Characterizing and Mitigating the Impact of Telluric Absorption in Precise Radial Velocities

Sharon Xuesong Wang (王雪凇)1, Natasha Latouf2, Peter Plavchan2, Bryson Cale2, Cullen Blake3, Étienne Artigau4,5,6,7,8, Jonathan Gagné7,8, and Angelle Tanner10

1 Department of Astronomy, Tsinghua University, Beijing 100084, People’s Republic of China; sharonw@mail.tsinghua.edu.cn
2 George Mason University, 4400 University Drive, Fairfax, VA 22030, USA
3 University of Pennsylvania, Philadelphia, PA 19104, USA
4 Université de Montréal, Département de Physique, IREX, Montréal, QC H3C 3J7, Canada
5 Observatoire du Mont-Mégantic, Université de Montréal, Montréal, QC H3C 3J7, Canada
6 Johns Hopkins University, Applied Physics Laboratory, Laurel, MD, USA
7 Planétarium Rio Tinto Alcan, Espace pour la vie, 4801 av. Pierre-De Coubertin, Montréal, QC H1V 3V4, Canada
8 Institute for Research on Exoplanets, Université de Montréal, Département de Physique, C.P. 6128 Succ. Centre-ville, Montréal, QC H3C 3J7, Canada
9 Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA
10 Mississippi State University, Department of Physics and Astronomy, 355 Lee Boulevard, Mississippi State, MS 39762, USA

Received 2022 February 10; revised 2022 August 22; accepted 2022 September 22; published 2022 October 24

Abstract

Precise radial velocity (PRV) surveys are important for the search for Earth analogs around nearby bright stars, which induce a small stellar reflex motion with an RV amplitude of ~10 cm s⁻¹. Detecting such a small RV signal poses challenges to instrumentation, data analysis, and the precision of astrophysical models to mitigate stellar jitter. In this work, we investigate an important component in the PRV error budget—the spectral contamination from the Earth’s atmosphere (tellurics). We characterize the effects of telluric absorption on the RV precision and quantify its contribution to the RV error budget over time and across a wavelength range of 350 nm–2.5 μm. We use simulated solar spectra with telluric contamination injected, and we extract the RVs using two commonly adopted algorithms: dividing out a telluric model before performing cross-correlation or forward modeling the observed spectrum incorporating a telluric model. We assume various degrees of cleanliness in removing the tellurics. We conclude that the RV errors caused by telluric absorption can be suppressed to close to or even below 1–10 cm s⁻¹ in the blue optical region. At red through near-infrared wavelengths, however, the residuals of tellurics can induce an RV error on the meter-per-second level even under the most favorable assumptions for telluric removal, leading to significant systematic noise in the RV time series and periodograms. If the red-optical or near-infrared becomes critical in the mitigation of stellar activity, systematic errors from tellurics can be eliminated with a space mission such as EarthFinder.

Unified Astronomy Thesaurus concepts: Astronomy data analysis (1858); Exoplanet astronomy (486); Exoplanet detection methods (489); Earth atmosphere (437); Radial velocity (1332)

1. Introduction

Precise radial velocities (PRVs) are important for finding nearby Earth analogs that are the primary targets of future direct imaging missions (e.g., Howard & Fulton 2016; Dressing et al. 2019; Morgan et al. 2021). Recently, a number of new RV spectrographs with extremely high precision down to 10–30 cm s⁻¹ have come available, nearly able to detect Earth analogs around Sun-like stars (Wright & Robertson 2017). However, using a spectrograph capable of reaching 10 cm s⁻¹ does not mean achieving such a high precision on real targets, because there are still challenges in the data analyses and the mitigation of stellar jitter—the intrinsic RV signals produced by the stars themselves, unrelated to planets (Plavchan et al. 2015; Fischer et al. 2016; Halverson et al. 2016). Mitigating stellar jitter requires a suite of efforts, including making advances in the observational and theoretical understanding of a variety of stellar astrophysical phenomena (Crass et al. 2021). Similarly, reducing the error budget in the data analyses is also a multifaceted problem and includes tackling many numerical and statistical challenges interwoven with observational constraints. One of the most important and urgent issues in PRV data analyses is dealing with the Earth’s atmospheric contamination, considered one of the “seven pillars holding up the tent” of the overall PRV error budget (Fischer et al. 2016).

The Earth’s atmosphere imprints numerous absorption and emission lines (hereafter telluric lines) on top of an emitted stellar spectrum, and these lines do not exhibit the same Doppler shift that the stellar lines do due to the Earth’s spin and its motion around the Sun. Therefore, the blending between the telluric lines and the stellar lines would lead to errors when estimating the stellar Doppler shift. Telluric contamination has been a well-known bottleneck for achieving higher RV precision in the near-infrared (NIR; Bean et al. 2010) at the meter-per-second (m s⁻¹) level. However, in the optical, which is relatively free of telluric lines, there are in fact prevalent shallow water absorption lines called the “microtelluric” lines, and multiple bands of oxygen lines. These microtelluric lines (flux depths <2% and mostly <1%) at visible wavelengths can contribute to the RV error budget at the ~20 cm s⁻¹ level (Artigau et al. 2014; Cunha et al. 2014; Lisogorskiy et al. 2019).

Eliminating the impact of telluric lines in PRVs can be challenging for multiple reasons. While emission lines can, in principle, be removed via sky subtraction, it is still very
challenging to achieve a thorough removal due to the variability of these emission lines and the inherent noise in low signal-to-noise ratio (SNR) sky spectra obtained using a sky calibration fiber (e.g., low SNR). The only exception is when using spectrographs fed with diffraction-limited images through single-mode fibers, which would have a significant reduction in sky emission lines (as well as moon contamination; Crepp et al. 2016) thanks to its small point-spread function. More challenging yet are the absorption lines, especially in the NIR, where deep lines are common even within the few spectral windows in between the saturated lines. Early work often took observations of telluric standard stars (hot stars with relatively flat continua) to characterize and subtract the telluric absorption lines (Vacca et al. 2003), and the best effort so far on this front has achieved a residual of around 2% for water lines in the 0.9–1.35 μm window (Sameshima et al. 2018), but often around 5% or so elsewhere. However, those authors note that the time and spatial variability of water absorption puts stringent requirements on the telluric standard star observations.

There are several packages that could generate synthetic telluric absorption lines to fit the observations, e.g., TAPAS (Bertaux et al. 2014), Telfit (Gullikson et al. 2014), MolecFit (Smette et al. 2015), and other packages based on HITRAN (Rothman et al. 2013; Gordon et al. 2017). The best effort in modeling the observed spectra using synthetic lines are relatively successful, with a residual level of around 1%–2%, for example, in fitting a high-resolution, high-fidelity solar spectrum (Baker et al. 2020). Ulmer-Moll et al. (2019) compared several methods and also the most commonly used software packages to correct for telluric absorption lines using synthetic spectra, and their best results typically have residuals of around 3%–7% (1%–2% at the very best), on a similar level with that of Sameshima et al. (2018). Modeling residuals at this level could lead to an RV error on the order of 20–50 cm s$^{-1}$ or even more (Sithajan et al. 2016), similar to the impact of microtellurics in the optical region.

There are multiple challenges in improving the modeling precision of telluric absorption lines. Water lines, copious and highly variable in time (though less so along lines of sight; Li et al. 2018), are the most challenging lines to model. It is difficult to know the amount of precipitable water vapor (PWV) in the atmosphere during an observation a priori, which is an important model parameter governing the line depths. Although there are ways to measure PWV through remote sensing data from satellites (Meier Valdés et al. 2021) and using GPS (Baker et al. 2017), these methods are not yet precise enough to a submillimeter level. Other challenges with modeling the absorption lines, especially the relatively deeper ones, include imperfections in the line lists and line profiles (Seifahrt et al. 2010; Baker et al. 2020), as well as Doppler shifts of the lines due to the wind (Figueira et al. 2012).

Overall, telluric contamination is a large term in the PRV error budget that is challenging to mitigate, but this is also a term that can be eliminated completely by going to space, which is one of the motivations behind the NASA Probe Mission Concept Study, EarthFinder (Plavchan et al. 2020). As part of the EarthFinder study, our work set out to investigate the gain in terms of the RV precision improvement when eliminating telluric contamination completely by going to space. To be specific, our goals are: \(^{11}\) (1) to quantify the RV precision limit set by the telluric contamination in a broad wavelength range, from 350 nm to 2.5 μm, on a Sun-like star; (2) to characterize the effectiveness of commonly used methods for mitigating tellurics (i.e., dividing out the telluric lines and forward modeling). We perform simulations with synthetic spectra and extract RVs from these spectra to assess the RV precision limit set by the telluric contamination for ground-based instruments. With our controlled simulations, we isolate the effects of tellurics from other factors such as photon noise, stellar activity/jitter, and instrumental effects.

The paper is structured as follows. In Section 2 we describe our methods in synthesizing the observed spectral time series, extracting radial velocities while mitigating telluric lines, and combining RVs across different wavelengths. Section 3 describes our main results and the interpretation, including how tellurics enter the RV error budget across the wavelengths, what correlates with telluric-induced errors, the net contribution of tellurics to the RV error budget in different photometric bands, and how tellurics affect the RV time-series measurements. Section 4 enumerates the caveats in this study, summarizes our major conclusions, and outlines a few aspects of important future directions.

2. Methodology

Overall, our simulation aims to reproduce an ensemble of spectra of a Sun-like star through the Earth’s atmosphere throughout a year. We then extract RVs from these spectra using different telluric mitigation methods, and combine the RVs from different wavelengths through a weighted-average scheme. We then compare the RVs extracted using different methods and telluric mitigation strategies and compare their effectiveness. Figure 1 illustrates the workflow of our method.

2.1. Simulating a Year of Observed Spectra

We simulated a set of 365 spectra of a Sun-like star, one per night over a year. There are two spectral components in each of our simulated spectra: the stellar spectrum and the telluric spectrum. We use the Kurucz solar spectrum generated by ATLAS9 (Kurucz 2005), and we keep the solar spectrum constant throughout our simulation, i.e., no stellar line variation due to stellar jitter. We multiply the telluric absorption spectrum onto the solar spectrum using synthetic telluric spectra generated by TAPAS (Bertaux et al. 2014). The telluric spectrum is different for each simulated night, as described below.

There are two time-variable components in our simulated spectra:

(1) The telluric spectrum: We varied the depths of the telluric lines according to three things: the atmospheric temperature and pressure profiles of the observing season, the airmass of each observation, and the amount of PWV for each simulated night. We assumed the observations were taken on a telescope on Kitt Peak and adopted the corresponding atmospheric profiles. To mimic the change caused by atmospheric temperature and pressure profiles in the telluric spectra, we generated the telluric spectra using TAPAS for a selected date of each of the twelve months throughout the year of 2017 (usually the 15th of each month, unless the atmospheric profile data are unavailable). To be specific, for each month, we generated a suite of spectra for water, methane, carbon dioxide, nitrous oxide, oxygen, and ozone molecules separately, without additional line broadening or Rayleigh scattering, and at the

\(^{11}\) We limit the scope of our work to telluric absorption, as it is significantly more challenging to deal with than the emission lines.
最高分辨率和采样可能，在1.0的视宁度。我们注意到，只有深度线才会在季节内显示出可见的差异，这可能是因为TAPAS获取的气象条件的变化。

此外，我们随机从均值为5.0 mm，宽度为2.0 mm的高斯分布中抽取值来表示PWV，并相应地缩放了水线。PWV值被抽取成一个分布，该分布类似于2017年基特峰天文台的PWV值，如Suominet（例如，Wang et al. 2019）。12 我们将所有六种分子的光谱通过相乘组合起来，然后将透射光谱乘以相应的视宁度值。视宁度分布类似于RV标准星HD 10700，或泰坦，如Keck高分辨率横条光谱仪（例如，Butler et al. 2017）。图2显示了我们的输入PWV和视宁度值的直方图。

Figure 1. Flowchart illustrating our simulation methodology, including the generation of synthetic spectra and the RV extraction process using the cross-correlation function (CCF) and the forward-modeling (FM) methods with both correct and incorrect line profiles. The process of RV extraction is repeated each night for 365 nights for each method and line profile, and then they are combined to form a single nightly RV.

最高分辨率和采样可能，在1.0的视宁度。我们注意到，只有深度线才会在季节内显示出可见的差异，这可能是因为TAPAS获取的气象条件的变化。

此外，我们随机从均值为5.0 mm，宽度为2.0 mm的高斯分布中抽取值来表示PWV，并相应地缩放了水线。PWV值被抽取成一个分布，该分布类似于2017年基特峰天文台的PWV值，如Suominet（例如，Wang et al. 2019）。12 我们将所有六种分子的光谱通过相乘组合起来，然后将透射光谱乘以相应的视宁度值。视宁度分布类似于RV标准星HD 10700，或泰坦，如Keck高分辨率横条光谱仪（例如，Butler et al. 2017）。图2显示了我们的输入PWV和视宁度值的直方图。

12 https://www.suominet.ucar.edu/
around the solar system barycenter) of the “observed star” for that night following
\[
\sin(\text{date}/(2\pi \cdot 365.15) - \pi/2) \cdot 30 \text{ km s}^{-1},
\]
which roughly matches with the BC values throughout a year for Tau Ceti as observed at Keck. Then, we shifted the telluric spectra according to the BC value to offset the telluric spectra from the stellar spectrum in RV space to mimic the Earth’s motion. We chose to shift the telluric spectrum and leave the stellar spectrum untouched because the TAPAS synthetic spectra have a higher fidelity than the Kurucz solar spectrum, and thus this would minimize the numerical errors in shifting (interpolating and resampling).

Finally, we convolved the spectrum to a spectral resolution of \( R = 120,000 \), which is representative of the spectral resolution of the next-generation ground-based PRV instruments. We did not bin the spectrum further down to mimic the CCD pixels but chose to preserve the original binning of the Kurucz solar spectrum to avoid further numerical errors. As a result, the “pixel size” in our simulated spectra is about 0.0007–0.005 nm, a factor of three smaller than the typical CCDs used for PRV instruments today. We chose not to add photon noise or blaze functions in our simulated spectra to avoid adding more RV errors in order to isolate the effects of tellurics. Then, we extracted the RVs through two methods, using the cross-correlation method or forward modeling.

2.2. Extracting RVs and Correcting for the Telluric Absorption

We extracted the RVs from our simulated spectra with two methods that are commonly employed in the community: the cross-correlation function (CCF) method, where RVs are estimated from the CCF between the observed and a template spectrum (or a mask), or the forward-modeling method, where the observed spectra are modeled with input spectral templates and several free parameters, including the Doppler shift. In each method, we used a perfect stellar template that is identical to our input stellar spectrum (i.e., the synthetic solar spectrum) to minimize additional errors unrelated to the tellurics. Whenever needed, we assumed perfect knowledge of the wavelength solution of the observed spectra and the spectral point-spread function (PSF; also called the line-spread function or the spectrograph response function), an assumption not too far from reality for the latest generation of ultrastabilized RV spectrographs calibrated by laser frequency combs. With each method, we apply a set of telluric mitigating strategies with various degrees of accuracy in eliminating tellurics.

Before extracting the RVs, we divided each spectrum into 230 chunks (“orders”), from 0.3 nm at 350 nm to \( \sim 20 \) nm at 2.5 \( \mu \)m, with the chunk size growing linearly versus wavelength. The change of the chunk size is to mimic the length of a typical spectral order and also to compensate for the decrease of Doppler content toward the NIR. Computing RVs in the unit of spectral orders is also convenient for characterizing the RV errors induced by tellurics as a function of wavelength. This is also close to reality where RVs are typically computed independently for different spectral orders and then combined together later on.

We describe each method and the associated mitigation strategy in detail below:

1. No correction: RVs are extracted via the CCF method as first described by Baranne et al. (1996), but using a stellar template instead of a mask and stepping through the RV space instead of fitting a Gaussian to the CCF, similar to Zechmeister et al. (2018).13 We do not perform any correction on the tellurics, so this method characterizes the net contribution of tellurics to the RV error budget when left untreated. We note that using a forward-modeling method gives the same results, so we simply present the results from the CCF method here for the “No correction” method.

2. CCF after telluric division (referred to as “CCF with division” or “Division below”): RVs are extracted via the CCF method, but before the CCF is computed, a telluric spectrum is removed from each order via mathematical division. The telluric spectrum being divided out is exactly the same as the one used for generating the simulated spectrum for each night, but convolved to \( R = 120,000 \) to match the resolution of the observed spectrum. This approach is the most commonly used

13 We had performed a sanity check on our CCF algorithm to ensure that an RV of \( 0 \) m s\(^{-1} \) is returned from all orders when extracting RVs from a simulated solar spectrum with no tellurics. This confirms that we do not introduce numerical errors beyond the machine precision or due to any edge effects in our choice of chunking up the spectra or in our CCF algorithm.
reported by Ulmer-Moll et al.\cite{Ulmer-Moll2019}. Maunakea line by a power law. The rms of the residual of this simulated spectra is 2%, which is around the typical value for all lines and smaller than the typical residuals (2\%–5\%) reported by Ulmer-Moll et al.\cite{Ulmer-Moll2019}.

...method to mitigate tellurics in PRVs for stabilized spectrographs such as HARPS and SPIRou (Artigau et al.\cite{Artigau2021}; Cretignier et al.\cite{Cretignier2021}).

3. CCF after telluric division, but with an incorrect telluric line profile (“CCF with wrong profiles” or “Division K-profile”): RVs are extracted in the same way as the CCF with division method described above. However, instead of using the exact same telluric spectrum as the input, we use a spectrum with a different line profile setting than the input telluric spectrum. Our input telluric spectrum was generated assuming the atmospheric conditions over Kitt Peak and, here in this method, we divide out a spectrum generated using conditions over Maunakea and hence differing in the underlying line profiles. We scale the Maunakea telluric spectrum to match the line depths in the simulated observed spectrum to mimic a modeling process,\cite{Wang2022} with the exception of ozone absorption, which we omit completely from the division, because the ozone absorption is more like a modulation on the continuum level and is very hard to know a priori. A comparison of the two line profiles is shown for an example telluric line in Figure 3. The difference between the two sets of lines is mostly due to atmospheric pressure difference effects between the two sites. Introducing such a difference is also similar to having the line width wrong. This introduction of a line profile mismatch is to mimic our lack of knowledge on the telluric lines, similar to having a modeling residual of around 1\% for typical lines (somewhat smaller than the best effort in the literature, e.g., Ulmer-Moll et al.\cite{Ulmer-Moll2019}).

Figure 3. Comparison of the line profiles for a CO2 line near 1602 nm for two observatories: Kitt Peak (used in simulated spectra) and Maunakea (used as input to fit the observed spectrum). Both spectra are at airmass = 1. The Maunakea scaled spectrum is the best-fit version when fitting the Kitt Peak profile by scaling the Maunakea line by a power law. The rms of the residual of this fit is 1.2\%, which is around the typical value for all lines and smaller than the typical residuals (2\%–5\%) reported by Ulmer-Moll et al.\cite{Ulmer-Moll2019}.

4. Forward modeling with the perfect input telluric spectrum: RVs are extracted via the forward-modeling method as described in Butler et al.\cite{Butler1996}. This method builds a model using the stellar template and the synthetic telluric spectra (both the same as used in simulation input), with the only free parameter being the Doppler shift (RV), to fit the observed spectrum via a least-squares minimization algorithm.\cite{Wang2022} We again assume perfect knowledge on the wavelength solution and the spectral PSF. This method serves as the sanity check for our work and, as expected, it returns 0 m s$^{-1}$ (to a precision of 10$^{-15}$ m s$^{-1}$) in all orders across all nights, meaning no RV errors. We do not show the results from this method for the rest of the paper.

5. Forward modeling with an incorrect telluric line profile (“modeling with wrong profiles” or “modeling K-profile”): RVs are extracted following the same methodology as described above, however we again introduce a different telluric line profile than the one used to generate the synthetic spectra. Relatively similar to the CCF with wrong profiles method, we first estimate a set of best-fit line depths for the six molecules in our telluric template spectra (with the wrong line profile) by performing a first iteration forward-modeling fit to the observed spectrum. In this first iteration, the telluric line depth for each molecular species is a free parameter in the modeling, but only for spectral orders where the Doppler information content in the lines of this particular molecular species exceeds 1 m s$^{-1}$ (estimated following Butler et al.\cite{Butler1996} and assuming SNR = 1000 per pixel). As in “CCF with wrong profiles,” we ignore the ozone completely. If the lines do not carry enough Doppler information (e.g., too shallow or too few), we freeze the line depth to zero and do not attempt to retrieve the line depth for this species from this particular order. Next, all the best-fit line depths are averaged to report a final depth for each species, which are then fixed for all orders in the second (and final) iteration of forward modeling. We adopt this two-step modeling because we find a

\footnote{This is equivalent to knowing the line depths to some degree (though not perfectly) prior to the final RV extraction, a fairly common situation in reality (e.g., Artigau et al.\cite{Artigau2021}).}

\footnote{We use the \texttt{LMFIT} package in Python with the L-BFGS algorithm, which we observed to have the best convergence properties for our calculations.}
significant improvement in the RV precision by adding a second iteration.

2.3. Combining RVs from Different Orders

After RV extraction, we have a matrix of RVs across 230 orders and 365 nights. We then combine the RVs of the 230 orders from a single night into one reported RV value for the night. The combination is done through a weighted average, and the weight for each order consists of two components added in quadrature: the RV scatter across 365 nights for this order (caused by tellurics) and the photon-limited RV error. We bring back photon errors here so that the weights are more realistic, because otherwise the spectral orders with tellurics will be downweighted significantly, eliminating errors induced by tellurics almost completely. The RV scatter caused by tellurics is simply estimated by computing the standard deviation of the RVs of each order across 365 nights. The photon-limited RV error is estimated following Butler et al. (1996) and assuming SNR = 100 per pixel, similar to the typical SNR achieved per resolution element in typical real RV observations. At this SNR, the photon-limited RV error for an order is typically on the order of 1–10 m s⁻¹, and about half of the orders have the photon-limited RV error as the dominant term in their weights. Such an weighted averaging scheme represents the typical strategy in combining RVs across multiple orders.

3. Results and Interpretation

Overall, there are two ways that tellurics contribute to the RV error budget. One is through peak pulling (e.g., Wright et al. 2013), where the telluric contamination lines would “pull” the stellar spectral lines when performing CCF or modeling, as illustrated in Figure 4. In other words, the existence of a telluric line changes the slope of the stellar line, which is the origin of the Doppler information in a line, and this leads to a biased measurement of the RVs. Any residuals in dividing or modeling the tellurics would cause RV errors in a similar fashion. This is very analogous to how the sky/solar contamination affects RV precision (Roy et al. 2020).

The second way is more subtle and only occurs when dividing out the telluric lines. One of the common treatments for tellurics when performing CCF technique to extract RVs is to divide out a telluric model, as we do in our simulations (e.g., Artigau et al. 2014). However, this is mathematically incorrect, because the stellar+telluric spectrum is a convolution between the spectral PSF (IP) and a product between the unconvolved stellar spectrum, S, and the unconvolved telluric spectrum, T—i.e., multiplication first and then convolution, or the observed spectrum, can be expressed as (S · T) * IP. Mathematically, convolution does not distribute over multiplication, or (S · T) * IP ≠ (S * IP) · (T * IP), as illustrated in Figure 5. Therefore, dividing by a telluric model (T * IP) will not result in a clean stellar spectrum (S * IP).

As a result, using division to remove telluric contamination, even if using a perfectly accurate telluric model, will lead to additional errors in the spectrum and RV errors. Such an error is larger for deeper lines and is thus more severe for the NIR region. This division error does not exist in the forward-modeling results, as the algorithm follows the correct order of multiplication then convolution when modeling the observed spectra. We note that since our sampling scale is smaller than the typical CCD pixel scales in real instruments, these convolution+division errors might be underestimated. This error also increases with a lower spectral resolution.

In the following subsections, we first take a detailed look into how tellurics affect the RVs as a function of wavelength, and then we take a more holistic view of the tellurics’ overall contribution to the RV error budget under different scenarios. We then discuss the impact of wind shifts in oxygen lines and why they should not impact our results significantly.

3.1. Errors Induced by Tellurics versus Wavelength

As we divide the stellar spectrum into 230 orders, we first take a look at how tellurics affect the RVs in each order. Figure 6 is an illustration of the systematic RV errors when tellurics are ignored in two different orders—one in the optical and one in the NIR. We choose to plot RV versus BC instead of day of year because any systematic errors as a result of the line-pulling effect would manifest more directly as a correlation between the RVs and the BCs. This is due to the fact that the BCs dictate the position of the stellar lines on the detector, since usually the bulk of a star’s Doppler motion comes from the barycentric motion (∼±30 km s⁻¹ in amplitude).

The right panels of Figure 6 show that there are two ways that the tellurics enter the RV error budget, similar to the effects of solar contamination (Roy et al. 2020): first, they add scatter to the measured RVs; and second, they add a highly coherent systematic signal. This coherent systematic signal has a characteristic width in the wavelength space of a few kilometers per second, which is the typical stellar line width, as a result of the peak pulling effect, as the telluric lines “scan” back and forth against the stellar lines in the stellar rest frame during the course of one solar year’s worth of observations. When the stellar lines dominate over the telluric lines in terms of the Doppler information content, as in the top panels of Figure 6, the systematic errors are small, and vice versa (as shown in the bottom panels).

We take the standard deviation of the RVs reported for each order over a year as a metric to describe the scatter caused by tellurics at different wavelengths. The three panels in Figure 7 illustrate the RV scatter as a function of wavelength for various RV extraction methods and telluric mitigation strategies. Each point plotted is the standard deviation16 of the RVs reported by

---

16 We note that using the median absolute deviation (MAD) gives consistent results, similarly for the RV scatter quoted in the tables and plots in the next subsection as well.
a single order across a year (i.e., the scatter in the right panels of Figure 6). Overall, the amplitudes of the RV scatters trace the telluric bands—in wavelength regions where the telluric absorption is heavy (e.g., the places between the NIR photometric bands), the RV scatter is significantly larger than in the others. The optical region experiences less RV scatter than the NIR, as expected.

Dividing out the telluric absorption lines clearly improves the RV precision across the entire spectrum, but the CCF division method does not remove the effects of tellurics completely, even though it was a division using a perfectly accurate telluric model. This is a result of the residual “convolution-division” errors, as illustrated in Figure 5.

The top right panel of Figure 7 compares the RV scatters using the forward-modeling method with the CCF method, both using the wrong line profiles. The RV scatters of the no-correction method are also plotted as a fiducial baseline for comparison. Note that we did not plot the results for the forward-modeling method with no correction, because it produces the same results as the CCF with no treatment method, just as expected. The results for the forward modeling with the perfect telluric model are also not plotted, because this method produces essentially 0 m s$^{-1}$ (within machine precision) for all orders across all simulated spectra (i.e., it completely eliminates the effects of telluric contamination). It outperforms the CCF with perfect division because forward modeling is not subject to the convolution-division error. For this same reason, the forward modeling with the wrong profiles method outperforms CCF with the wrong profiles in a typical order with moderate to deep telluric lines.

The bottom panel of Figure 7 shows an overall comparison for all methods, with the RV scatter values now smoothed over five orders and plotted as lines for clarity. In particular, CCF division (with perfect lines) outperforms forward modeling with the wrong profiles in regions where telluric lines are shallow, especially in the optical region. This is because, when the telluric lines are shallow, the convolution-division error is small compared with the residual error caused by the wrong line profile. As shown in Figure 5, the rms of the convolution-division error is much smaller than 1% for shallow lines (and scales roughly with line depths), while the line profile mismatch induces a typical rms of about 1% (Figure 3). For a similar reason, the CCF division and the forward-modeling methods essentially perform equally well in the optical region bluer than 600 nm when using the wrong line profile. Another reason why the CCF division (with perfect lines) performs better than the other methods in the optical region blueward of 700 nm is that the continuum absorption caused by O$_3$ was divided out completely, as we used the input telluric model. By contrast, the ozone absorption was ignored in the CCF division or forward-modeling methods with the wrong line profile. It is beyond the scope of this work to constrain the absorption of ozone in observations, which is equivalent to a very small change in the continuum level in the observed spectrum.

### 3.2. What Correlates with Telluric-induced RV Errors

In this section, we examine the effects of telluric contamination on the final RVs calculated by combining RVs from all orders. How the RVs are combined is described in Section 2, and, briefly, it is a weighted average with the weights estimated using the RV precision limit set by the photon-limited Doppler content and the RV scatter caused by tellurics.

The first two rows of plots in Figure 8 show the correlation between the final RV (a weighted sum across our entire spectral range) versus the airmass or PWV for each night. The RVs from the CCF division method are not plotted as they show similar patterns. The top plots of Figure 8 show that the amount of scatter in the RVs is only very weakly correlated with the airmass. This is because the final RVs are dominated by the RVs reported in the optical region (350–600 nm), especially the bluer part, with ample solar lines and relatively few telluric lines. The telluric...
lines in the blue/green optical region are basically all water lines, whose depths are mostly governed by the PWV. PWV variations across nights can be easily more than a factor of two, much more than the change in airmass. This is also reflected in the middle plots of Figure 8, where the RV scatter is clearly correlated strongly with the PWV.

In the bottom plots of Figure 8, we show the final RVs versus the barycentric velocity, BC, with the PWV color coded. The telluric contamination induces a small coherent systematic error with an amplitude of order a centimeter per second in the RVs from all three methods. BC stands out as the variable that has the most clear and strongest correlation with the RV error, with PWV as the second one. Airmass does not seem to have a strong correlation with the RVs even in the regions where there are no water lines (that is, if we plot the same set of plots with RVs derived using spectra from a water-free region instead of our entire wavelength range, the top plots would still look similar).

There are two reasons why the amplitude of this systematic error did not decrease significantly with our treatment of telluric lines, and it might seem surprising that the amplitude of the RV systematics in the forward-modeling method (lower right panel) is even larger than that for the no correction or CCF division method. This is due to our adopted weighting scheme. First, the final RVs are dominated by the best-behaving orders in the blue/green part of the optical region, where CCF or forward modeling with the wrong profiles perform similarly to the CCF no correction method, as shown in the bottom panel of Figure 7. Second, CCF or forward modeling with the wrong profiles did not correct for the continuum absorption in the optical region caused by the ozone, which is the dominant source of systematics when water lines are absent in the blue. This also means that our results and conclusions would basically remain unchanged if we use a different set of randomly generated PWV and airmass values.

Why do the systematic patterns from different orders not cancel out after combining the RVs? As illustrated in Figure 6, different orders exhibit different systematic errors with varying patterns and amplitudes. The characteristic widths of these systematic signals are similar, though not the same, across all orders, since they are tied to the line widths. Therefore, when adding all these systematic errors together, they persist as a coherent systematic error. This is analogous to adding a finite series of quasi-periodic curves with similar periods but varying amplitudes and phases, and the weighted average of these waves would have a coherent structure as well (with a reduced amplitude).
The high correlation between RV, BC, and PWV shown in the lower left panel of Figure 8 is intriguing: this means that if the stellar spectrum and the telluric absorption lines in a series of observations are precisely known, then the RV bias induced by tellurics could in principle be predicted via simulations like this work. However, this could be hard to implement in reality as it is hard to know the stellar and telluric spectra very well, but it nonetheless presents a possible alternative path forward in mitigating tellurics.

### 3.3. Tellurics’ Contribution to the RV Error Budget in Different Photometric Bands

Next, we examine the total amount of scatter in the RVs induced by tellurics over a wavelength range. We first see how much the telluric contamination would contribute to the RV error budget for several new generation high-precision RV instruments, as listed in Table 1. We caution the readers that there are several important assumptions here that would make such estimates not accurate for these instruments, but here they serve to make a comparison between them instead. First, we did not consider the various instrumental resolutions, which typically differ from the resolution of our simulated spectra. Second, we assume an SNR = 100 per pixel in our simulated spectra when combining the RVs, while real observations would have a variety of SNRs (especially given that the resolution and the pixel sampling factors of different instruments also differ).

Comparing the first column, “No Correction,” with the column listing the photon-limited precision in Table 1, it is evident that for the instruments operating in the broad optical region, tellurics do not contribute significantly to the error budget, assuming proper weighting is applied when combining the RVs. In the NIR, however, tellurics affect the RV precision significantly. Even after dividing out or modeling the tellurics, the RV precision still could not reach the photon-limited precision, with a scatter of about 1 m s$^{-1}$. Overall, the three mitigation methods perform on a similar level, with CCF division outperforming the other two (again recall that forward modeling with perfect tellurics eliminates the telluric-induced errors completely). This is consistent with the results shown in Figure 7.

If we take a more careful look and see how the tellurics affect the RV precision in narrower bands, we would see that it is not all bad in the NIR. Table 2 lists the RV scatter caused by tellurics in several commonly used photometric bands. I, Z, and Y bands in the short NIR could actually reach very close to the photon-limited precision after treating the tellurics. Interestingly, in these three bands, modeling with the wrong profile is worse than division with the wrong profile in certain bands, which seems at odds with the results for different orders in Figure 7. This is because of weighting.

In the I, Z, and Y bands, the division performs as well as the modeling in regions almost free of tellurics. In regions with tellurics, division is significantly worse, so these orders receive very low weights in Division. However, the same orders received higher weights, as they were not as worse and
contributed more significantly to the final RVs of Modeling, making its RV rms worse. For example, in the Z band, there are several orders that are reporting a roughly meter-per-second precision (840–900 nm) for both division and modeling, and the rest of the orders are typically 10 m s\(^{-1}\) for modeling, but ∼50+ ms\(^{-1}\) for division. If we put more equal weights between the orders or weight the orders only by their RV scatter from tellurics, modeling would report better results than division, as expected by looking at Figure 7. However, such a weighting scheme is rarely adopted in real practice: equal weighting is certainly far from optimized, and weighting only by the telluric-induced errors is unrealistic, because it is hard to distinguish from the total RV scatter how much is the true photon-limited error versus the residual impact of tellurics in real observations. Still, the fact that forward modeling does not outperform the CCF method, as we would have expected based on Figure 7, suggests that our weighting scheme, a commonly adopted one in real practice, can be improved.

Both Tables 1 and 2 list the standard deviation of the combined RVs. If we use the median absolute deviation instead of the standard deviation, the scatter would decrease considerably, by up to a factor of two. This is not surprising since the RV scatter is not Gaussian but a systematic pattern as shown in the bottom panels of Figure 8. Therefore, different summary statistics would report a different value for the RV scatter. To capture this complication and to provide a straightforward illustration, we summarize the RV variations across different photometric bands in Figure 9, where both the range of the RV systematics (95\% percentile) and the standard deviation of RVs in any band are plotted as the two ends of each vertical line. Interestingly, our results in the Y, J, and H band using the K-profile methods are consistent with the findings of Hirano et al. (2020), where they investigated the effects of inaccurate telluric modeling and concluded that it adds 1-2 m s\(^{-1}\) to the RV scatter, and they also found that the scatter is the worst in the J band.

### 3.4. How Telluric-induced RV Errors Manifest in Time Series

Similar to how telluric-induced RV errors manifest as coherent systematic patterns in the RV–BC plane, these errors appear as correlated noise with structures in the final RV time
Table 1
Summary of the RV Scatters for Different Instruments using Different RV Extraction Methods

| Instrument (wavelengths) | No Correction (m s\(^{-1}\)) | CCF Division (m s\(^{-1}\)) | Division K-profile (m s\(^{-1}\)) | Modeling K-profile (m s\(^{-1}\)) | Photon-limited Precision (m s\(^{-1}\)) | Instrument Precision\(^{a}\) (m s\(^{-1}\)) | Std. Star Precision\(^{a}\) (m s\(^{-1}\)) |
|--------------------------|-------------------------------|-----------------------------|---------------------------------|---------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|
| EFE Visible Arm (380–900 nm) | 0.035 | 0.021 | 0.021 | 0.032 | 0.059 | ... | ... |
| EFE NIR Arm (900–2500 nm) | 2.432 | 0.761 | 1.196 | 1.220 | 0.143 | ... | ... |
| ESPRESSO (380–788 nm) | 0.034 | 0.020 | 0.018 | 0.023 | 0.062 | 0.1\(^{b}\) | ~0.5\(^{b}\) |
| EXPRESS (380–680 nm) | 0.031 | 0.013 | 0.021 | 0.024 | 0.061 | 0.1\(^{c}\) | ~0.5\(^{c}\) |
| NEID (380–930 nm) | 0.034 | 0.019 | 0.021 | 0.033 | 0.059 | 0.3\(^{d}\) | 0.76\(^{d}\) |
| CARMENES Visible Arm (520–960 nm) | 0.169 | 0.098 | 0.111 | 0.179 | 0.107 | 1.7\(^{e}\) | 1–2\(^{e}\) |
| CARMENES NIR Arm (960–1710 nm) | 2.359 | 0.659 | 1.161 | 1.144 | 0.229 | 3.7\(^{e}\) | 3–4\(^{e}\) |

Notes. EFE stands for EarthFinder Equivalent on the ground. The RVs from different orders were combined using a weighted average, taking into account both the scatter caused by the tellurics and the photon-limited RV precision assuming an SNR of 100 per pixel using a Kurucz solar spectrum.

\(^{a}\) By "Instrument Precision," we refer to the internal precision achieved by the hardware, and the "Std. Star Precision" refers to the best or typical on-sky performance reported by each instrument observing RV standard stars over a relatively long time baseline.

\(^{b}\) From Pepe et al. (2021), and the standard star precision is the on-sky rms over a couple of months.

\(^{c}\) Instrument precision from Blackman et al. (2020). The standard star precision is the lowest RV rms in a suite of standard star observations over a few months as in Zhao et al. (2022).

\(^{d}\) Instrument precision from the design estimate in Halverson et al. (2016). The quoted standard star precision is the solar rms over a few months from Lin et al. (2022), and the solar nightly binned RV rms is 41 cm s\(^{-1}\).

\(^{e}\) From Bauer et al. (2020), and the standard star precision is taken as the lowest RV rms in a distribution of the CARMENES survey stars.

Table 2
Summary of RV rms Precision Lower Limits for Several Common Photometric Bands

| Photometric Band | No Correction (m s\(^{-1}\)) | Division (m s\(^{-1}\)) | Division K-profile (m s\(^{-1}\)) | Modeling K-profile (m s\(^{-1}\)) | Photon-limited Precision (m s\(^{-1}\)) |
|-----------------|-----------------------------|-------------------------|---------------------------------|---------------------------------|--------------------------------------|
| B (398–492 nm) | 0.022 | 0.001 | 0.008 | 0.008 | 0.094 |
| V (507–595 nm) | 0.167 | 0.036 | 0.072 | 0.072 | 0.219 |
| R (589–727 nm) | 0.446 | 0.283 | 0.358 | 0.454 | 0.285 |
| I (732–881 nm) | 0.968 | 0.427 | 0.645 | 0.746 | 0.396 |
| Z (0.8–1.0 \(\mu m\)) | 1.283 | 0.790 | 0.541 | 1.210 | 0.416 |
| Y (0.96–1.1 \(\mu m\)) | 2.414 | 0.778 | 1.272 | 1.217 | 0.993 |
| J (1.1–1.3 \(\mu m\)) | 5.816 | 2.821 | 5.193 | 3.473 | 1.027 |
| H (1.5–1.8 \(\mu m\)) | 5.789 | 1.169 | 1.597 | 1.586 | 0.610 |
| K (2.0–2.4 \(\mu m\)) | 58.358 | 2.187 | 7.686 | 6.778 | 1.093 |

Note. The RVs were combined in the same fashion as in Table 1.

series as well, as shown in Figure 10. Similar to the bottom panels of Figure 8, the characteristic widths of the systematic patterns are determined by the typical line widths. In addition, the patterns in Figure 10 are symmetric against the median date of the year due to the way we set up our BC values (see Section 2). Notably, in the optical, the amplitudes of the telluric-induced RV errors are small compared to the RV signals of an Earth-like planet (the top panel of Figure 10), even when tellurics are completely ignored and left untreated (largely thanks to the weighting process that eliminated the bad regions). By contrast, tellurics could pose significant challenges to the detection of an Earth-like planet in the NIR (the bottom panel), creating systematic noise much larger than 10 cm s\(^{-1}\) even with the typical best-effort treatment (CCF division).
These coherent systematic noises would add strong features to the periodograms for the NIR RVs at a periodicity of one year and its harmonics, while being almost negligible in the optical, as illustrated in Figure 11.

3.5. Effects of Wind Shifts in Oxygen Lines

In addition to testing the effects of the wrong line profile, we have also invested the effects of wind shifts in oxygen lines, as reported in Figueira et al. (2012).

We randomly added a wind shift in the oxygen lines with a wind speed of either $+20$ or $-20 \text{ m s}^{-1}$ in our simulated spectra, and we measured the RVs using the CCF division method. We then compared the RVs reported in orders 68, 69, 70, and 71 (about 620–640 nm in wavelengths), which have no telluric lines, strong oxygen and water lines ($\sim 50\% - 70\%$ in line depths), and moderate oxygen lines (typically $\sim 10\%$ in line depths), and shallow (micro) water lines (except for the broad ozone absorption), respectively. The RV scatter caused by tellurics in these four orders, if left untreated, are on the order of 1 cm s$^{-1}$ for order 68, 1 m s$^{-1}$ for order 71, and 10 m s$^{-1}$ for orders 69 and 70. With the CCF division method, the RV scatters are reduced to less than a centimeter per second for order 68, the centimeter-per-second level for order 71, and the meter-per-second level for orders 69 and 70. There was no change in the RVs extracted with the CCF division method in orders 68 and 71 before and after we added in the wind shifts in oxygen lines, as expected, since they do not have oxygen lines, and there were only roughly centimeter-per-second level or $<10\%$ changes at most in the RVs of orders 69 and 70. We thus conclude that wind shifts do not introduce significant additional RV errors on top of the original bias caused by the oxygen lines.

To validate our result, we increased the wind speed to $\pm 200 \text{ m s}^{-1}$, and we found that again the RVs reported by orders 67 and 70 do not change, as expected, but the RVs change by $\sim 0.1 - 1 \text{ m s}^{-1}$ for orders 68 and 69. The results are the same for orders 67 and 70 when we increased the wind speed to $\pm 2000 \text{ m s}^{-1}$ and the RV changes increased for orders 68 and 69. A wind speed of 200 or 2000 m s$^{-1}$ is unrealistic for the Earth's atmosphere, of course. We conclude that wind shifts do not have a significant impact on our results.

The results are not surprising given the small speed of wind in reality ($\sim 10 - 20 \text{ m s}^{-1}$). The typical resolution of PRV instruments and line widths are at the 2–5 km s$^{-1}$ level. Assuming a Gaussian line shape for simplicity, then the change in line flux induced by a small shift scale as the velocity shift divided by the line width (see, e.g., Equation (6) of Butler et al. 1996 and integrate over the line assuming a Gaussian shape). Therefore, systematic structures due to winds are typically $<1/200$ of the absorption line depth (10 m s$^{-1}$ divided by 2 km s$^{-1}$). For a telluric line with a depth of 30% such as the deep line in Figure 5, this is an effect at the 0.15% (1/200 x 30%) level, which is significantly smaller than the errors caused by the postconvolution division of telluric lines (at 0.2% in rms or $\sim 1\%$ in amplitude; see the residuals in Figure 5) or the typical residuals in modeling the telluric lines in real observations ($\sim 2\%$).

4. Discussion and Conclusion

4.1. Caveats and Additional Limitations Set by Tellurics

We enumerate the caveats of this study below, including several additional RV error terms that are induced by the telluric contamination but not considered in our simulations above. Because of these caveats and limitations, the numbers presented in this work should not be taken as the hard limit for the RV precision ceiling set by tellurics, but rather an estimate of where the ceiling could hit in real work. Improvements in instruments and algorithms could bring this ceiling down, but...
complications in real observations (such as the emission lines) could bring the ceiling up again.

(1) First, we assumed a very idealized level of precision, accuracy, and our knowledge on the extracted, observed spectra. We did not include photon noise in our simulated spectra in order to isolate the effects of telluric contamination. In reality, a limited SNR would translate into a limited ability of dividing out or modeling the telluric lines, and thus the existence of photon noise could further hinder our ability to eliminate the effects of tellurics. In addition, our simulation often assumes perfect knowledge of the spectral continuum or the blaze function, and we also assumed a constant SNR across each order, which is unrealistic for most RV instruments. The assumption of a constant SNR over an order probably has a rather minimal impact on the results, since this is equivalent to weighting the stellar lines somewhat differently when computing the final RVs. These weights do not differ significantly between different lines, given the relative small change in SNR within an order (typically a factor of two or so). They are also more or less random for each instrument (and when comparing between different instruments) because of the randomness in the wavelength ranges of the spectral orders.

A more likely cause in additional RV errors is the change in the blaze function. The exact shape of the blaze function could change due to a variety of factors, such as changes in the instruments and in the incident light (for example, due to the time-dependent Rayleigh scattering, even changes in the polarization; e.g., Halverson et al. 2015). Our work does not take such effects into account as this is beyond the scope of this study, but we do see some of the effects of a changing...
continuum in the blue region: in the calculations with the forward modeling using the wrong telluric profiles, we did not correct for the continuum absorption caused by the ozone in the blue optical region, which caused additional RV errors compared with the CCF division method that corrects for this continuum change (see the bottom panel of Figure 7).

(2) We also assumed perfect knowledge of the true underlying stellar spectrum, i.e., having a perfect stellar template or mask. In realistic situations, stellar spectra of very high SNR can be recovered from stacking many observations, but telluric lines cannot be completely eliminated from the stacked spectrum and spectral normalization can be challenging. Residual tellurics in the stellar template could induce additional errors as they would “latch” onto the telluric lines in the observed spectrum at each epoch (e.g., Wang et al. 2019). Therefore, constructing a clear stellar template is vital to minimizing the impact of tellurics in the RV but can be a challenging task, especially for the NIR. In particular, it is extra challenging to separate the telluric lines from stellar lines for targets near the ecliptic poles that would have a small BC variation (and, unfortunately, such targets are very often within the continuous viewing zones of space missions such as TESS and JWST). Future work on constructing accurate stellar templates would no doubt bring significant improvement in the RV precision, and it is also potentially critical for mitigating...
stellar jitter by anchoring an accurate “median” among the varying stellar spectral time series.

(3) Furthermore, we adopted a very simple scenario of telluric contamination and the level of inaccuracy in modeling the tellurics. First, we created a simple imitation to capture the mismatch between the real telluric lines and the model lines (see the description about the forward-modeling method in Section 2), while in reality the discrepancy could be larger. We have not tested the effects at various degrees or types of mismatches. Second, we did not include sky emission lines, such as the OH and Na lines, which could be challenging to remove due to their high variability and high intensity, especially for fiber-fed spectrographs without efficient sky fibers, and particularly challenging in the NIR. Better algorithms to mitigate our lack of knowledge of the varying telluric absorption and emission lines is one of the critical next steps in this line of work, and we discuss some promising future leads in the next subsection.

(4) There are also some caveats in the algorithms for extracting RVs in our work, especially when fitting the tellurics. Modeling the tellurics, the stellar spectrum, and the RV simultaneously is a nonlinear optimization problem, and we have not tested our algorithm extensively beyond comparing a few options and making a judgment call on choosing the “apparent best one” in terms of the level of goodness of fit and stability in convergence under different initial conditions. We chose to ignore an additional correction on the continuum for simplicity, but such an additional term could fix the errors due to our omission of ozone absorption in the blue and further improve the RV precision (though it complicates the algorithm in terms of better convergence). We also chose to model the tellurics separately in each spectral order, instead of performing a global fit of all orders, and we also did not fit a certain molecular species if it did not have enough absorption as quantified in its level of Doppler content. This is to follow the common practice in extracting RVs and also to easily isolate the effects of tellurics across different wavelengths, but it is unclear whether this is the best practice for obtaining a better fit to the tellurics and a higher RV precision. A more advanced algorithm for extracting RVs in the context of residual telluric lines is another key step in mitigating tellurics, or any form of spectral contamination.

(5) We have limited the scope of this paper to working with spectra with a resolution of $R = 120,000$, which is the common value for the current state-of-the-art RV spectrographs. However, a higher spectral resolution could bring additional benefits in terms of mitigating the tellurics (certainly for the intrinsic Doppler precision and mitigating stellar jitter; e.g., Davis et al. 2017). For example, the “convolution” errors caused by division shown in Figure 5 would decrease as the spectral resolution increases and the stellar and telluric lines are relatively more resolved from each other. A higher resolution might also help with constructing a more accurate telluric model and thus reducing the errors caused by an incomplete knowledge of the telluric lines, such as the ones we see with the “K-profile” methods in this work. However, the resolution adopted in this work ($R = 120,000$) is already high enough to resolve most of the telluric lines very well. Nonetheless, for some of the intrinsically heavily blended lines, such as the water lines near 1.5 μm, a higher resolution up to $R \sim 200,000$ does seem to bring a visible though limited improvement in resolving them.

To summarize, the first three caveats above (no noise, perfect template, simple inaccuracy in telluric modeling) would result in larger RV errors caused by tellurics, bringing the RV precision ceiling further up. The last two caveats (algorithm, spectral resolution) have the potential to reduce the RV errors by better modeling the telluric lines, especially in the NIR, but probably not significantly so in the broad optical range in real-life scenarios, given how close the RV errors are to the photon-limited precision (Table 1). However, it is hard to quantify the impacts of these caveats without further investigation (e.g., more simulations under different scenarios, such as with photon noise included), which is beyond the scope of this paper.

Finally, all calculations and discussions in this work are limited to solar analogs. RV measurements of stars of different stellar types will differ, as the Doppler information content of each star distributes differently across the wavelengths, and they also interact differently with the telluric lines as a result. Probing how the effects of tellurics change with different stellar types is also very important, especially for the NIR spectrographs that primarily target M dwarfs. This is the focus of a follow-up paper of this work (e.g., Latouf et al. 2022).

4.2. Conclusion and Future Work

In this work, we characterized the effects of telluric absorption in the broad optical and NIR bands using simulations, and we compared the efficacy of two commonly adopted methods for mitigating tellurics under different assumptions, including simulating the lack of knowledge of the telluric line profile. We found that tellurics contribute to the RV error budget on the level of a few centimeters per second in the optical regime, but this could increase to the meter-per-second level in the NIR, depending on the exact wavelength range. In addition to adding scatter to the RVs, tellurics induce coherent systematic signals, which could be more problematic for the search of periodic planetary signals.

We found that, under realistic assumptions where the telluric lines are not perfectly known, both methods, CCF+division and forward modeling, perform reasonably well at mitigating tellurics, especially in the NIR. In the optical bands, it is important and effective to simply lower the weights of the spectral regions that are affected by tellurics, or eliminate them completely. One could easily argue that the weighting scheme adopted in this work is not the optimized one, for example, as discussed in Section 3.3 when comparing the combined RVs from the CCF and forward-modeling methods using the wrong profiles. With an appropriate weighting or elimination scheme, it would not be necessary to invoke sophisticated mitigating methods such as the ones presented in this paper to reach a $10 \text{ cm s}^{-1}$ precision in the broad optical region.

In addition, it is more important to have an accurate model for the telluric lines in the observed spectrum. Under the assumptions in this work where we adopted a set of wrong line profiles for the tellurics (inducing a residual error on the level of a few percent), the RV errors induced by the tellurics show a significant increase, especially in the NIR. This suggests that modeling the telluric lines more accurately, for example, to better than a couple of percent, should bring visible improvement in the RV precision.

The purpose of this work is to assess the impacts of tellurics and some of the commonly adopted mitigation strategies. There is certainly work to be done to explore better strategies for
mitigating tellurics. For example, as mentioned above, perhaps a higher spectral resolution would bring an improvement in disentangling the tellurics, and there might be an optimal spectral resolution beyond which the gain in telluric mitigation would no longer be significant.

From the software side, there is still plenty of room for improvement: algorithms for constructing the most accurate template from observations in the existence of tellurics, algorithms for modeling the telluric lines more accurately, and also better algorithms to solve for the RVs in the telluric-contaminated spectra. For example, the RV extraction code wobble by Bedell et al. (2019) shows an innovative path toward solving the telluric problem. The algorithm adopted by wobble makes minimal assumptions regarding the prior knowledge of the stellar template or the telluric lines. Its pure machine-learning, data-driven approach offers an alternative path toward disentangling the stellar and telluric lines. This approach is similar to the PCA-based telluric modeling presented in Artigau et al. (2014) and used in the CFHT-SPIRou pipeline (e.g., Donati et al. 2020; Moutou et al. 2020), but with a more data-driven way to construct a telluric-free stellar template. Although wobble suffers from “convolution” errors as it assumes simple multiplication of spectra, it should, in principle, perform as well as the division method presented in this work, which can reach or is close to the photon-limited precision in the majority of the common photometric bands. wobble has been very successful in the optical band, and it is yet to be seen whether it could tackle the more challenging NIR region (e.g., Gan et al. 2022).

Another innovative data-driven method was presented by Rajpaul et al. (2020), where all stellar spectra were modeled with Gaussian processes and RVs were measured through pairwise comparison, and telluric contamination was part of the “variance” in the spectra. This has the advantage of not requiring any prior knowledge of tellurics and incorporating the impact of tellurics in the final error estimation, but similar to wobble this algorithm is yet to be tested for the NIR. A plausible path might be to use the synthetic telluric models as inputs to help the data-driven algorithms to arrive at a more accurate solution for the tellurics (e.g., perhaps along the lines of the method developed by Cretignier et al. 2021), while taking into account the uncertainties in telluric modeling, which would in turn generate a more accurate stellar model and more accurate RVs as a result.

In addition, improvements can be made in weighting the spectral regions contaminated by tellurics. In this work, weights were only adopted on a per-order basis, but an order covers a relatively wide spectral range (of order nanometers). Suborder weights or even pixel-based weights or rejections have the potential to further improve the RV precision while only suffering a relatively small SNR loss, and there might be a balance between rejecting telluric pixels versus retaining SNR. A recent work by Artigau et al. (2022) is an example adopting this philosophy, where they performed line-by-line RV analysis and assigned a different weight for each spectral line. Finally, it is noteworthy that in the NIR, tellurics could produce more errors than the stellar jitter. While stellar jitter remains the bottleneck for getting down to 10 cm s$^{-1}$ in the optical (e.g., Fischer et al. 2016), it is thought to be less of a problem in the NIR due to the decreased magnitude of spot contrast (e.g., Cale et al. 2021; Tran et al. 2021). However, the systematics induced by the tellurics in the NIR can get easily up to 1 m s$^{-1}$ and above (e.g., Figure 10), on par with the amplitudes of stellar jitter of mature and relatively quiet stars. Mitigating the effects of tellurics is thus vitally important for achieving sub-meter-per-second RV precision in the NIR. If NIR proves to be critical in mitigating the stellar jitter to the centimeter-per-second level, then achieving a centimeter-per-second precision would become critical for the NIR. This means great efforts should be made to deal with the telluric contamination in the NIR, or that the launch of a space-based NIR spectrograph, such as EarthFinder, becomes very appealing.

We thank the anonymous referee for their comments, which improved the quality of the paper.

N.L. gratefully acknowledges support from an NSF GRFP. P.P.P. acknowledges support from NASA (EarthFinder Probe Mission Concept Study 16-APROBES16-0020, Exoplanet Research Program Award #80NSSC20K0251, JPL Exoplanet Exploration Program and Research and Technology Development) and the NSF (Astronomy and Astrophysics Grants #1716202 and 2006517).

This work is being supported by CNES (Centre National des Études Spatiales) and CNRS (Centre National de la Recherche Scientifique). TAPAS is a service maintained by ETHER data center. ETHER acknowledges for TAPAS the use of the HITRAN database and theLBLRTM radiative transfer code, the use of ECMWF data, and the ETHER data center. We acknowledge useful discussions with Iouli Gordon and Larry Rothman.

ORCID iDs

Sharon Xuesong Wang (王雪淞) @ https://orcid.org/0000-0002-6937-9034

Natasha Latouf @ https://orcid.org/0000-0001-8079-1882

Peter Plavchan @ https://orcid.org/0000-0002-8864-1667

Bryson Cale @ https://orcid.org/0000-0001-6279-0595

Cullen Blake @ https://orcid.org/0000-0002-6096-1749

Étienne Artigau @ https://orcid.org/0000-0003-3506-5667

Carey M. Lisse @ https://orcid.org/0000-0002-9548-1526

Jonathan Gagné @ https://orcid.org/0000-0002-2592-9612

Jonathan Crass @ https://orcid.org/0000-0002-1503-2852

References

Artigau, É., Astudillo-Defru, N., Delfosse, X., et al. 2014, Proc. SPIE, 9149, 1716202 and 2006517

Bertaux, J. L., Lallement, R., Ferron, S., Boonne, C., & Bodichon, R. 2014, A&A, 564, A46

Blackett, R. T., Fischer, D. A., Jurgenson, C. A., et al. 2020, AJ, 159, 238

Butler, R. P., Marcy, G. W., Williams, E., et al. 1996, PASP, 108, 500

Butler, R. P., Vogt, S. S., Laughlin, G., et al. 2017, AJ, 153, 208

Cale, B., Reefe, M., Plavchan, P., et al. 2021, AJ, 162, 295

Crandall, G., Gaia, B. S., Leifer, S., et al. 2021, arXiv:2107.14291

Crepp, J. R., Crass, J., King, D., et al. 2016, Proc. SPIE, 9908, 990819

Kaminski, A., Hartman, H., et al. 2010, ApJ, 713, 410

Bedell, M., Hogg, D. W., Foreman-Mackey, D., Montet, B. T., & Lugr, R. 2019, AJ, 158, 164

Berta, T. L., Lallement, R., Ferron, S., Boonne, C., & Bodichon, R., 2014, A&A, 564, A46

Blackman, R. T., Fischer, D. A., Jurgenson, C. A., et al. 2020, AJ, 159, 238

Butler, R. P., Marcy, G. W., Williams, E., et al. 1996, PASP, 108, 500

Butler, R. P., Vogt, S. S., Laughlin, G., et al. 2017, AJ, 153, 208

Cale, B., Reefe, M., Plavchan, P., et al. 2021, AJ, 162, 295

Crass, J., Gaudi, B. S., Leifer, S., et al. 2021, arXiv:2107.14291

Crepp, J. R., Crass, J., King, D., et al. 2016, Proc. SPIE, 9908, 990819

Cretignier, M., Dumasque, X., Haro, N. C., & Pepe, F. 2021, A&A, 653, A43

Cunha, D., Santos, N. C., Figueira, P., et al. 2014, A&A, 568, A35
Davis, A. B., Cisewski, J., Dumusque, X., Fischer, D. A., & Ford, E. B. 2017, ApJ, 846, 59
Donati, J. F., Kouach, D., Moutou, C., et al. 2020, MNRAS, 498, 5684
Dressing, C., Stark, C. C., Plavchan, P., & Lopez, E. 2019, BAAS, 51, 268
Figueira, P., Kerber, F., Chacon, A., et al. 2012, MNRAS, 420, 2874
Fischer, D. A., Anglada-Escude, G., Arriagada, P., et al. 2016, PASP, 128, 066001
Gan, T., Lin, Z., Xu, X., Wang, S., et al. 2022, MNRAS, 511, 83
Gordon, I. E., Rothman, L. S., Hill, C., et al. 2017, JQSRT, 203, 3
Gullikson, K., Dodson-Robinson, S., & Kraus, A. 2014, AJ, 148, 53
Halverson, S., Roy, A., Mahadevan, S., & Schwab, C. 2015, ApJL, 814, L22
Halverson, S., Terrien, R., Mahadevan, S., et al. 2016, Proc. SPIE, 9908, 99086P
Hirano, T., Kuzuhara, M., Kotani, T., et al. 2020, PASJ, 72, 93
Howard, A. W., & Fulton, B. J. 2016, PASP, 128, 114401
Kurucz, R. L. 2005, MSAIS, 8, 14
Latouf, N., Xuesong-Wang, S., Cale, B., & Plavchan, P. 2022, arXiv:2206.05233
Li, D., Blake, C. H., Nidever, D., & Halverson, S. P. 2018, PASP, 130, 014501
Lin, A. S. J., Monson, A., Mahadevan, S., et al. 2022, AJ, 163, 184
Lisogorskyi, M., Jones, H. R. A., & Feng, F. 2019, MNRAS, 485, 4804
Meier Valdés, E. A., Morris, B. M., & Demory, B. O. 2021, A&A, 649, A132
Morgan, R., Savransky, D., Turmon, M., et al. 2021, JATIS, 7, 021220
Moutou, C., Dalal, S., Donati, J. F., et al. 2020, A&A, 642, A72
Pepe, F., Cristiani, S., Rebolo, R., et al. 2021, A&A, 645, A96
Plavchan, P., Latham, D., Gaudi, S., et al. 2015, arXiv:1503.01770
Plavchan, P., Vasishht, G., Beichman, C., et al. 2020, arXiv:2006.13428
Rajpaul, V. M., Aigrain, S., & Buchhave, L. A. 2020, MNRAS, 492, 3960
Rothman, L. S., Gordon, I. E., Babikov, Y., et al. 2013, JQSRT, 130, 4
Roy, A., Halverson, S., Mahadevan, S., et al. 2020, AJ, 159, 161
Sameshima, H., Matsunaga, N., Kobayashi, N., et al. 2018, PASP, 130, 074502
Seifahrt, A., Käufl, H. U., Zangl, G., et al. 2010, A&A, 524, A11
Sithajan, S., Ge, J., & Wang, J. 2016, AAS Meeting Abstracts, 227, 137,19
Smette, A., Sana, H., Noll, S., et al. 2015, A&A, 576, A77
Tran, Q. H., Bowler, B. P., Cochran, W. D., et al. 2021, AJ, 161, 173
Ulmer-Moll, S., Figueira, P., Neal, J. J., Santos, N. C., & Bonnefoy, M. 2019, A&A, 621, A79
Vacca, W. D., Cushing, M. C., & Rayner, J. T. 2003, PASP, 115, 389
Wang, S. X., Wright, J. T., Bender, C., et al. 2019, AJ, 158, 216
Wright, J. T., & Robertson, P. 2017, RNAAS, 1, 51
Wright, J. T., Roy, A., Mahadevan, S., et al. 2013, ApJ, 770, 119
Zechmeister, M., Reiners, A., Amado, P. J., et al. 2018, A&A, 609, A12
Zhao, L. L., Fischer, D. A., Ford, E. B., et al. 2022, AJ, 163, 171