power. We implemented code with python using numpy (Harris et al. 2020), scipy (Virtanen et al. 2020), matplotlib (Hunter 2007), pandas (McKinney 2010), netCDF4 (Unidata 2020), xarray (Hoyer & Hamman 2017), and the skillmetrics package developed by Peter Rochford²⁸. For implementing Bayesian analysis, we used the computer language R (R Core Team 2019) including rethinking (McElreath 2016), ggplot2 (Wickham 2016) and rstan (Team 2020).

References

- AAO. 2013, Observer, 123
- Abahamid, A., Vernin, J., Benkhalidoun, Z., et al. 2004, *A&A*, 422, 1123
- Allen, M. R., & Ingram, W. J. 2002, *Nature*, 419, 228
- Baker, A. J., Schiemann, R., Hodges, K. I., et al. 2019, *J. Climate*, 32, 7763
- Baldwin, M. P., Gray, L. J., Dunkerton, T. J., et al. 2001, *Rev. Geophys.*, 39, 179
- Bauer, P., Stevens, B., & Hazeleger, W. 2021, *Nature Clim. Change*, 11, 80
- Bell, B., Hersbach, H., Berrisford, P., et al. 2020, ERA5 hourly data on pressure levels from 1950 to 1978 (preliminary version)
- Bessell, M. S., Bessell, S., M. 2005, *ARA&A*, 43, 293
- Betancourt, M. 2017, ArXiv [arXiv:1701.02434]
- Bischoff-Gauß, I., Kalthoff, N., & Fiebig-Wittmaack, M. 2006, *Theor. Appl. Climatol.*, 85, 227
- Bony, S., Stevens, B., Frierson, D. M., et al. 2015, *Nat. Geosci.*, 8, 261
- Bradley, E., Roberts, Jr., L., Bradford, L., et al. 2006, *PASP*, 118, 172
- Brookfield, R., Anderson, A., Cotton, D. V., & Ramage, C. 2020, Operations Report – AAT (Semester 19B), Report for User Committee, Tech. rep., Australian National University
- Buck, A. L. 1981, *J. Appl. Meteorol.*, 20, 1527
- Burtscher, L., Dalgleish, H., Barret, D., et al. 2021, *Nat. Astron.*, 5, 857
- Byrne, M. P., & O’Gorman, P. A. 2018, *Proc. Natl. Acad. Sci. U.S.A.*, 115, 4863
- Cantalloube, F., Milli, J., Böhm, C., et al. 2020, *Nat. Astron.*, 4, 826
- Cao, G., Giambelluca, T. W., Stevens, D. E., & Schroeder, T. A. 2007, *J. Climate*, 20, 1145
- Castro-Almazán, J. A., Muñoz-Tuñón, C., García-Lorenzo, B., et al. 2016, in *Proc. SPIE*, 9910, 99100P
- Catala, L., Crawford, S. M., Buckley, D. A. H., et al. 2013, *MNRAS*, 436, 590
- Cavazzani, S., Ortolani, S., & Zitelli, V. 2012, *MNRAS*, 419, 3081
- Chapman, I. M., Naylor, D. A., & Phillips, R. R. 2004, *MNRAS*, 354, 621
- Collins, M., Knutti, R., Arblaster, J., et al. 2013, *Long-term Climate Change: Projections, Commitments and Irreversibility*, Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Tech. rep., ed. T. F. Stocker (Cambridge University Press)
- Copernicus Climate Change Service (C3S). 2020, ERA5 hourly data on pressure levels from 1979 to present (Copernicus Climate Change Service Climate Data Store (CDS))
- Cruz-Gonzales, I., Avila, R., Tapia, M., et al. 2004, in *Proc. SPIE*, 5382
- Da Silva, S. C., Businger, S., & Schroeder, T. 2012, High altitude climate of the island of Hawaii (University of Hawaii)
- Davies, R., & Kasper, M. 2012, *ARA&A*, 50, 305
- Dempsey, J. T., Friberg, P., Jenness, T., et al. 2013, *MNRAS*, 430, 2534
- EC-Earth Consortium. 2018, EC-Earth-Consortium EC-Earth3P-HR model output prepared for CMIP6 HighResMIP

²⁷ <https://digital-strategy.ec.europa.eu/en/policies/destination-earth>, accessed on 2021 Jul. 06.

²⁸ <https://github.com/PeterRochford/SkillMetrics>, accessed on 2021 Sep. 22.

- ESO. 2011, The E-ELT Construction Proposal, Tech. rep., European Southern Observatory, Garching bei München
- ESO. 2015, La Silla — ESO's First Observatory, Tech. rep., European Southern Observatory
- European Centre for Medium-Range Weather Forecasts. 2014, ERA-20C Project (ECMWF Atmospheric Reanalysis of the 20th Century)
- Eyring, V., Bony, S., Meehl, G. A., et al. 2016, *Geoscientific Model Dev.*, **9**, 1937
- Falvey, M., & Rojo, P. M. 2016, *Theor. Appl. Climatol. J.*, **125**, 841
- Fried, D. L. 1966, *J. Opt. Soc. Am.*, **56**, 1372
- Gelman, A., & Rubin, D. B. 1992, *Stat. Sci.*, **7**, 457
- Giambelluca, T. W., & Schroeder, M. A. 1986, Rainfall Atlas of Hawaii, Tech. rep., Department of Land and Natural Resources Hawaii
- Giambelluca, T. W., Diaz, H. F., & Luke, M. S. 2008, *Geophys. Res. Lett.*, **35**, 12
- Giorgi, F. 2019, *J. Geophys. Res.: Atmos.*, **124**, 5696
- Giorgi, F., Jones, C., & Asrar, G. 2009, *World Meteorol. Organ. Bull.*, **58**, 175
- Gladstone, J. H., & Dale, T. P. 1863, *Philos. Trans. Roy. Soc. Lond.*, **153**, 317
- Gleckler, P. J., Taylor, K. E., & Doutriaux, C. 2008, *J. Geophys. Res. Atmos.*, **113**, D6
- Graham, E., Sarazin, M. S., Beniston, M., et al. 2004, in *Ground-based Telescopes*, *Proc. SPIE*, 5489
- Graham, E., Sarazin, M., Kurlandczyk, H., Neun, M., & Matzler, C. 2008, *Proc. SPIE, Ground-based and Airborne Telescopes II*, 7012, 70121Y
- Grenon, M. 1990, *ESO Messenger*, **61**, 11
- Haarsma, R. J., Hazeleger, W., Severijns, C., et al. 2013, *Geophys. Res. Lett.*, **40**, 1783
- Haarsma, R. J., Roberts, M. J., Vidale, P. L., et al. 2016, *Geosci. Model Dev.*, **9**, 4185
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, **585**, 357
- Haustein, K., Allen, M. R., Forster, P. M., et al. 2017, *Sci. Rep.*, **7**, 1
- He, J., & Soden, B. J. 2016, *J. Climate*, **29**, 4317
- Held, I. M., & Soden, B. J. 2006, *J. Climate*, **19**, 5686
- Hellemeier, J. A., Yang, R., Sarazin, M., & Hickson, P. 2019, *MNRAS*, **482**, 4941
- Hersbach, H., Bell, B., Berrisford, P., et al. 2018a, ERA5 hourly data on pressure levels from 1979 to present
- Hersbach, H., Bell, B., Berrisford, P., et al. 2018b, ERA5 hourly data on single levels from 1979 to present
- Hersbach, H., Bell, B., Berrisford, P., et al. 2020, *Q. J. Roy. Meteorol. Soc.*, **146**, 1
- Hoyer, S., & Hamman, J. J. 2017, *J. Open Res. Softw.*, **5**, 10
- Hunter, J. D. 2007, *Comput. Sci. Eng.*, **9**, 90
- IPCC. 2013, *Summary for Policymakers*, Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Tech. rep., eds. T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, & S. K. Allen (Cambridge University Press)
- IPCC. 2021, *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Tech. rep. (Cambridge University Press)
- Kagawa-Viviani, A. K., & Giambelluca, T. W. 2020, *J. Geophys. Res.: Atmos.*, **125**, e2019JD031571
- Kalinnikov, V. V., & Khutorova, O. G. 2017, *Ann. Geophys.*, **35**, 453
- Källberg, P. W., Simmons, A., Uppala, S., & Fuentes, M. 2004, The ERA-40 archive [revised October 2007], Tech. rep., ECMWF, Shinfield Park, Reading
- Kalnay, E., Kanamitsu, M., Kistler, R., et al. 1996, *Bull. Am. Meteorol. Soc.*, **77**, 437
- Kennedy, J., Titchner, H., Rayner, N., & Roberts, M. 2017, Earth System Grid Federation, <https://doi.org/10.22033/ESGF/input4MIPs.1221>
- Kerber, F., Rose, T., Chacón, A., et al. 2012, *SPIE*, 8446, 84463N
- Kerber, F., Querel, R. R., & Hanuschik, R. 2014, *Observatory Operations: Strategies, Processes, and Systems V*, 9149, 229
- Kidger, M. R., Rodríguez-Espinosa, J. M., Del Rosario, J. C., & Trancho, G. 1998, *New Astron. Rev.*, **42**, 537
- Knutti, R., Abramowitz, G., Collins, M., et al. 2010, *Good Practice Guidance Paper on Assessing and Combining Multi Model Climate Projections*, Meeting Report of the Intergovernmental Panel on Climate Change Expert Meeting on Assessing and Combining Multi Model Climate Projections, Tech. rep., IPCC Working Group I Technical Support, eds. T. F. Stocker, & D. Qin (Bern, Switzerland: University of Bern)
- Kornilov, V., Tokovinin, A. A., Vozyakova, O., et al. 2003, MASS: a monitor of the vertical turbulence distribution, *Proc. SPIE*, 4839
- Kriegler, E., Bauer, N., Popp, A., et al. 2017, *Global Environ. Change*, **42**, 297
- López, J. A., Gutiérrez, L., & Pedro Mártir, S. 2003, *Rev. Mex. Astron. Astrofís.*, **19**, 3
- Lyman, R., Cherubini, T., & Businger, S. 2020, *MNRAS*, **496**, 4734
- Mahrt, L., Thomas, C., Richardson, S., et al. 2013, *Boundary-Layer Meteorol.*, **147**, 179
- Masciadri, E., Vernin, J., & Bougeault, P. 1999, *A&AS*, **137**, 185
- Masciadri, E., Lascaux, F., Turchi, A., & Fini, L. 2017, *MNRAS*, **466**, 520
- McElreath, R. 2016, rethinking: Statistical Rethinking book package. R package version 1.60.
- McIlveen, R. 1992, *Fundamentals of Weather and Climate*, 2nd edn. (Oxford: Oxford University Press)
- McInnes, B., & Walker, M. F. 1974, *PASP*, **86**, 529
- McKinney, W. 2010, in Proceedings of the 9th Python in Science Conference
- Meehl, G. A. 1995, *Bull. Am. Meteorol. Soc.*, **76**, 951
- Meehl, G. A., Boer, G. J., Covey, C., Latif, M., & Stouffer, R. J. 2000, *Bull. Am. Meteorol. Soc.*, **81**, 313
- Meyer, M. R., Currie, T., Guyon, O., et al. 2018, ArXiv e-prints, [arXiv:1804.03218]
- Molinari, E., & Hernandez, N. 2012, in *Observatory Operations: Strategies, Processes, and Systems IV*, 8448 (SPIE), 844822
- Molinari, E., & Hernandez, N. 2014, in *Observatory Operations: Strategies, Processes, and Systems V*, 9149 (SPIE), 914927
- Moreno-Chamarro, E., Caron, L.-P., Loosveldt Tomas, S., et al. 2022, *Geosci. Model Dev.*, **15**, 269
- Muñoz-Tuñón, C., Vernin, J., & Varela, A. M. 1997, *A&AS*, **125**, 183
- Navarra, A., Stern, W. F., & Miyakoda, K. 1994, *J. Climate*, **7**, 1169
- Osborn, J., & Sarazin, M. 2018, *MNRAS*, **480**, 1278
- Otarola, A., De Breuck, C., Trouvillon, T., et al. 2019, *PASP*, **131**, 045001
- Parker, W. S. 2016, *Bull. Am. Meteorol. Soc.*, **97**, 1565
- Pathak, P., dit de la Roche, D. J. M. P., Kasper, M., et al. 2021, *A&A*, **652**, A121
- Payne, A. E., Demory, M.-E., Leung, L. R., et al. 2020, *Nat. Rev. Earth Environ.*, **1**, 143
- Peixoto, J. P., Oort, A. H., Covey, C., & Taylor, K. 1992, *Phys. Today*, **45**, 8
- Pena-Ortiz, C., Gallego, D., Ribera, P., Ordóñez, P., & Alvarez-Castro, M. D. C. 2013, *J. Geophys. Res.: Atmos.*, **118**, 2702
- Plauchu-Frayn, I., Colorado, E., Richer, M. G., & Herrera-Vázquez, C. 2020, *Rev. Mex. Astron. Astrofís.*, **56**, 295
- Querel, R. R., Kerber, F., Lo Curto, G., et al. 2010, in *Ground-based and Airborne Telescopes III*, *Proc. SPIE*, 7733, 773349
- Racine, R. 2005, *PASP*, **117**, 401
- R Core Team. 2019, R: a language and environment for statistical computing.
- Reichler, T., & Kim, J. 2008, *Am. Meteorol. Soc.*, **89**, 303
- Riahi, K., Rao, S., Krey, V., et al. 2011, *Climatic Change*, **109**, 33
- Riahi, K., van Vuuren, D. P., Kriegler, E., et al. 2017, *Global Environ. Change*, **42**, 153
- Ribeiro, L., Kretschmer, N., Nascimento, J., et al. 2015, *Hydrol. Sci. J.*, **60**, 1840
- Roberts, M. 2017, MOHC HadGEM3-GC31-HM model output prepared for CMIP6 HighResMIP
- Roberts, M. J., Vidale, P. L., Mizielinski, M. S., et al. 2015, *J. Climate*, **28**, 574
- Roberts, M. J., Hewitt, H. T., Hyder, P., et al. 2016, *Geophys. Res. Lett.*, **43**, 10,430
- Roberts, C. D., Senan, R., Molteni, F., Boussetta, S., & Keeley, S. 2017, CMIP6 Citation 'ECMWF ECMWF-IFS-HR model output prepared for CMIP6 HighResMIP'
- Roberts, M., Hewitt, H., Iovino, D., et al. 2018, in EGU General Assembly Conference Abstracts (Adrian New), 17903
- Roddier, F. 1981, *Progr. Opt.*, **19**, 281
- Sarazin, M. 1988, VLT Report Nr. 65: Comparison of Meteorological Conditions on Chilean Sites - Annual Summary 1986, Tech. rep., European Southern Observatory
- Sarazin, M. 2001, Atmospheric Time Constants at Paranal during VLTI VINCI & Siderostats Commissioning
- Sarazin, M., & Roddier, F. 1990, *A&A*, **227**, 294
- Sarazin, M., & VLT Site Selection Working Group. 1990, VLT Report Nr. 62, Tech. rep., European Southern Observatory
- Sarazin, M., Melnick, J., Navarrete, J., & Lombardi, G. 2008, *The Messenger*, **132**, 11
- Sarazin, M., Le Louarn, M., Ascenso, J., Lombardi, G., & Navarrete, J. 2013, 3rd AO4ELT Conference – Adaptive Optics for Extremely Large Telescopes
- Schiemann, R., Vidale, P. L., Hatcher, R., & Roberts, M. 2019, NERC HadGEM3-GC31-HM model output prepared for CMIP6 HighResMIP
- Schoeck, M., Els, S., Riddle, R., et al. 2009, *PASP*, **121**, 384
- Scoccimarro, E., Bellucci, A., & Peano, D. 2017, CMCC CMCC-CM2-VHR4 model output prepared for CMIP6 HighResMIP
- Shaffrey, L. C., Stevens, I., Norton, W. A., et al. 2009, *J. Climate*, **22**, 1861
- Stevens, B., Satoh, M., Auger, L., et al. 2019, *Progr. Earth Planet. Sci.*, **6**, 61
- Tallis, M., Bailey, V. P., Macintosh, B., et al. 2020, *J. Astron. Telescopes Instrum. Syst.*, **6**, 1
- Tatarski, V. I., Silverman, R. A., & Chako, N. 1961, *Phys. Today*, **14**, 46

- Tatarskii, V. I. 1971, *The Effects of the Turbulent Atmosphere on Wave Propagation* (Jerusalem: Israel Program for Scientific Translations)
- Taylor, K. E. 2001, *J. Geophys. Res. Atmos.*, **106**, 7183
- Taylor, K. E., Juckes, M., Balaji, V., et al. 2018, *CMIP6_global_attributes_filenames_CVs* - Google Docs
- Team, S. D. 2020, RStan: the R interface to Stan. R package version 2.19.3
- Teare, S., Thompson, L., Gino, M., & Palmer, K. 2000, *PASP*, **112**, 1496
- Trenberth, K. E., Fasullo, J., & Smith, L. 2005, *Climate Dyn.*, **24**, 741
- Unidata. 2020, Network Common Data Form (netCDF) version 1.5.3, <https://www.unidata.ucar.edu/software/netcdf/>
- Vannière, B., Demory, M. E., Vidale, P. L., et al. 2020, *Climate Dyn.*, **52**, 6817
- Vernin, J. 1986, in *Proc. SPIE*, 0628
- Vernin, J. & Muñoz-Tuon, C. 1994, *A&A*, **284**, 311
- Vernin, J., Muñoz-Tuñón, C., Sarazin, M., et al. 2011, *PASP*, **123**, 1334
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nat. Methods*, **17**, 261
- Voltaire, A. 2019, CNRM-CERFACS CNRM-CM6-1-HR model output prepared for CMIP6 HighResMIP
- von Storch, J.-S., Putrasahan, D., Lohmann, K., et al. 2017, MPI-M MPI-ESM1.2-XR model output prepared for CMIP6 HighResMIP
- Walker, C., Chinigò, D., & Dubow, S. 2019, *J. Southern Afr. Stud.*, **45**, 627
- Ward, D. J., & Galewsky, J. 2014, *J. Geophys. Res.: Earth Surf.*, **119**, 2048
- Wedi, N. P. 2014, *Philos. Trans. Roy. Soc. A: Math. Phys. Eng. Sci.*, **372**
- Wickham, H. 2016, *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York)
- Wiedner, M. C., Hills, R. E., Carlstrom, J. E., & Lay, O. P. 2001, *ApJ*, **553**, 1036
- Willett, K. M., Dunn, R. J. H., Thorne, P. W., et al. 2014, *Climate Past*, **10**, 1983
- Wu, T., Lu, Y., Fang, Y., et al. 2019, *Geoscientific Model Dev.*, **12**, 1573
- Wyant, M. C., Wood, R., Bretherton, C. S., et al. 2010, *Atmos. Chem. Phys.*, **10**, 4757
- Zhang, C., Wang Dr., Y., Lauer, A., & Hamilton, K. 2012, *Monthly Weather Rev.*, **140**, 3259

Appendix A: Data in tables

Appendix A.1: Skill scores

Table A.1. Skill score (Eq. 17) and skill score classification (Table 5) of **temperature** for all sites.

Site	ERA5 to in situ	PRIMAVERA (to in situ) simulation	PRIMAVERA to in situ	PRIMAVERA (to ERA5) simulation	PRIMAVERA to ERA5
Mauna Kea	0.86	atmos-past	0.56	atmos-past	0.60
Paranal	0.97	atmos-past	0.61	atmos-past	0.61
La Silla	0.94	coupled-past	0.74	coupled-past	0.74
Cerro Tololo	0.84	coupled-past	0.66	coupled-past	0.78
La Palma	0.99	atmos-past	0.89	atmos-past	0.89
Siding Spring	0.99	atmos-past	0.93	atmos-past	0.93
Sutherland	1.00	coupled-past	0.93	coupled-past	0.89
SPM	1.00	atmos-past	0.90	atmos-past	0.91

Notes. The skill scores are colour coded as follows: poor skill score as dark red, mediocre skill score as red, good skill score as orange and excellent skill score as yellow. The best skill score of the atmosphere-only or coupled PRIMAVERA simulation, depending on which scored higher, with respect to in situ data and to ERA5 data is indicated.

Table A.2. As Table A.1, but for **relative humidity**.

Site	ERA5 to in situ	PRIMAVERA (to in situ) simulation	PRIMAVERA to in situ	PRIMAVERA (to ERA5) simulation	PRIMAVERA to ERA5
Mauna Kea	0.80	atmos-past	0.15	atmos-past	0.23
Paranal	0.84	atmos-past	0.47	atmos-past	0.47
La Silla	0.71	atmos-past	0.36	atmos-past	0.37
Cerro Tololo	0.57	atmos-past	0.28	atmos-past	0.26
La Palma	0.69	atmos-past	0.48	coupled-past	0.46
Siding Spring	0.82	atmos-past	0.53	atmos-past	0.43
Sutherland	0.96	coupled-past	0.29	coupled-past	0.21
SPM	0.89	atmos-past	0.55	atmos-past	0.51

Table A.3. As Table A.1, but for **specific humidity**.

Site	ERA5 to in situ	PRIMAVERA (to in situ) simulation	PRIMAVERA to in situ	PRIMAVERA (to ERA5) simulation	PRIMAVERA to ERA5
Mauna Kea	0.82	atmos-past	0.23	atmos-past	0.34
Paranal	0.87	atmos-past	0.53	atmos-past	0.54
La Silla	0.87	atmos-past	0.80	atmos-past	0.79
Cerro Tololo	0.89	atmos-past	0.79	atmos-past	0.70
La Palma	0.83	atmos-past	0.55	coupled-past	0.52
Siding Spring	0.90	atmos-past	0.82	atmos-past	0.79
Sutherland	0.99	atmos-past	0.76	atmos-past	0.75
SPM	0.98	atmos-past	0.81	atmos-past	0.82

Table A.4. As Table A.1, but for **PWV**. '...' indicates missing in situ data.

Site	ERA5 to in situ	PRIMAVERA (to in situ) simulation	PRIMAVERA to in situ	PRIMAVERA (to ERA5) simulation	PRIMAVERA to ERA5
Mauna Kea	0.51	atmos-past	0.16	atmos-past	0.34
Paranal	0.98	atmos-future	0.49	atmos-past	0.48
La Silla	0.62	atmos-past	0.41	atmos-past	0.42
Cerro Tololo	atmos-past	0.60
La Palma	atmos-past	0.59
Siding Spring	atmos-past	0.82
Sutherland	atmos-past	0.75
SPM	atmos-past	0.86

Table A.5. As Table A.1, but for **cloud cover**. ‘...’ indicates missing in situ data.

Site	ERA5 to in situ	PRIMAVERA (to in situ) simulation	PRIMAVERA to in situ	PRIMAVERA (to ERA5) simulation	PRIMAVERA to ERA5
Mauna Kea	0.13	coupled-past	0.07	atmos-past	0.18
Paranal	0.36	atmos-past	0.09	atmos-past	0.21
La Silla	0.58	atmos-past	0.44	atmos-past	0.58
Cerro Tololo	0.68	coupled-past	0.66	atmos-past	0.69
La Palma	0.54	coupled-past	0.03	atmos-past	0.06
Siding Spring	0.72	atmos-past	0.14	atmos-past	0.16
Sutherland	coupled-past	0.34
SPM	atmos-past	0.44

Table A.6. As Table A.1, but for **200-hPa-wind-speed seeing**. ‘...’ indicates missing in situ data.

Site	ERA5 to in situ	PRIMAVERA (to in situ) simulation	PRIMAVERA to in situ	PRIMAVERA (to ERA5) simulation	PRIMAVERA to ERA5
Mauna Kea	0.39	coupled-past	0.31	atmos-past	0.61
Paranal	0.06	coupled-past	0.05	atmos-past	0.66
La Silla	0.27	atmos-past	0.18	atmos-past	0.50
Cerro Tololo	0.27	atmos-past	0.23	atmos-past	0.46
La Palma	0.11	coupled-past	0.28	coupled-past	0.49
Siding Spring	atmos-past	0.68
Sutherland	atmos-past	0.42
SPM	0.08	coupled-past	0.10	atmos-past	0.67

Table A.7. As Table A.1, but for **seeing model**. ‘...’ indicates missing in situ data.

Site	ERA5 to in situ	PRIMAVERA (to in situ) simulation	PRIMAVERA to in situ	PRIMAVERA (to ERA5) simulation	PRIMAVERA to ERA5
Mauna Kea	0.19	coupled-past	0.29	coupled-past	0.52
Paranal	0.26	coupled-past	0.10	coupled-past	0.52
La Silla	0.50	atmos-past	0.27	coupled-past	0.57
Cerro Tololo	0.54	atmos-past	0.33	coupled-past	0.55
La Palma	0.02	coupled-past	0.01	coupled-past	0.57
Siding Spring	atmos-past	0.62
Sutherland	atmos-past	0.56
SPM	0.16	coupled-past	0.19	coupled-past	0.75

Appendix A.2: Future climate trends

Table A.8. Trends of **temperature** (in °C per decade) for ERA5 (1979-2019), PRIMAVERA historical (1950-2014) simulations and PRIMAVERA future (2015-2050) simulations.

Site	ERA5	coupled-past	atmos-past	coupled-future	atmos-future
Mauna Kea	0.03 ^{+0.06} _{-0.06}	0.14 ^{+0.01} _{-0.01}	0.11 ^{+0.02} _{-0.02}	0.57 ^{+0.04} _{-0.04}	0.34 ^{+0.05} _{-0.05}
Paranal	0.23 ^{+0.06} _{-0.06}	0.13 ^{+0.02} _{-0.02}	0.15 ^{+0.02} _{-0.02}	0.54 ^{+0.04} _{-0.04}	0.46 ^{+0.05} _{-0.05}
La Silla	0.20 ^{+0.06} _{-0.06}	0.12 ^{+0.02} _{-0.02}	0.15 ^{+0.02} _{-0.02}	0.47 ^{+0.05} _{-0.06}	0.41 ^{+0.08} _{-0.08}
Cerro Tololo	0.15 ^{+0.07} _{-0.06}	0.11 ^{+0.02} _{-0.02}	0.14 ^{+0.02} _{-0.02}	0.48 ^{+0.05} _{-0.05}	0.42 ^{+0.07} _{-0.07}
La Palma	0.19 ^{+0.09} _{-0.09}	0.12 ^{+0.04} _{-0.04}	0.11 ^{+0.04} _{-0.04}	0.44 ^{+0.11} _{-0.11}	0.37 ^{+0.10} _{-0.10}
Siding Spring	0.24 ^{+0.08} _{-0.08}	0.12 ^{+0.02} _{-0.02}	0.12 ^{+0.03} _{-0.03}	0.44 ^{+0.07} _{-0.07}	0.38 ^{+0.06} _{-0.07}
Sutherland	0.13 ^{+0.08} _{-0.07}	0.11 ^{+0.03} _{-0.02}	0.11 ^{+0.03} _{-0.02}	0.53 ^{+0.07} _{-0.07}	0.41 ^{+0.07} _{-0.07}
SPM	0.21 ^{+0.08} _{-0.08}	0.15 ^{+0.02} _{-0.02}	0.11 ^{+0.03} _{-0.03}	0.46 ^{+0.06} _{-0.06}	0.40 ^{+0.07} _{-0.07}

Notes. The 89% percentile credible interval of predicted values sampled from the posterior distribution (Sect. 5.3) is given in brackets.

Table A.9. As Table A.8, but for **relative humidity** (in % per decade).

Site	ERA5	coupled-past	atmos-past	coupled-future	atmos-future
Mauna Kea	$0.23^{+0.33}_{-0.33}$	$-0.03^{+0.10}_{-0.09}$	$-0.03^{+0.11}_{-0.12}$	$-0.13^{+0.26}_{-0.27}$	$-0.05^{+0.36}_{-0.36}$
Paranal	$-0.22^{+0.21}_{-0.19}$	$-0.07^{+0.07}_{-0.07}$	$0.02^{+0.08}_{-0.08}$	$-0.02^{+0.16}_{-0.17}$	$0.04^{+0.21}_{-0.22}$
La Silla	$-0.84^{+0.23}_{-0.24}$	$-0.12^{+0.08}_{-0.08}$	$-0.11^{+0.10}_{-0.10}$	$-0.11^{+0.22}_{-0.21}$	$-0.13^{+0.26}_{-0.27}$
Cerro Tololo	$-0.68^{+0.20}_{-0.21}$	$-0.10^{+0.09}_{-0.08}$	$-0.14^{+0.09}_{-0.08}$	$-0.21^{+0.20}_{-0.20}$	$-0.20^{+0.25}_{-0.25}$
La Palma	$-0.12^{+0.25}_{-0.26}$	$-0.07^{+0.13}_{-0.13}$	$-0.05^{+0.12}_{-0.13}$	$-0.27^{+0.30}_{-0.32}$	$-0.22^{+0.35}_{-0.34}$
Siding Spring	$-0.35^{+0.29}_{-0.28}$	$-0.04^{+0.10}_{-0.09}$	$-0.01^{+0.10}_{-0.10}$	$-0.03^{+0.26}_{-0.24}$	$-0.36^{+0.25}_{-0.25}$
Sutherland	$-0.35^{+0.25}_{-0.25}$	$0.01^{+0.07}_{-0.07}$	$-0.00^{+0.08}_{-0.08}$	$-0.26^{+0.20}_{-0.21}$	$-0.20^{+0.23}_{-0.23}$
SPM	$-0.30^{+0.32}_{-0.31}$	$-0.07^{+0.11}_{-0.11}$	$-0.01^{+0.13}_{-0.12}$	$0.07^{+0.31}_{-0.31}$	$0.03^{+0.29}_{-0.30}$

Table A.10. As Table A.8, but for **specific humidity** (in g kg^{-1} per decade).

Site	ERA5	coupled-past	atmos-past	coupled-future	atmos-future
Mauna Kea	$0.05^{+0.03}_{-0.03}$	$0.02^{+0.01}_{-0.01}$	$0.01^{+0.01}_{-0.01}$	$0.07^{+0.03}_{-0.03}$	$0.04^{+0.04}_{-0.04}$
Paranal	$-0.04^{+0.04}_{-0.04}$	$0.00^{+0.01}_{-0.01}$	$0.02^{+0.01}_{-0.01}$	$0.06^{+0.03}_{-0.03}$	$0.06^{+0.03}_{-0.04}$
La Silla	$-0.19^{+0.04}_{-0.04}$	$0.01^{+0.01}_{-0.01}$	$0.01^{+0.02}_{-0.02}$	$0.11^{+0.04}_{-0.04}$	$0.07^{+0.06}_{-0.06}$
Cerro Tololo	$-0.14^{+0.03}_{-0.03}$	$0.00^{+0.01}_{-0.01}$	$0.00^{+0.01}_{-0.01}$	$0.10^{+0.03}_{-0.03}$	$0.07^{+0.04}_{-0.04}$
La Palma	$0.03^{+0.04}_{-0.04}$	$0.02^{+0.02}_{-0.02}$	$0.03^{+0.02}_{-0.02}$	$0.06^{+0.04}_{-0.04}$	$0.06^{+0.05}_{-0.05}$
Siding Spring	$-0.01^{+0.05}_{-0.05}$	$0.05^{+0.01}_{-0.01}$	$0.06^{+0.02}_{-0.02}$	$0.20^{+0.04}_{-0.04}$	$0.06^{+0.04}_{-0.05}$
Sutherland	$-0.05^{+0.04}_{-0.04}$	$0.05^{+0.01}_{-0.01}$	$0.04^{+0.01}_{-0.01}$	$0.15^{+0.04}_{-0.03}$	$0.12^{+0.04}_{-0.04}$
SPM	$-0.03^{+0.07}_{-0.07}$	$0.02^{+0.02}_{-0.02}$	$0.01^{+0.03}_{-0.03}$	$0.10^{+0.06}_{-0.06}$	$0.08^{+0.06}_{-0.06}$

Table A.11. As Table A.8, but for **PWV** (in mmH_2O per decade).

Site	ERA5	coupled-past	atmos-past	coupled-future	atmos-future
Mauna Kea	$0.09^{+0.05}_{-0.06}$	$0.04^{+0.01}_{-0.01}$	$0.03^{+0.02}_{-0.02}$	$0.14^{+0.05}_{-0.05}$	$0.11^{+0.07}_{-0.07}$
Paranal	$0.04^{+0.07}_{-0.07}$	$0.01^{+0.02}_{-0.02}$	$0.06^{+0.03}_{-0.03}$	$0.17^{+0.05}_{-0.05}$	$0.17^{+0.09}_{-0.09}$
La Silla	$-0.19^{+0.06}_{-0.06}$	$0.01^{+0.01}_{-0.01}$	$0.04^{+0.02}_{-0.03}$	$0.16^{+0.04}_{-0.04}$	$0.14^{+0.08}_{-0.08}$
Cerro Tololo	$-0.18^{+0.06}_{-0.06}$	$0.01^{+0.02}_{-0.02}$	$0.04^{+0.03}_{-0.03}$	$0.23^{+0.06}_{-0.06}$	$0.19^{+0.10}_{-0.10}$
La Palma	$0.09^{+0.06}_{-0.06}$	$0.04^{+0.01}_{-0.01}$	$0.05^{+0.02}_{-0.02}$	$0.09^{+0.04}_{-0.04}$	$0.20^{+0.06}_{-0.06}$
Siding Spring	$0.10^{+0.11}_{-0.11}$	$0.09^{+0.04}_{-0.03}$	$0.14^{+0.04}_{-0.04}$	$0.42^{+0.09}_{-0.09}$	$0.20^{+0.14}_{-0.13}$
Sutherland	$0.02^{+0.07}_{-0.07}$	$0.08^{+0.03}_{-0.02}$	$0.08^{+0.03}_{-0.03}$	$0.28^{+0.07}_{-0.06}$	$0.29^{+0.09}_{-0.09}$
SPM	$0.00^{+0.11}_{-0.12}$	$0.05^{+0.04}_{-0.04}$	$0.03^{+0.06}_{-0.06}$	$0.19^{+0.10}_{-0.10}$	$0.16^{+0.13}_{-0.14}$

Table A.12. As Table A.8, but for **cloud cover** (in % per decade).

Site	ERA5	coupled-past	atmos-past	coupled-future	atmos-future
Mauna Kea	$-0.09^{+0.33}_{-0.33}$	$-0.06^{+0.11}_{-0.11}$	$0.02^{+0.11}_{-0.11}$	$-0.07^{+0.25}_{-0.25}$	$-0.13^{+0.25}_{-0.25}$
Paranal	$0.04^{+0.33}_{-0.33}$	$-0.13^{+0.10}_{-0.09}$	$0.07^{+0.11}_{-0.11}$	$-0.14^{+0.27}_{-0.24}$	$-0.14^{+0.27}_{-0.29}$
La Silla	$-0.08^{+0.35}_{-0.35}$	$-0.05^{+0.10}_{-0.10}$	$0.07^{+0.11}_{-0.11}$	$0.10^{+0.27}_{-0.27}$	$-0.17^{+0.30}_{-0.29}$
Cerro Tololo	$-0.13^{+0.35}_{-0.33}$	$-0.05^{+0.09}_{-0.09}$	$0.06^{+0.10}_{-0.10}$	$0.15^{+0.28}_{-0.26}$	$-0.10^{+0.30}_{-0.30}$
La Palma	$0.05^{+0.28}_{-0.29}$	$-0.06^{+0.08}_{-0.08}$	$-0.06^{+0.08}_{-0.09}$	$-0.28^{+0.22}_{-0.21}$	$-0.22^{+0.22}_{-0.22}$
Siding Spring	$-0.17^{+0.32}_{-0.32}$	$-0.11^{+0.11}_{-0.10}$	$-0.06^{+0.12}_{-0.11}$	$-0.26^{+0.30}_{-0.31}$	$-0.32^{+0.32}_{-0.31}$
Sutherland	$-0.18^{+0.31}_{-0.30}$	$-0.00^{+0.09}_{-0.09}$	$0.06^{+0.10}_{-0.10}$	$-0.35^{+0.25}_{-0.23}$	$-0.15^{+0.28}_{-0.26}$
SPM	$-0.21^{+0.34}_{-0.34}$	$-0.12^{+0.10}_{-0.10}$	$0.05^{+0.13}_{-0.13}$	$0.22^{+0.28}_{-0.26}$	$-0.13^{+0.30}_{-0.32}$

Table A.13. As Table A.8, but for **seeing model (13)** (in arcseconds per decade).

Site	ERA5	coupled-past	atmos-past	coupled-future	atmos-future
Mauna Kea	$-0.00^{+0.00}_{-0.00}$	$-0.00^{+0.00}_{-0.00}$	$0.00^{+0.00}_{-0.00}$	$-0.00^{+0.00}_{-0.00}$	$-0.00^{+0.00}_{-0.00}$
Paranal	$0.02^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$	$0.00^{+0.00}_{-0.00}$	$0.00^{+0.01}_{-0.01}$	$-0.01^{+0.01}_{-0.01}$
La Silla	$0.01^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$	$0.00^{+0.00}_{-0.00}$	$0.01^{+0.01}_{-0.01}$	$0.00^{+0.01}_{-0.01}$
Cerro Tololo	$0.01^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$	$0.00^{+0.00}_{-0.00}$	$0.01^{+0.01}_{-0.01}$	$0.00^{+0.01}_{-0.01}$
La Palma	$0.01^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$	$0.00^{+0.00}_{-0.00}$	$0.00^{+0.01}_{-0.01}$	$0.01^{+0.01}_{-0.01}$
Siding Spring	$-0.00^{+0.01}_{-0.01}$	$-0.01^{+0.01}_{-0.01}$	$-0.00^{+0.01}_{-0.01}$	$-0.00^{+0.01}_{-0.01}$	$0.00^{+0.01}_{-0.01}$
Sutherland	$0.00^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$	$0.00^{+0.00}_{-0.00}$	$0.01^{+0.01}_{-0.01}$	$0.01^{+0.01}_{-0.01}$
SPM	$0.00^{+0.00}_{-0.00}$	$-0.00^{+0.00}_{-0.00}$	$-0.00^{+0.00}_{-0.00}$	$-0.00^{+0.00}_{-0.00}$	$-0.00^{+0.01}_{-0.01}$

Table A.14. As Table A.8, but for **200-hPa-wind-speed seeing (12)** (in arcseconds per decade).

Site	ERA5	coupled-past	atmos-past	coupled-future	atmos-future
Mauna Kea	$-0.01^{+0.01}_{-0.01}$	$-0.00^{+0.00}_{-0.00}$	$0.00^{+0.00}_{-0.00}$	$0.00^{+0.01}_{-0.00}$	$0.00^{+0.01}_{-0.01}$
Paranal	$0.02^{+0.01}_{-0.01}$	$0.01^{+0.01}_{-0.01}$	$0.01^{+0.01}_{-0.01}$	$0.01^{+0.01}_{-0.01}$	$-0.00^{+0.01}_{-0.01}$
La Silla	$0.02^{+0.01}_{-0.01}$	$0.01^{+0.00}_{-0.00}$	$0.01^{+0.01}_{-0.01}$	$0.01^{+0.01}_{-0.01}$	$0.01^{+0.01}_{-0.01}$
Cerro Tololo	$0.02^{+0.01}_{-0.01}$	$0.01^{+0.00}_{-0.00}$	$0.01^{+0.00}_{-0.00}$	$0.01^{+0.01}_{-0.01}$	$0.01^{+0.01}_{-0.01}$
La Palma	$-0.00^{+0.02}_{-0.02}$	$0.00^{+0.01}_{-0.01}$	$0.00^{+0.01}_{-0.01}$	$-0.01^{+0.02}_{-0.02}$	$0.01^{+0.02}_{-0.02}$
Siding Spring	$-0.01^{+0.02}_{-0.02}$	$-0.00^{+0.01}_{-0.01}$	$0.00^{+0.01}_{-0.01}$	$0.01^{+0.02}_{-0.02}$	$0.03^{+0.03}_{-0.02}$
Sutherland	$0.03^{+0.01}_{-0.01}$	$0.01^{+0.00}_{-0.00}$	$0.01^{+0.01}_{-0.01}$	$0.02^{+0.01}_{-0.01}$	$0.02^{+0.01}_{-0.01}$
SPM	$-0.00^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$	$0.00^{+0.00}_{-0.00}$	$0.01^{+0.01}_{-0.01}$	$-0.00^{+0.01}_{-0.01}$

Appendix B: Diurnal cycle

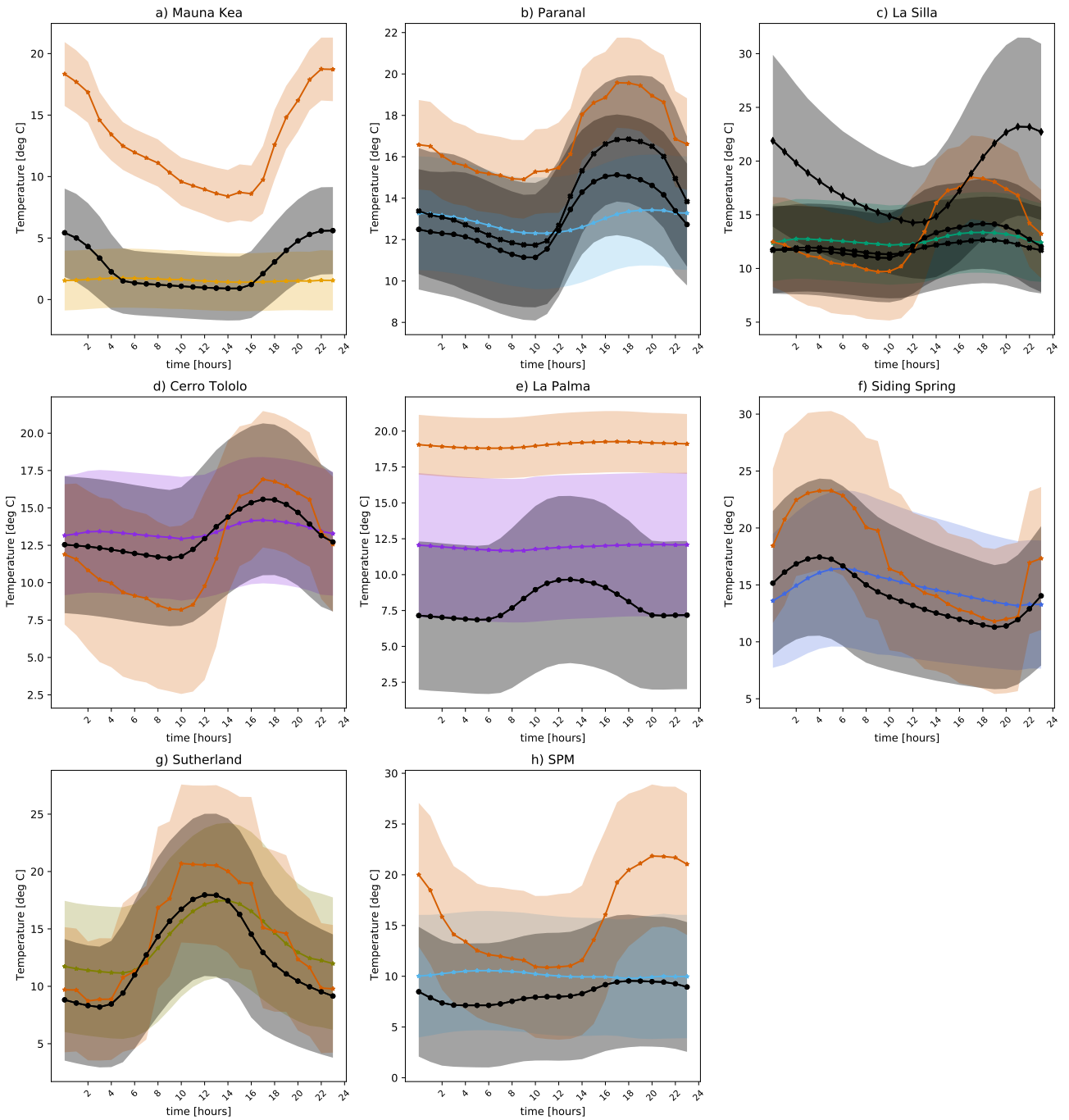


Fig. B.1. Diurnal cycle of temperature for in situ data (black markers) and ERA5 data (coloured markers). The coloured lines indicate the ERA5 two-metre temperature (orange) and ERA5 data on pressure levels 600 hPa (dark-yellow; a), 750 hPa (bright blue; b and h), 775 hPa (green; c), 800 hPa (blue-violet; d and e), 850 hPa (olive; g) and 900 hPa (royalblue; f).. The black x-filled markers in b) represent additional in situ data measured 20 m below the telescope platform at Paranal observatory. The x-filled and the diamond markers in c) represent additional in situ data 30 m above ground and directly at the ground station of La Silla observatory, respectively. The grey and coloured shadings represent the standard deviation of the in situ and ERA5 data based on hourly averaged data, respectively. Time is in universal time (UT).

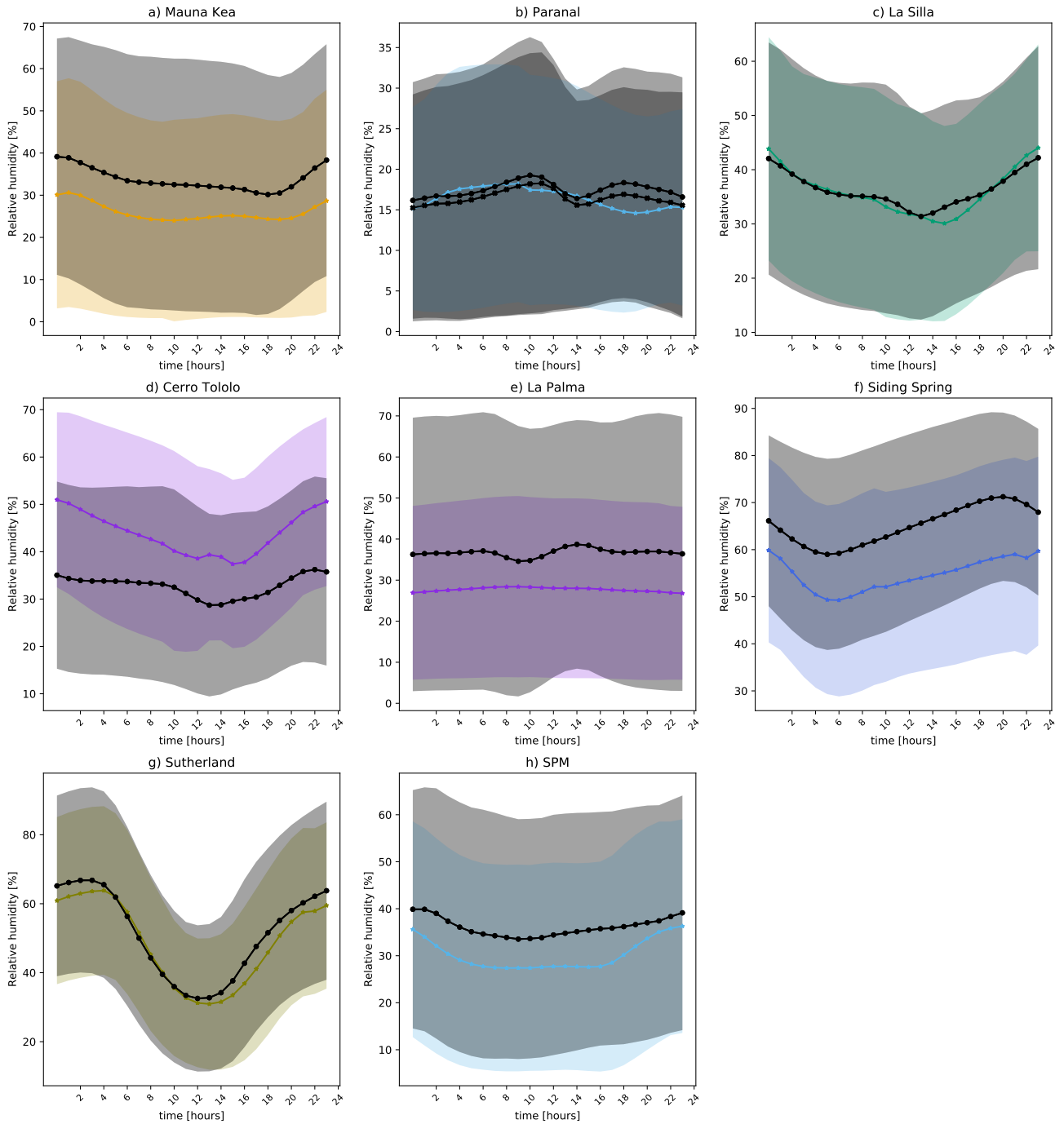


Fig. B.2. Diurnal cycle of relative humidity for in situ data (black markers) and ERA5 data (coloured markers). The coloured lines indicate ERA5 data on pressure levels 600 hPa (dark-yellow; a), 750 hPa (bright blue; b and h), 775 hPa (green; c), 800 hPa (blue-violet; d and e), 850 hPa (olive; g) and 900 hPa (royalblue; f). The black x-filled markers (b) represent additional in situ data measured 20 m below the telescope platform at Paranal observatory. The grey and coloured shadings represent the standard deviation of the in situ and ERA5 data based on hourly averaged data, respectively. Time is in universal time (UT).

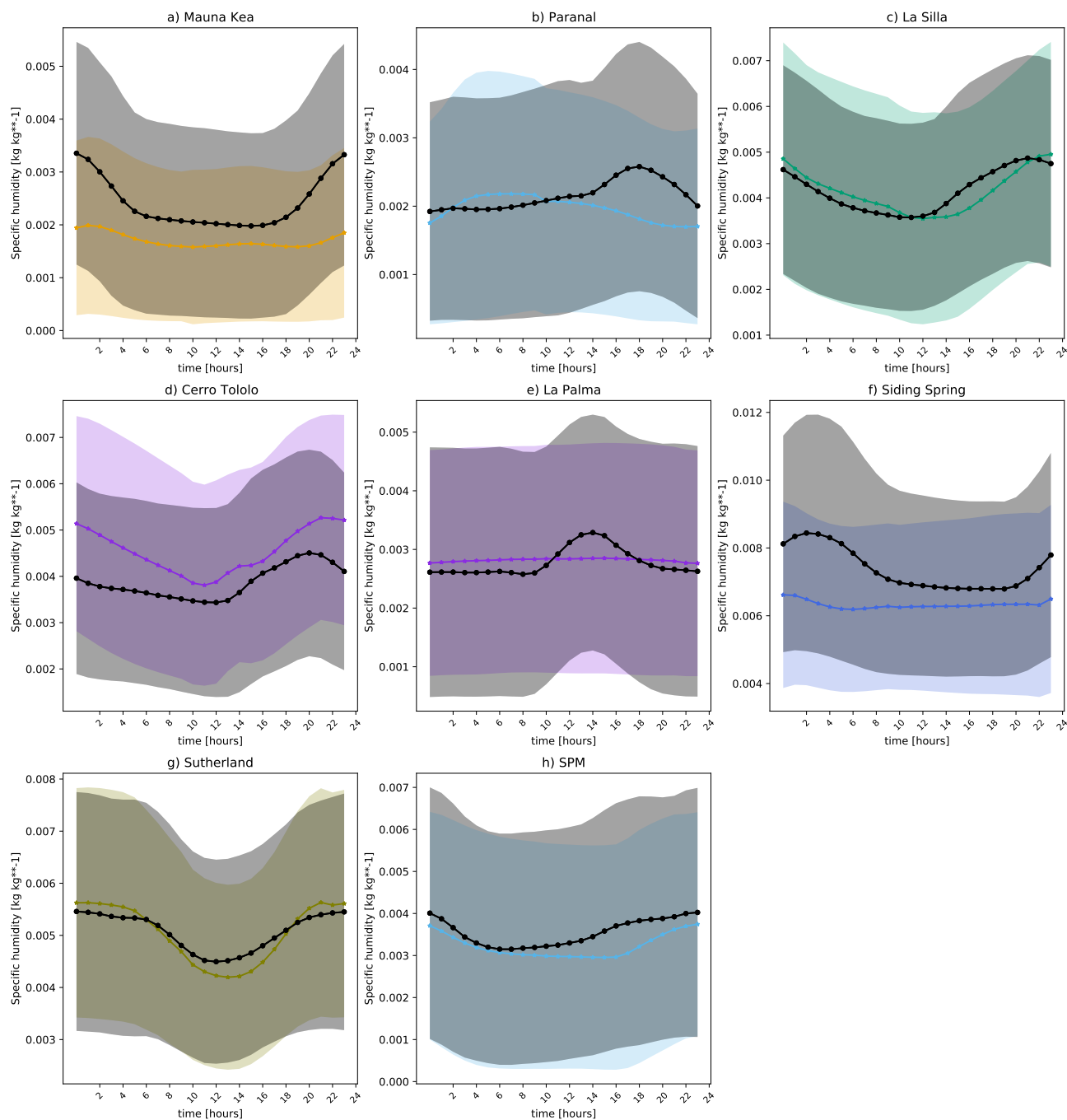


Fig. B.3. Diurnal cycle of specific humidity for in situ data (black markers) and ERA5 data (coloured markers). The coloured lines indicate ERA5 data on pressure levels 600 hPa (dark-yellow; a), 750 hPa (bright blue; b and h), 775 hPa (green; c), 800 hPa (blue-violet; d and e), 850 hPa (olive; g) and 900 hPa (royalblue; f). The grey and coloured shadings represent the standard deviation of the in situ and ERA5 data based on hourly averaged data, respectively. Time is in universal time (UT).

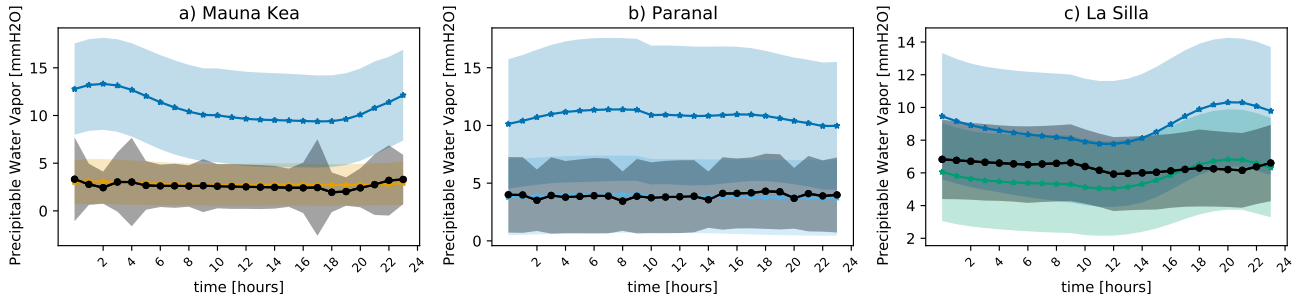


Fig. B.4. Diurnal cycle of PWV for in situ data (black markers) and ERA5 data (coloured markers). The ERA5 integrated PWV is shown as well as the ERA5 total column water output (*tcw*, blue). The grey and coloured shadings represent the standard deviation of the in situ and ERA5 data based on hourly averaged data, respectively. Time is in universal time (UT). In situ data are missing for Cerro Tololo, La Palma, Siding Spring, Sutherland and SPM. Data for La Silla is interpolated from 3-hourly data. The integration limit for the integrated PWV is 600 hPa for Mauna Kea (a, dark-yellow line), 750 hPa for Paranal (b, bright blue line) and 775 hPa for La Silla (c, green line).

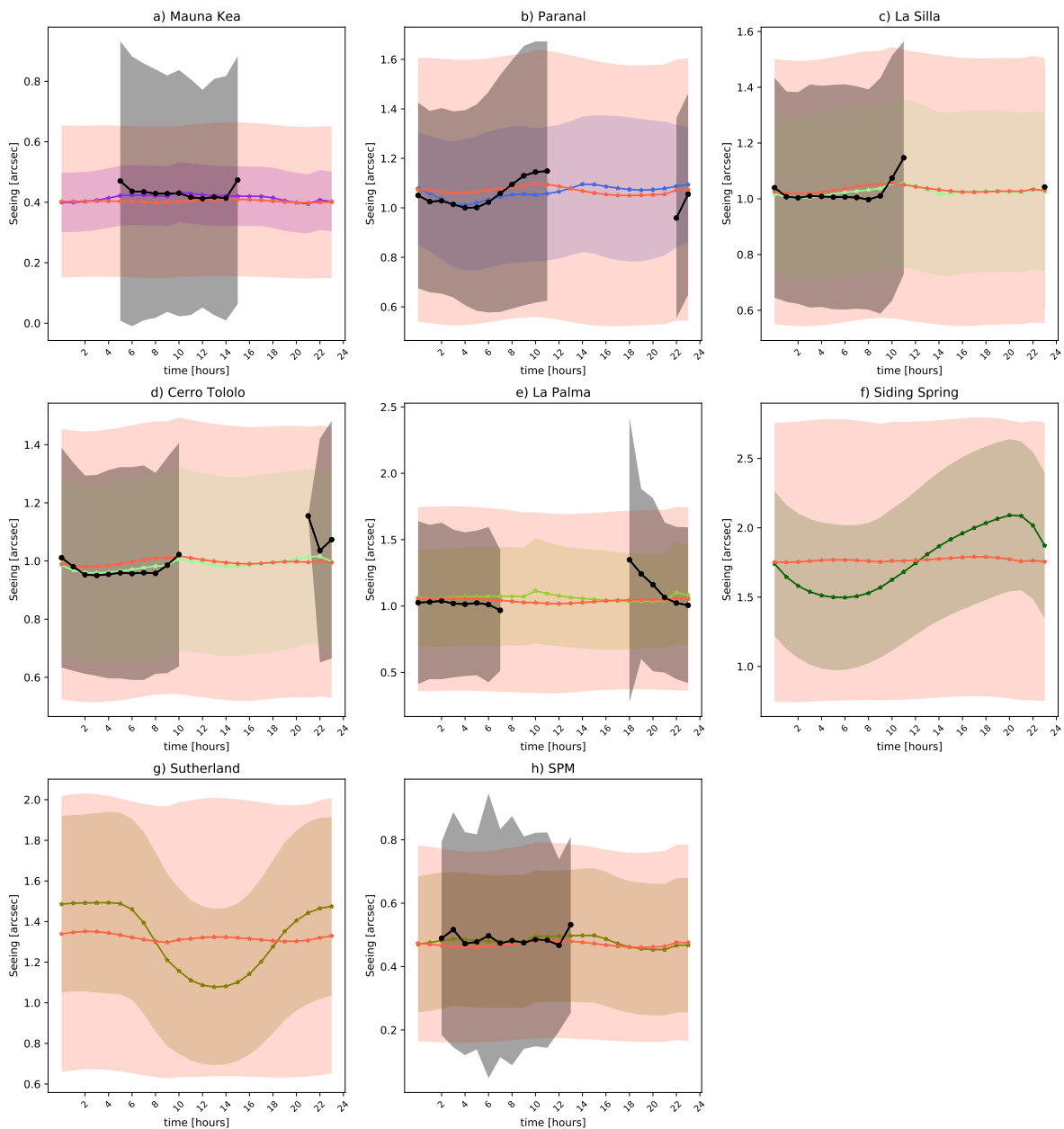


Fig. B.5. As Fig. B.1, but for astronomical seeing. For ERA5, the 200-hPa-wind-speed seeing (Eq. 12, tomato-red markers) and the seeing model (Eq. 13, coloured markers) are shown. Time is in universal time (UT). The lower integration limits for the seeing model are 800 hPa (blue-violet; a), 825 hPa (pale-green; c and d), 850 hPa (olive; g and h) and 900 hPa (royalblue; b), 950 hPa (dark-green, f) and 975 hPa (yellow-green, e). In situ data are missing for Siding Spring and Sutherland.

Appendix C: Orography

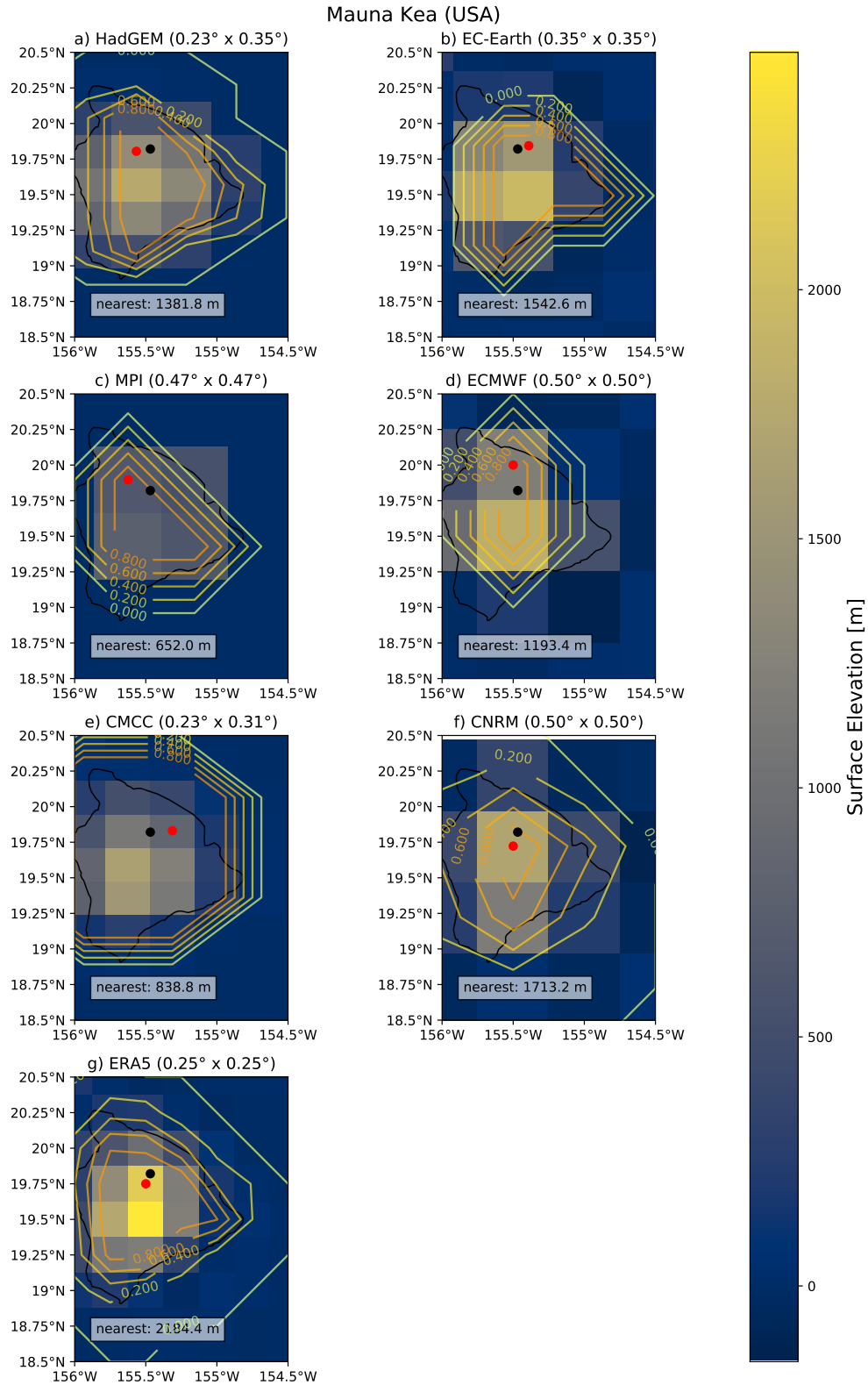


Fig. C.1. Surface elevation (in m) of Mauna Kea for PRIMAVERA (a-f) and ERA5 (g). Longitude is on the x-axis, latitude on the y-axis, both in degrees. The horizontal resolution of each model in this regional excerpt is given in degrees as latitude x longitude. The land-sea masks are shown with coloured contour lines, while the black contours represent the real border between land and sea. Black dots indicate the location of the Canada France Hawaii telescope. The red dots indicate the models' nearest grid point to the CFHT. Its elevation is 4204 m. For a global map view of the site location, see Figure 1.

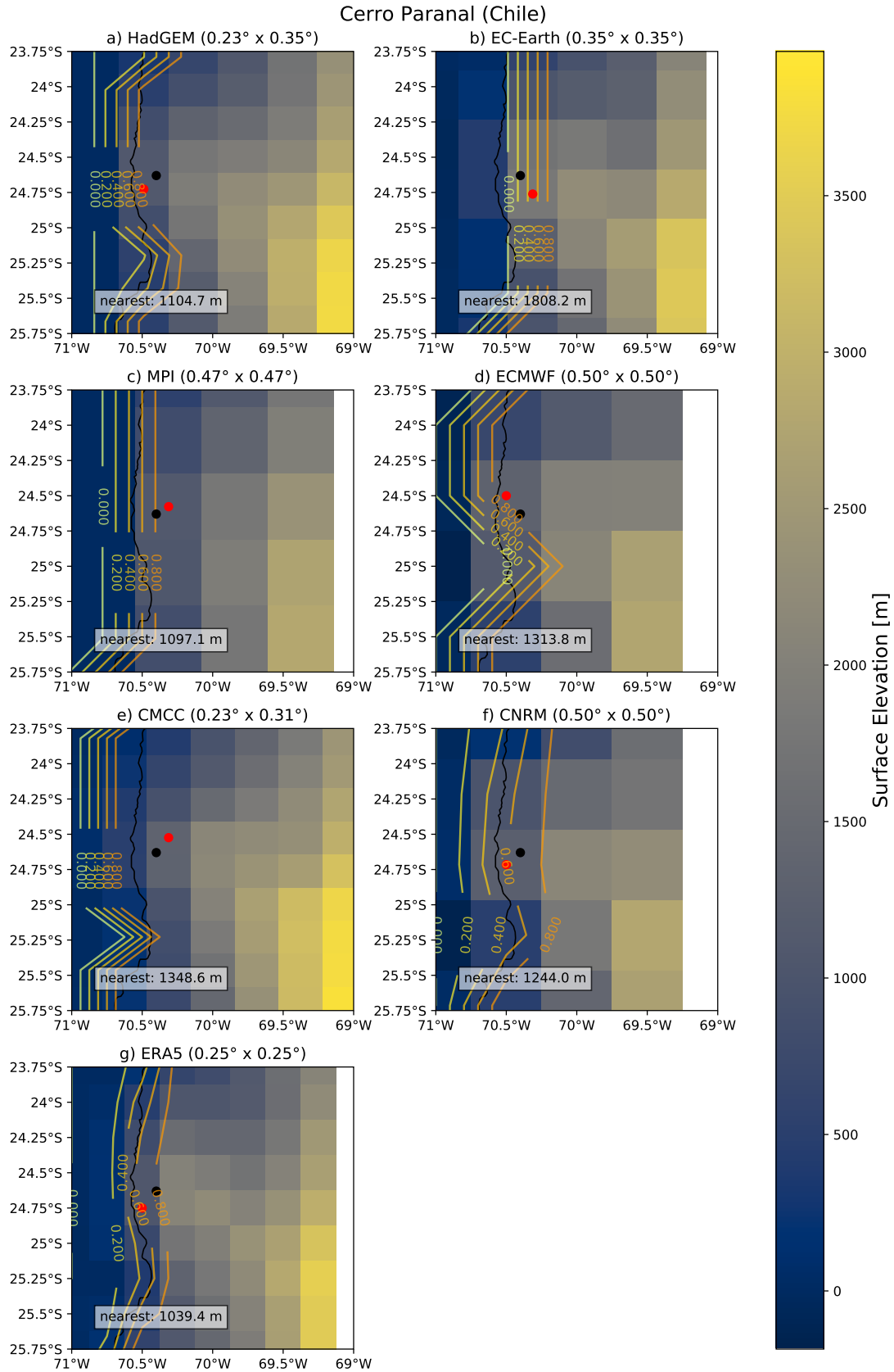


Fig. C.2. Surface elevation (in m) of Cerro Paranal for PRIMAVERA (a-f) and ERA5 (g). Longitude is on the x-axis, latitude on the y-axis, both in degrees. The horizontal resolution of each model in this regional excerpt is given in degrees as latitude×longitude. The land-sea masks are shown with coloured contour lines, while the black contours represent the real border between land and sea. Black dots indicate the location of the Paranal observatory (ESO). The red dots indicate the models' nearest grid point to the Paranal Observatory. Its elevation is 2635 m. For a global map view of the site location, see Figure 1.

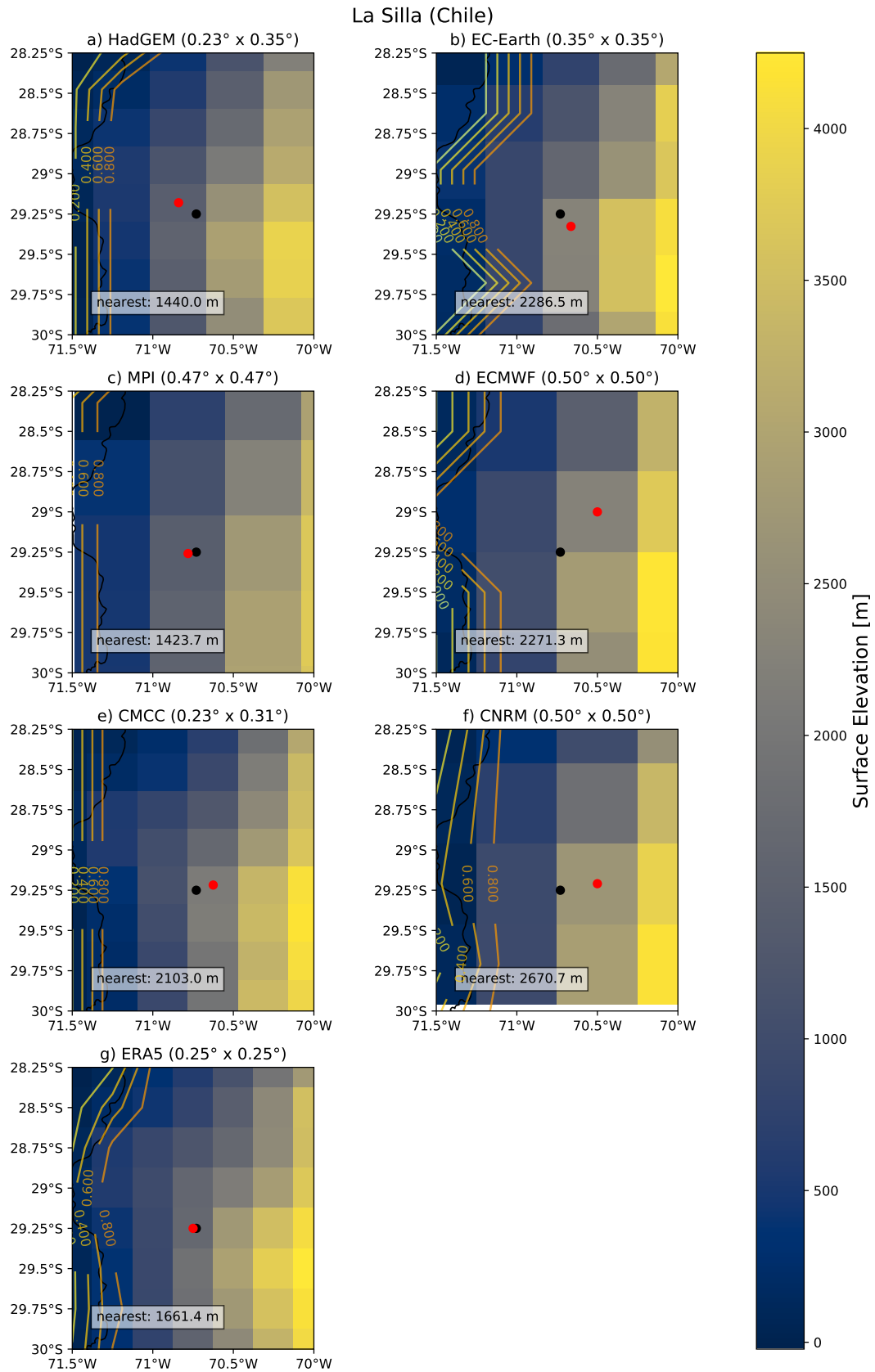


Fig. C.3. Surface elevation (in m) of La Silla for PRIMAVERA (a-f) and ERA5 (g). Longitude is on the x-axis, latitude on the y-axis, both in degrees. The horizontal resolution of each model in this regional excerpt is given in degrees as latitude×longitude. The land-sea masks are shown with coloured contour lines, while the black contours represent the real border between land and sea. Black dots indicate the location of the La Silla observatory. The red dots indicate the models’ nearest grid point to the La Silla observatory. Its elevation is 2400 m. For a global map view of the site location, see Figure 1.

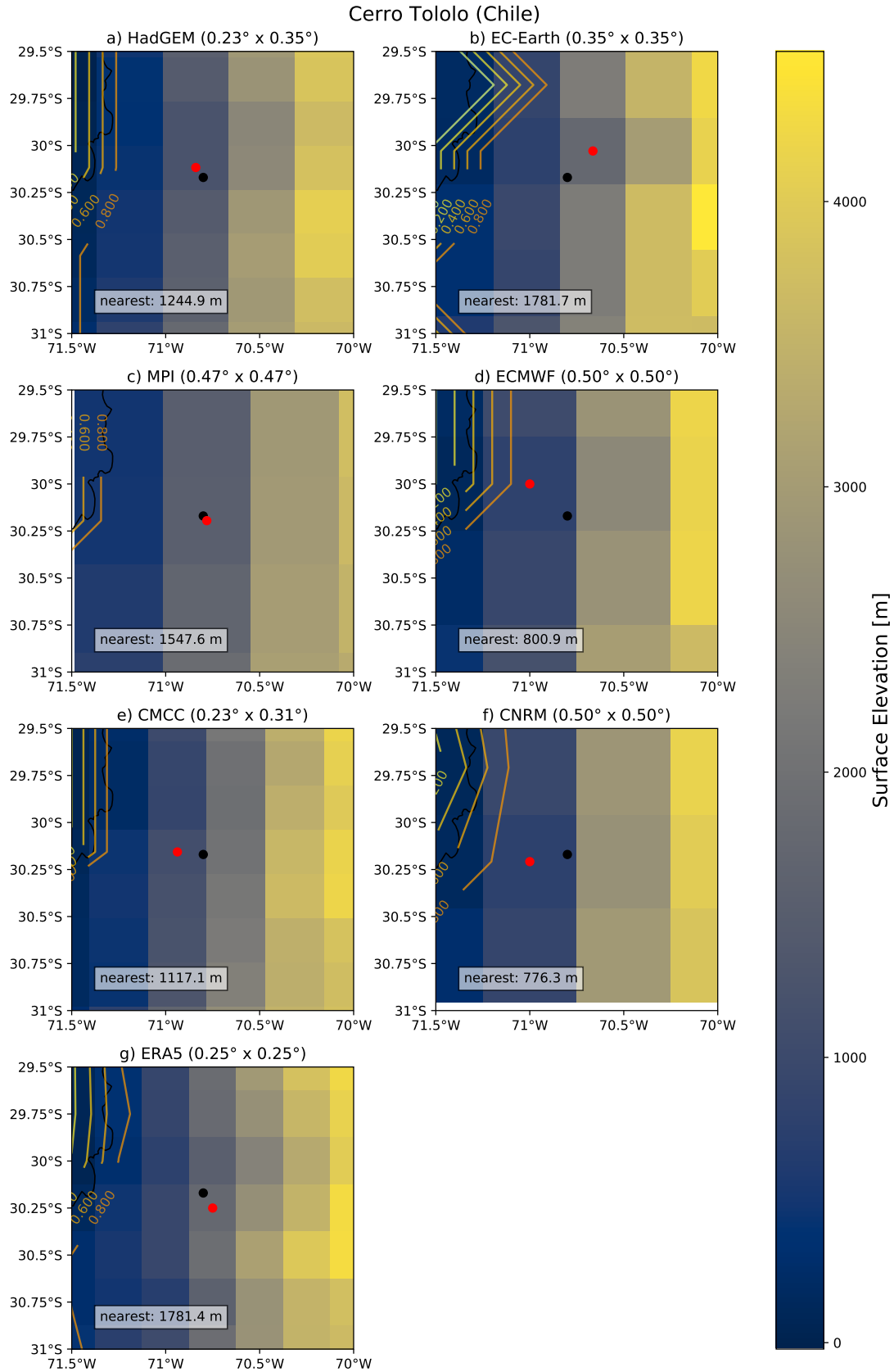


Fig. C.4. Surface elevation (in m) of Cerro Tololo for PRIMAVERA (a-f) and ERA5 (g). Longitude is on the x-axis, latitude on the y-axis, both in degrees. The horizontal resolution of each model in this regional excerpt is given in degrees as latitude×longitude. The land-sea masks are shown with coloured contour lines, while the black contours represent the real border between land and sea. Black dots indicate the location of the Cerro Tololo inter-American observatory (CTIO). The red dots indicate the models' nearest grid point to the CTIO. Its elevation is 2207 m. For a global map view of the site location, see Figure 1.

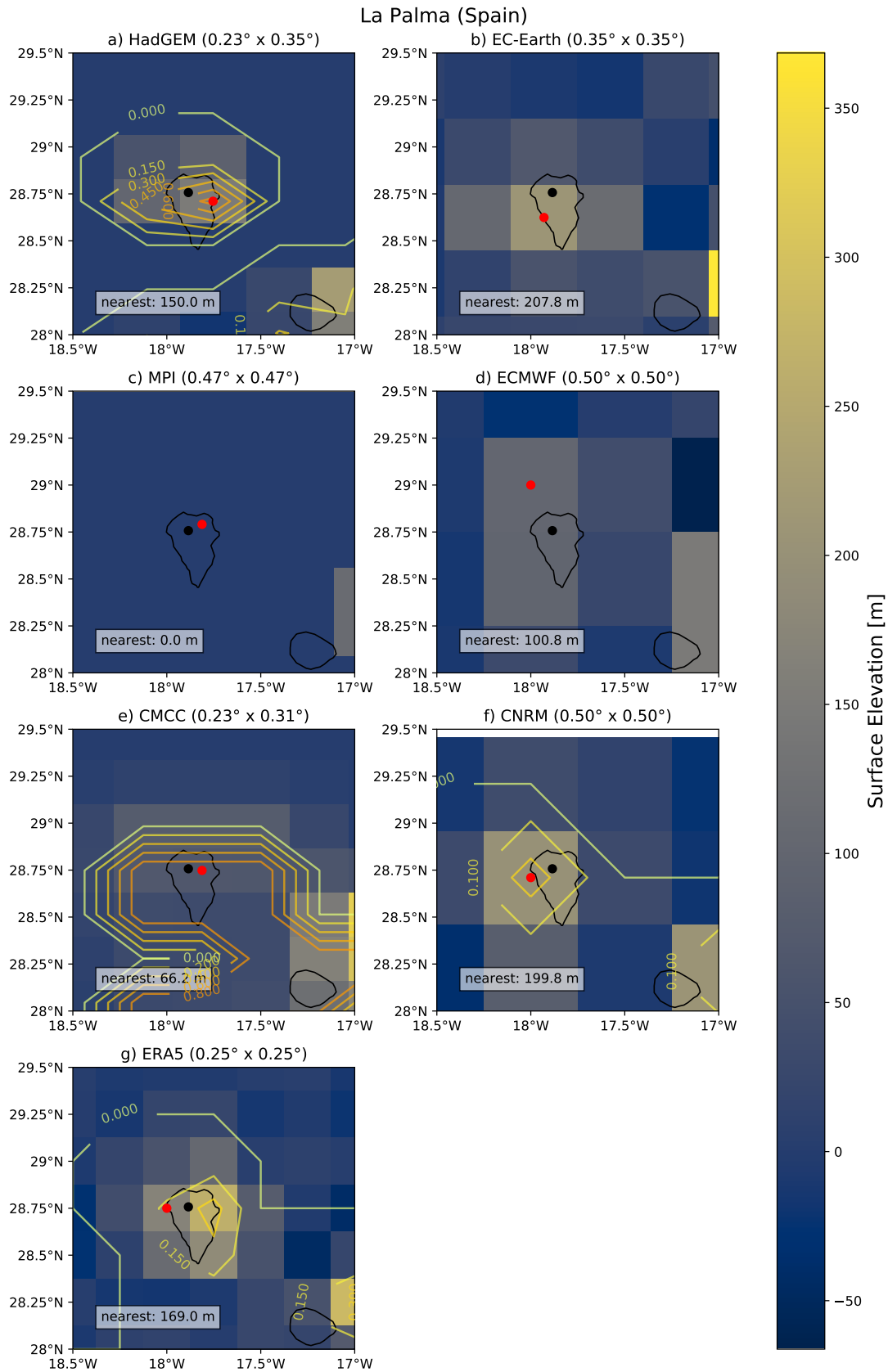


Fig. C.5. Surface elevation (in m) of La Palma for PRIMAVERA (a-f) and ERA5 (g). Longitude is on the x-axis, latitude on the y-axis, both in degrees. The horizontal resolution of each model in this regional excerpt is given in degrees as latitude×longitude. The land-sea masks are shown with coloured contour lines, while the black contours represent the real border between land and sea. Black dots indicate the location of the nordic optical telescope (NOT). The red dots indicate the models' nearest grid point to the NOT. Its elevation is 2382 m. We note that EC-EARTH, MPI and ECMWF do not classify it as an island. For a global map view of the site location, see Figure 1.

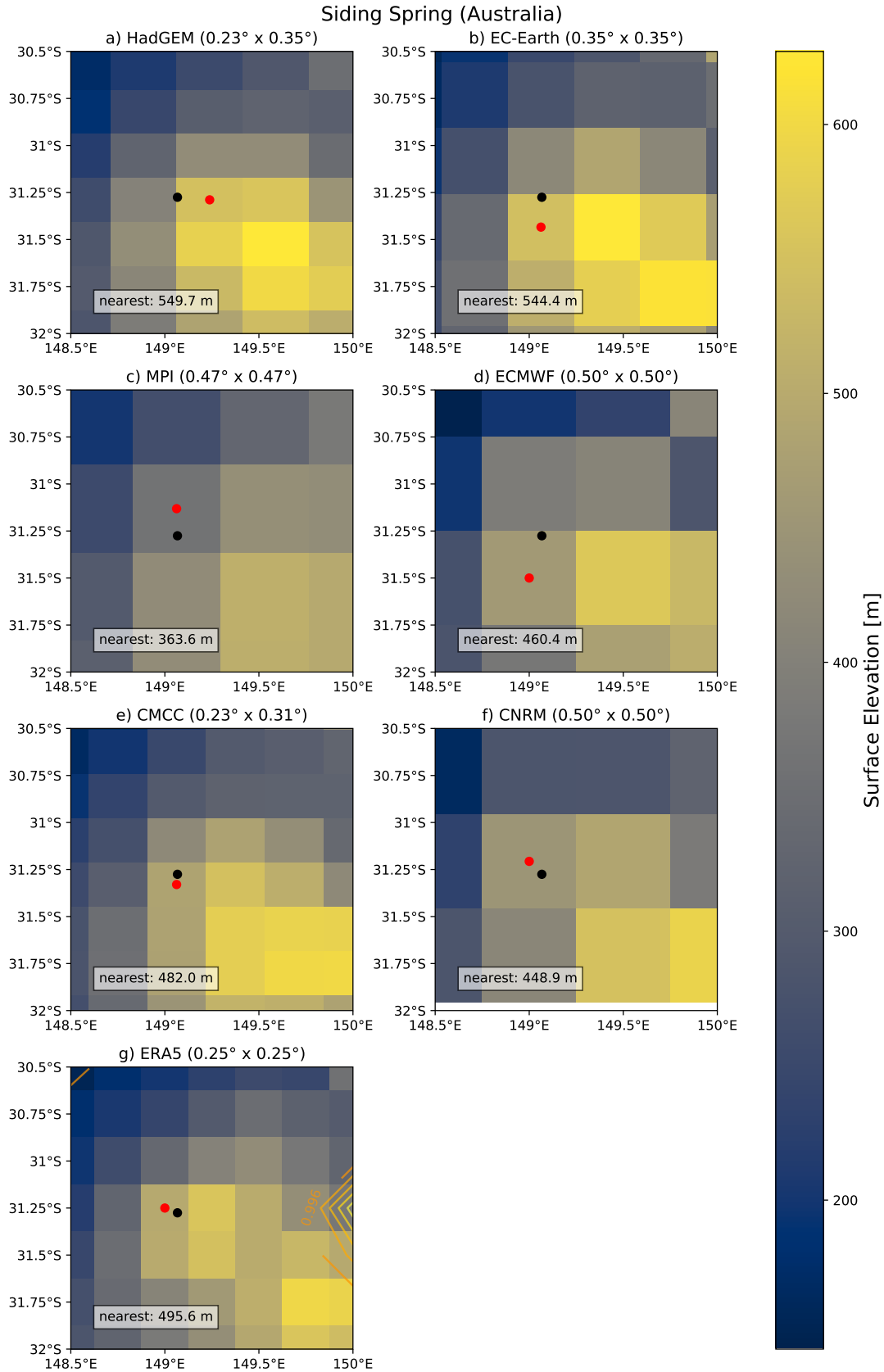


Fig. C.6. Surface elevation (in m) of Siding Spring for PRIMAVERA (a-f) and ERA5 (g). Longitude is on the x-axis, latitude on the y-axis, both in degrees. The horizontal resolution of each model in this regional excerpt is given in degrees as latitude×longitude. The land-sea masks are shown with coloured contour lines, while the black contours represent the real border between land and sea. Black dots indicate the location of the Anglo-Australian telescope (AAT). The red dots indicate the models' nearest grid point to the AAT. Its elevation is 1134 m. For a global map view of the site location, see Figure 1.

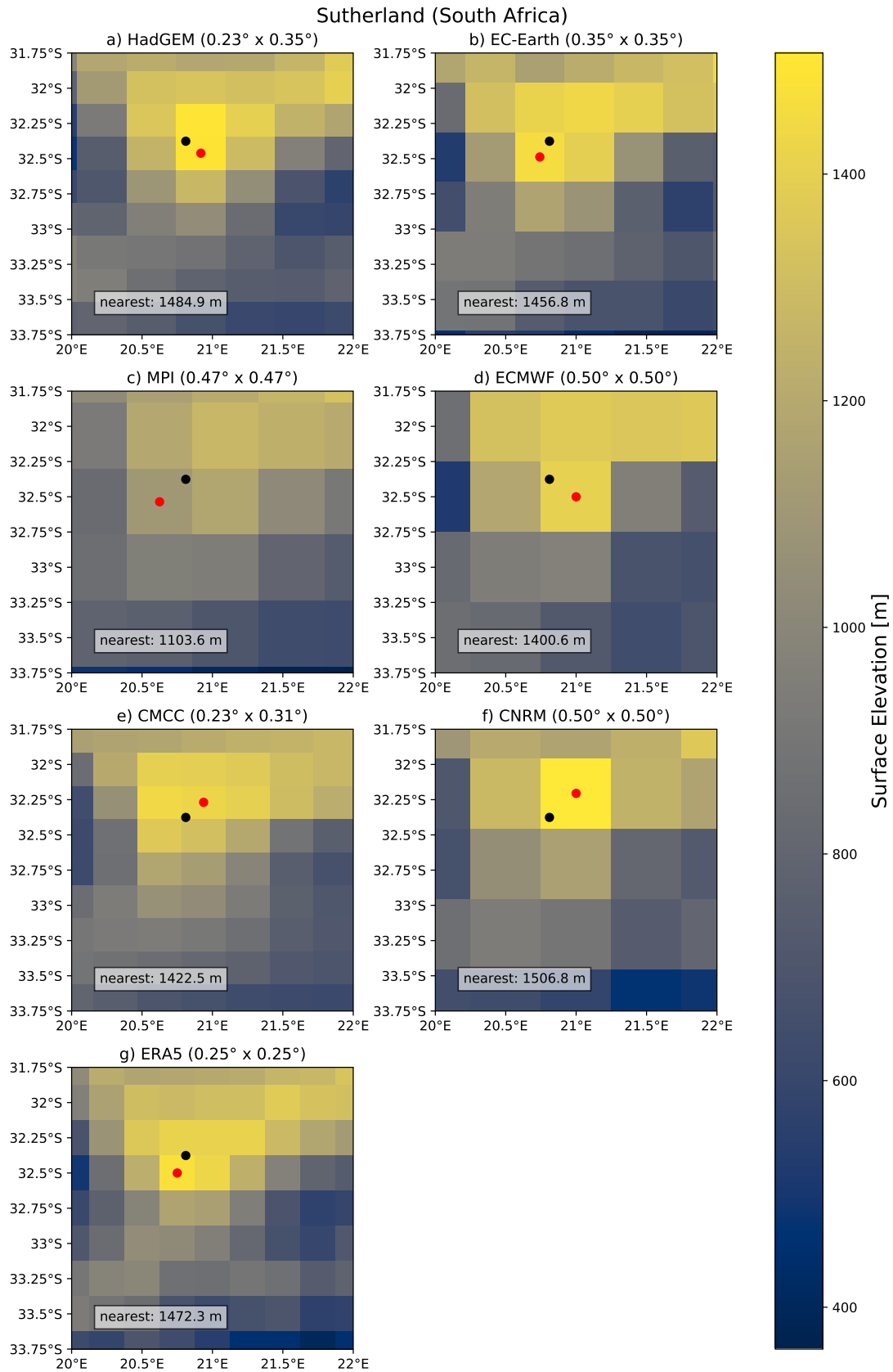


Fig. C.7. Surface elevation (in m) of Sutherland for PRIMAVERA (a-f) and ERA5 (g). Longitude is on the x-axis, latitude on the y-axis, both in degrees. The horizontal resolution of each model in this regional excerpt is given in degrees as latitude×longitude. The land-sea masks are shown with coloured contour lines, while the black contours represent the real border between land and sea. Black dots indicate the location of the south african large telescope (SALT). The red dots indicate the models' nearest grid point to the SALT. Its elevation is 1798 m. For a global map view of the site location, see Figure 1.

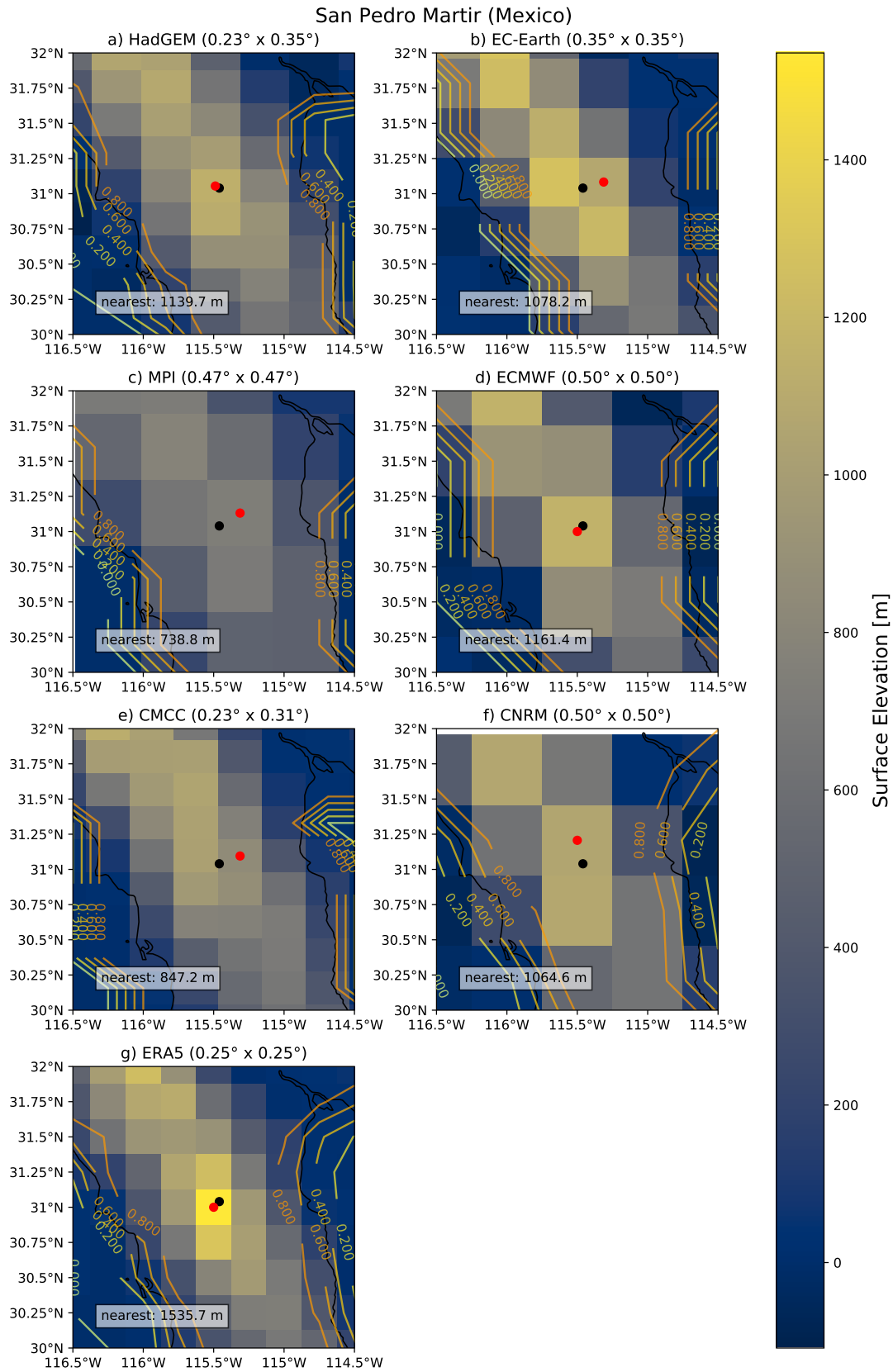


Fig. C.8. Surface elevation (in m) of SPM for PRIMAVERA (a-f) and ERA5 (g). Longitude is on the x-axis, latitude on the y-axis, both in degrees. The horizontal resolution of each model in this regional excerpt is given in degrees as latitude×longitude. The land-sea masks are shown with coloured contour lines, while the black contours represent the real border between land and sea. Black dots indicate the location of the national astronomical observatory (OAN). The red dots indicate the models' nearest grid point to the OAN. Its elevation is 2800 m. For a global map view of the site location, see Figure 1.

Appendix D: Vertical resolution of climate models output

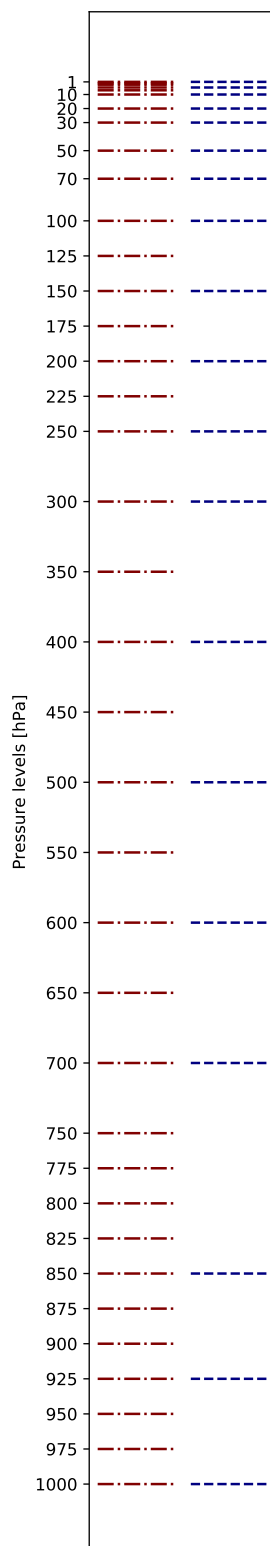


Fig. D.1. Available pressure levels for ERA5 (red dash-dotted) and PRIMAVERA (blue dashed) data. The pressure levels between 1 hPa and 10 hPa are 1, 2, 3, 5, 7 and 10 hPa for ERA5, and 1, 5 and 10 hPa for PRIMAVERA.

Appendix E: Vertical profile of C_n^2

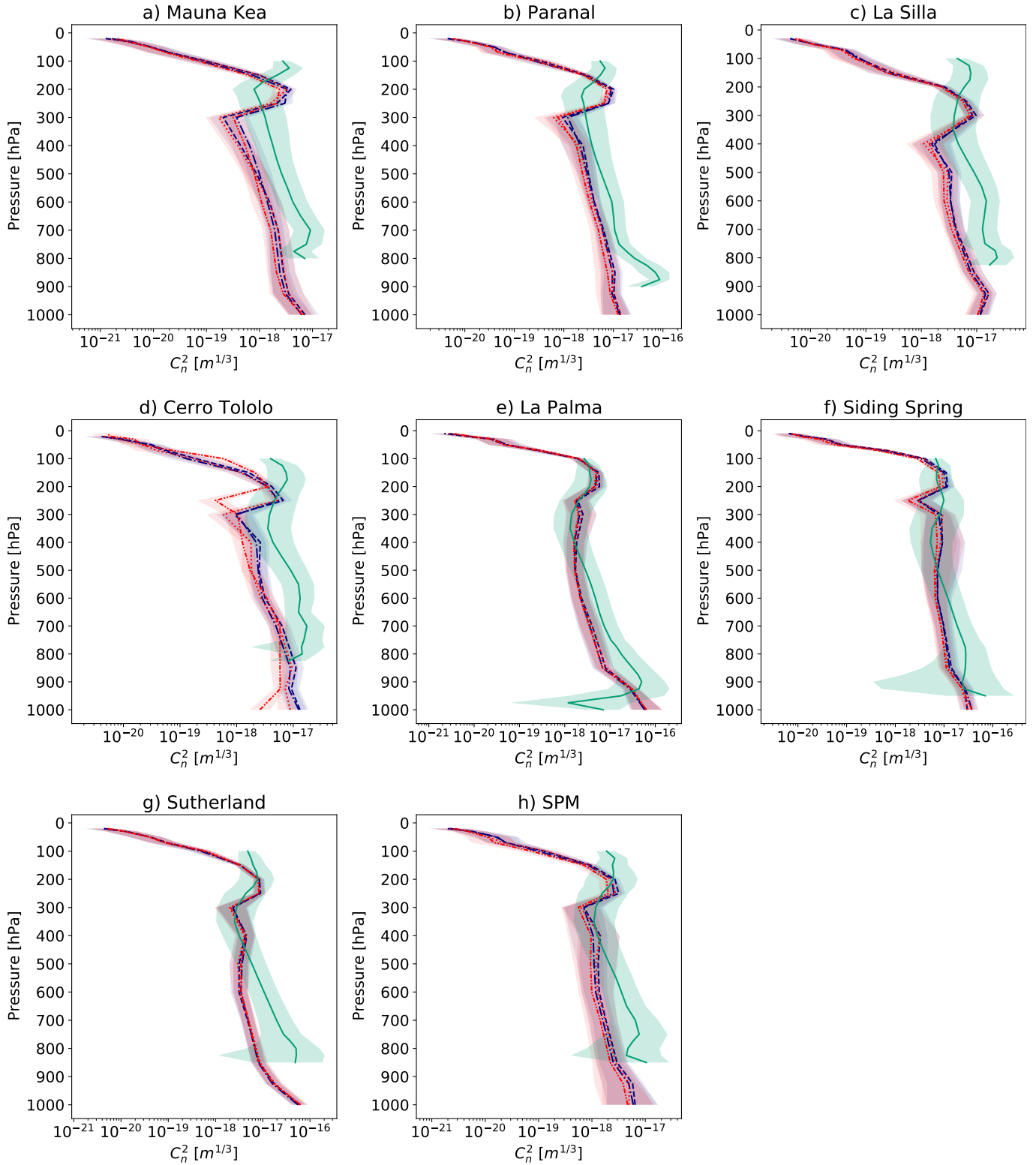


Fig. E.1. Vertical profile of the refractive index structure constant C_n^2 for all sites in ERA5 (1979-2019, green) and PRIMAVERA coupled-past (dashed blue line), atmos-past (dash-dotted blue line), coupled-future (red dotted line) and atmos-future (red dash-dot-dotted line) simulations. The higher the C_n^2 , the higher the seeing value. The lines indicate the median, and the spread is represented by the interquartile range (IQR, shading) based on monthly mean data. Results for PRIMAVERA are computed as the model ensemble mean of the monthly-mean median.

Appendix F: Vertical profile of specific humidity

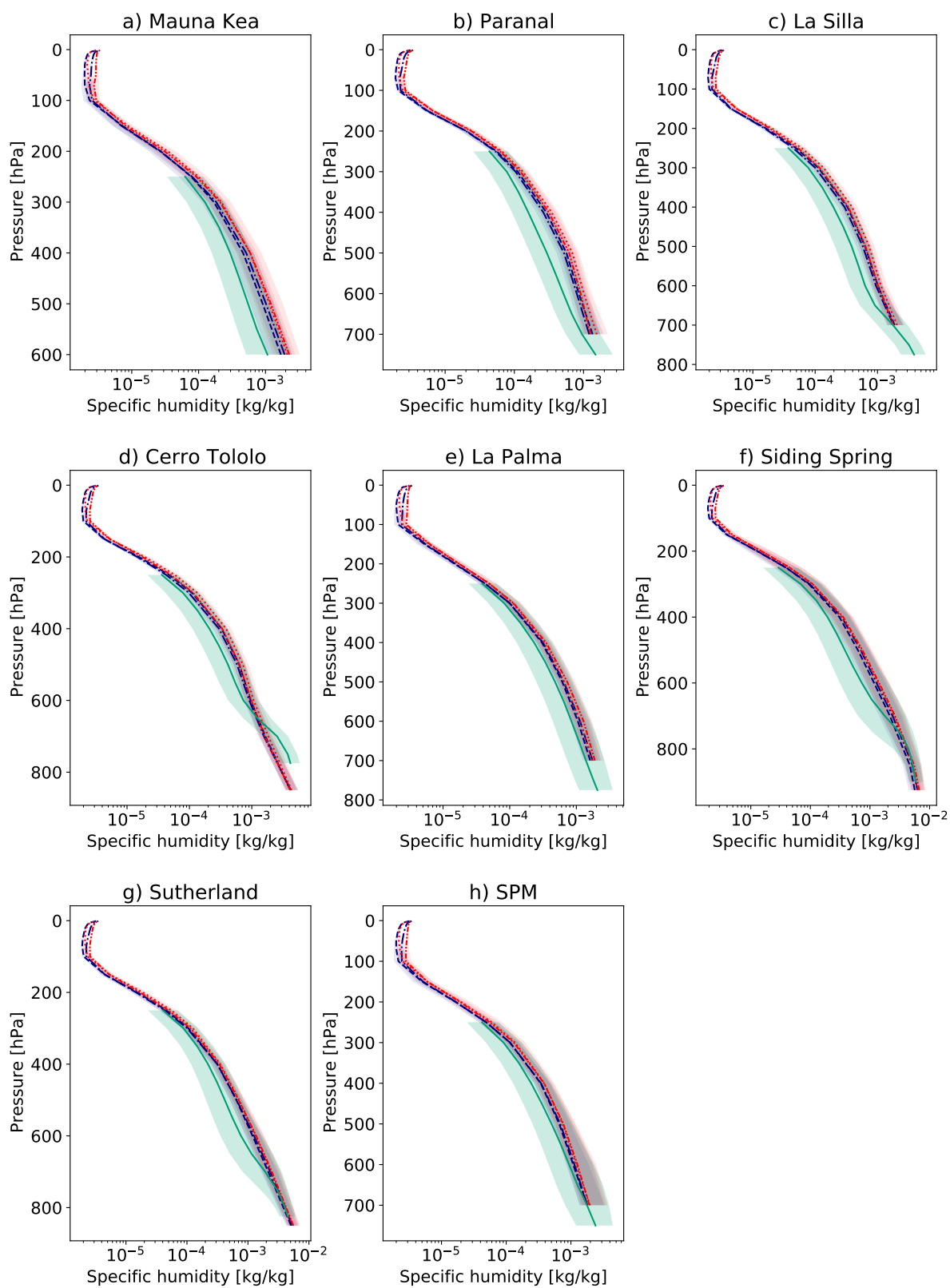


Fig. F.1. As Fig. E.1, but for specific humidity.

Appendix G: Vertical profile of horizontal wind speeds, temperature and geopotential height

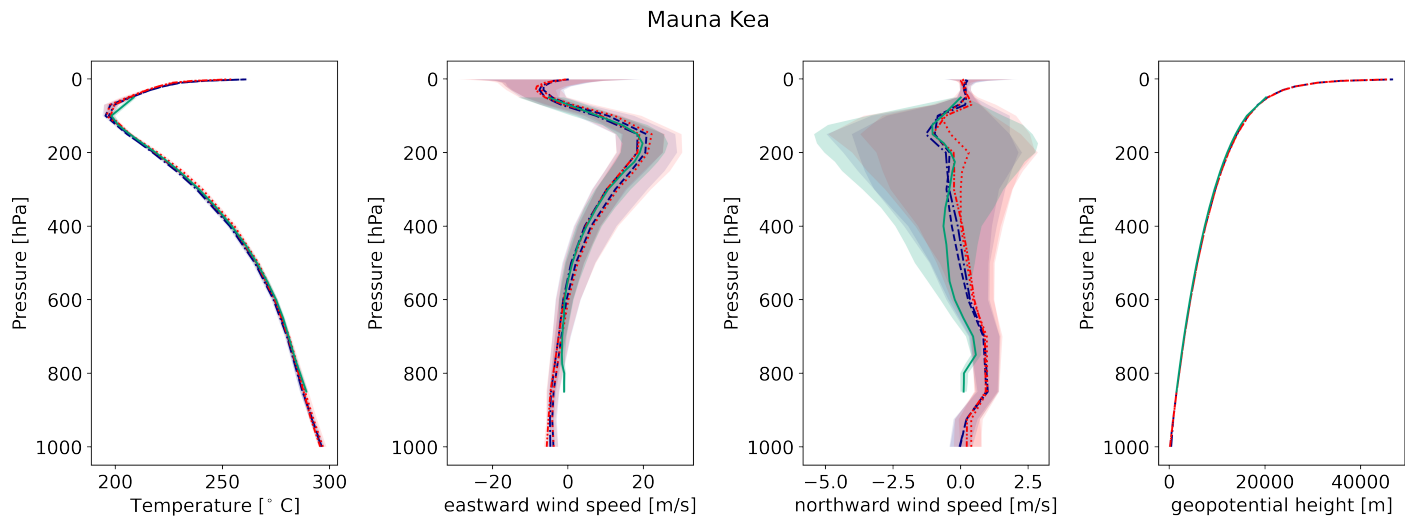


Fig. G.1. As Fig. E.1, but for temperature (left), eastward (middle left) and northward (middle right) wind components and geopotential height (right) of Mauna Kea.

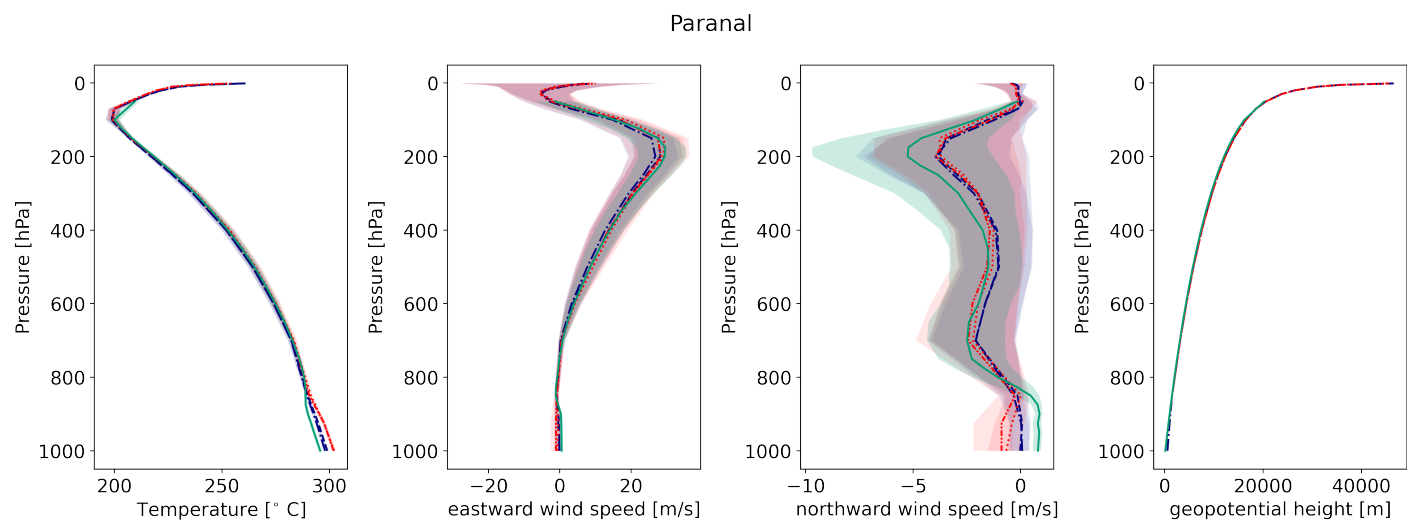


Fig. G.2. As Fig. G.1, but for Cerro Paranal.

La Silla

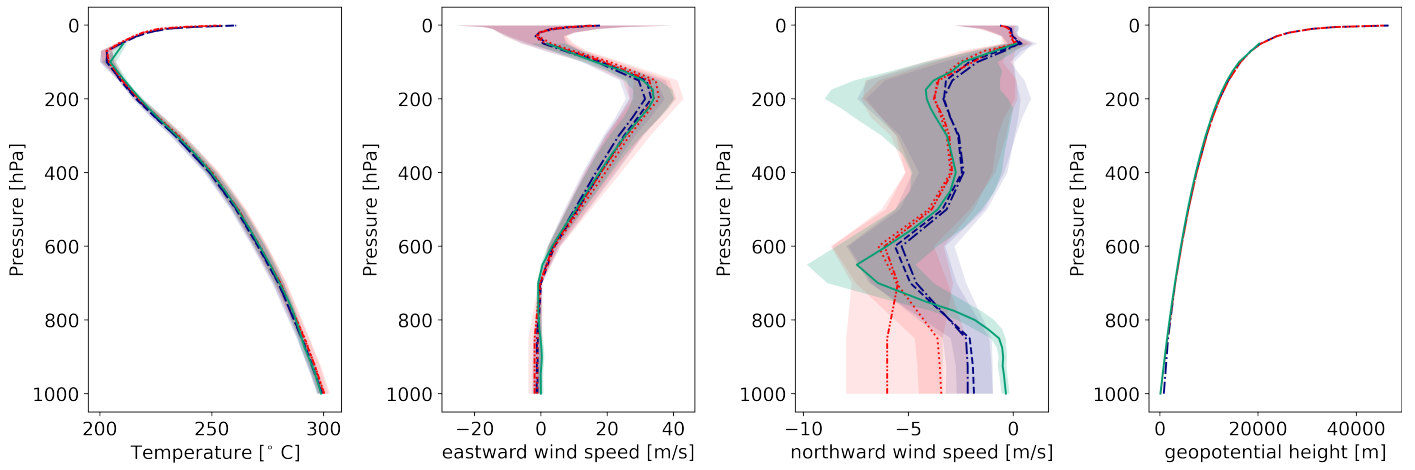


Fig. G.3. As Fig. G.1, but for La Silla.

Cerro Tololo

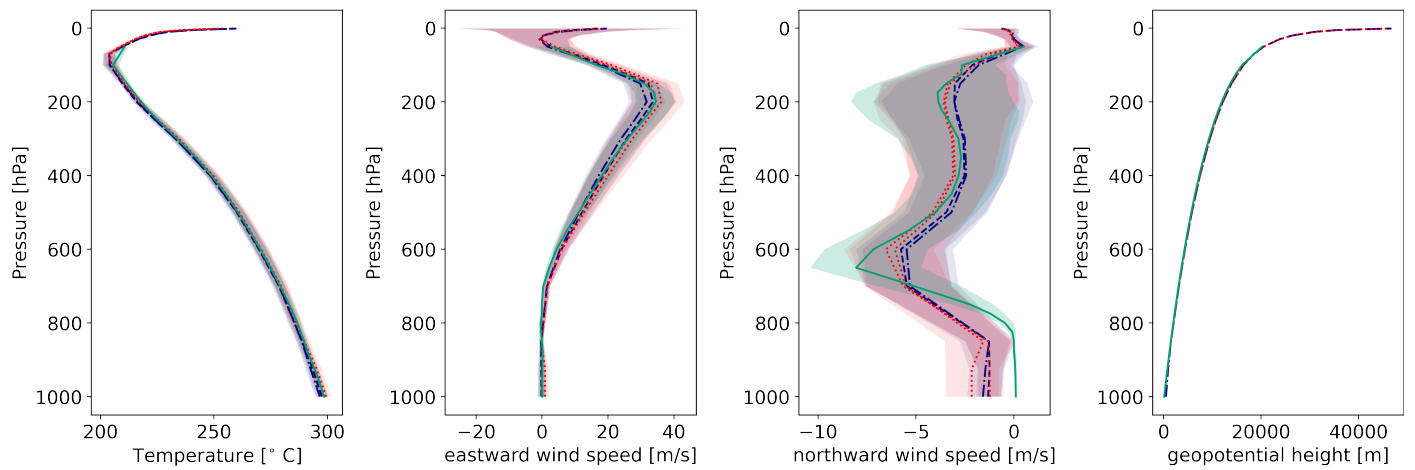


Fig. G.4. As Fig. G.1, but for Cerro Tololo.

La Palma

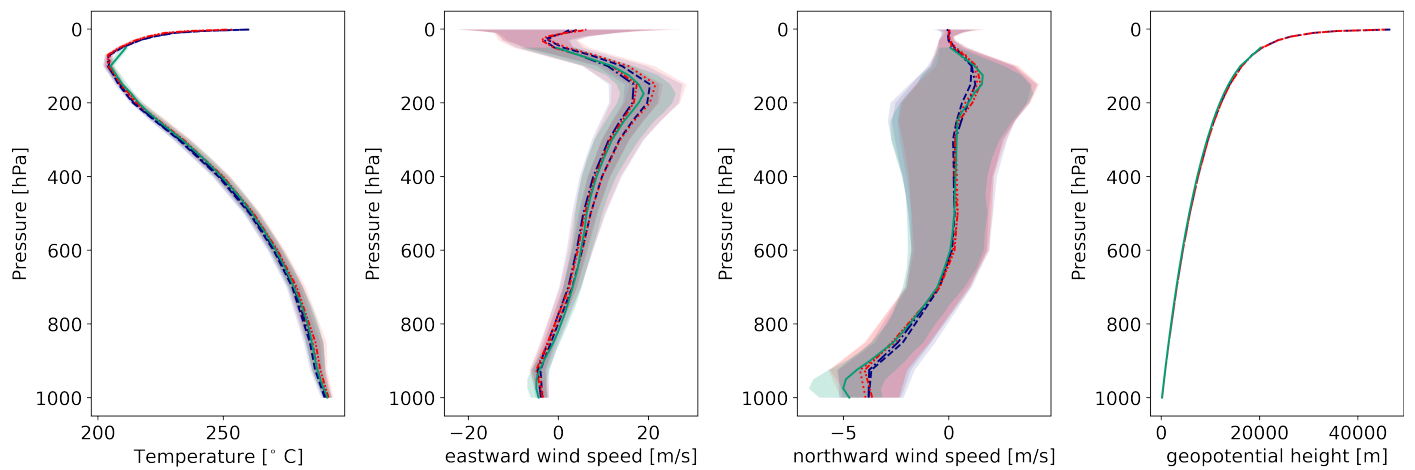


Fig. G.5. As Fig. G.1, but for La Palma.

Siding Spring

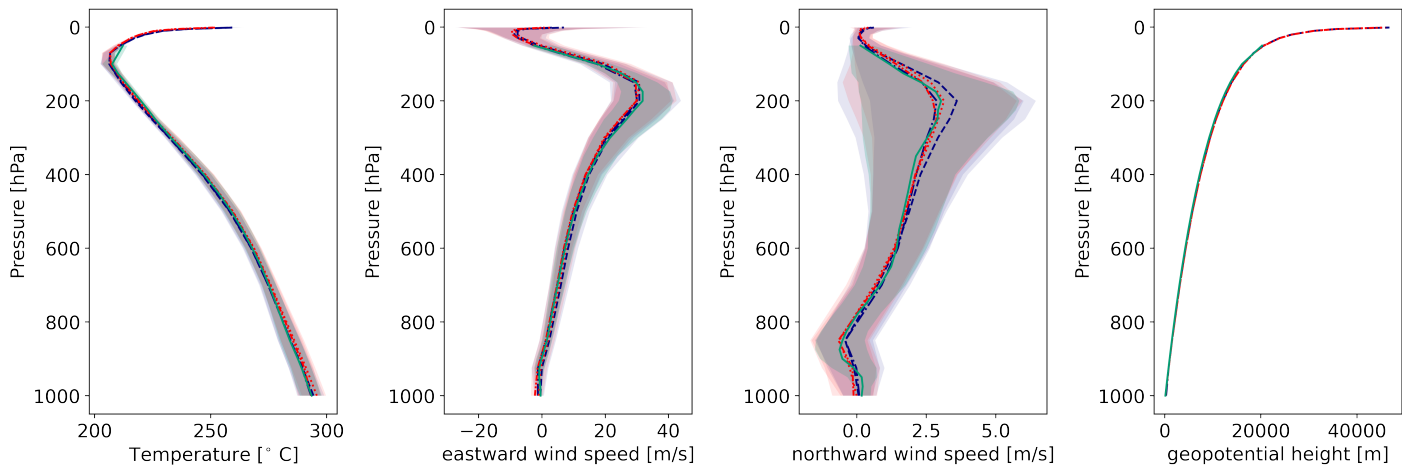


Fig. G.6. As Fig. G.1, but for Siding Spring.

Sutherland

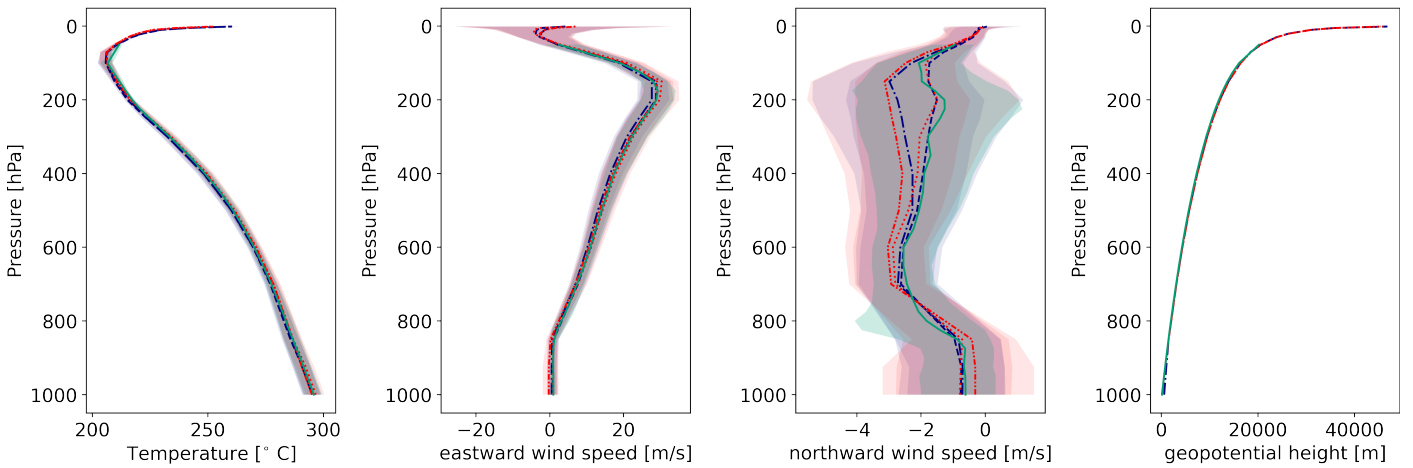


Fig. G.7. As Fig. G.1, but for Sutherland.

SPM

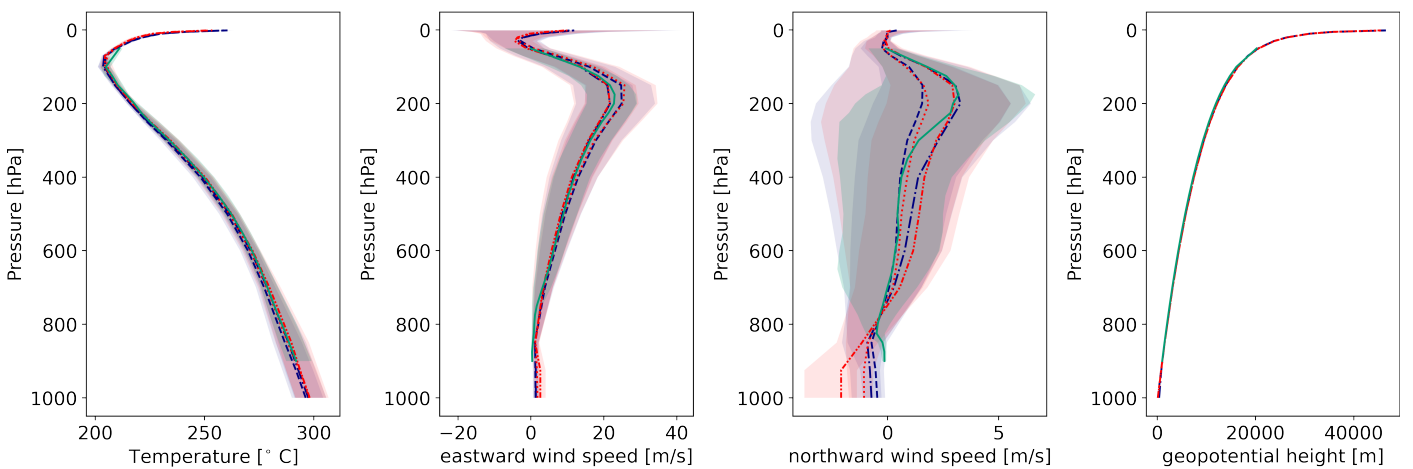


Fig. G.8. As Fig. G.1, but for SPM.