The Astrophysical Journal Supplement Series, 271:24 (11pp), 2024 March https://doi.org/10.3847/1538-4365/ad1e55

IRIS^{2+}: A Comprehensive Database of Stratified Thermodynamic Models in the Low Solar Atmosphere

Alberto Sainz Dalda[1,2], Aaryan Agrawal[3], Bart De Pontieu[1,4,5], and Milan Gošić[1,2]
[1] Lockheed Martin Solar & Astrophysics Laboratory, 3251 Hanover Street, Palo Alto, CA 94304, USA
[2] Bay Area Environmental Research Institute, NASA Research Park, Moffett Field, CA 94035, USA
[3] Henry M. Gunn High School, 780 Arastradero Road, Palo Alto, CA 94306, USA
[4] Rosseland Center for Solar Physics, University of Oslo, P.O. Box 1029 Blindern, NO-0315 Oslo, Norway
[5] Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, NO-0315 Oslo, Norway
Received 2023 March 12; revised 2024 January 8; accepted 2024 January 12; published 2024 February 27

Abstract

We present the most comprehensive collection of stratified-in-optical-depth thermodynamic models of the lower solar atmosphere. This database, named IRIS^{2+}, consists of 40,320 synthetic representative profiles (RPs) and their corresponding representative model atmospheres (RMAs). The latter result from inverting the observed RPs, which are the result of clustering multiline spectral profiles observed by the Interface Region Imaging Spectrograph in 126 active regions (ARs). Each AR was clustered in 320 RPs considering five solar features: umbra, pore or pore-like, penumbra, plage, and quiet Sun. The multiline spectral profiles contain 12 lines with encoded physical information from the top of the chromosphere to the midphotosphere. These 12 spectral lines, six chromospheric and six photospheric, were simultaneously inverted with the Stockholm Inversion Code. Therefore, the stratified-in-optical-depth low solar atmosphere models obtained in this study are better constrained than those obtained from the nonsimultaneous inversion of individual lines. Each tuple in the database contains an inverted RP, its corresponding RMA, the solar feature label where the observed RP was clustered, its heliocentric angle, and its date and time of observation. The database may be used as an inversion code for any of the lines present in the RPs, and the RMAs may be used for synthesizing spectral lines in the chromosphere and photosphere or as mean constraints for numerical models of the solar atmosphere.

Unified Astronomy Thesaurus concepts: The Sun (1693); Solar photosphere (1518); Solar chromosphere (1479); Radiative transfer (1335)

1. Introduction

The study of the Sun has played a prominent role in the knowledge we have about stars and in the development of the techniques and ideas needed to gain that knowledge. Thanks to its proximity, we are able to resolve many structures on the photosphere, chromosphere, transition region, and corona. Current instrumentation and post-processing data analysis techniques allow us to recover in great detail the physical conditions of solar features in scales as small as 100 km, with a temporal cadence of a few seconds, and with a high accuracy in the inferred magnetic field and the thermodynamic parameters.

There are two main ways to gain knowledge about the solar atmosphere. Forward modeling synthesizes the spectra from thermodynamic or magnetothermodynamic models by solving the radiative transfer (RT) problem. Then, those spectra are compared with the observed one in the solar feature that the models are simulating. It should be noted that the spectra obtained are unique and rely only on the assumptions made to create the numerical model and those related to the RT. The second method, called inversion of spectro(polarimetric) data recovers a model atmosphere from iteratively fitting an observed profile to a synthetic profile, which is obtained by solving the RT from an initial guess model atmosphere and slight subsequent modifications on it. These small changes in the parameters of the model atmosphere try to minimize the difference between the observed profile and the synthetic profile, which is evaluated through the loss function (see Equation (1)). This process continues until the loss function falls below a given threshold or a maximum number of iterations is reached. This method does not provide a unique solution, because it sensitively depends on the initial guess model and could suffer from common issues with least-squares fitting, such as the solution converging to a local minimum. Because of that, we (should) say, in the case of a good fit, that the recovered model atmosphere is compatible with the observed profile, and therefore that these physical conditions may occur in the solar atmosphere. A recent comparison between inferred values using the inversion method in a numerical model of the solar atmosphere and the actual values in that model has yielded average differences in the photosphere as small as $\delta|B| < 50$ G, $\delta\gamma < 20°$, $\delta T < 30$ K, and $\delta v_{\rm los} < 0.3$ km s^{-1} for the magnetic field strength and inclination, temperature, and line-of-sight (LOS) velocity, respectively (Quintero Noda et al. 2023). A detailed review of inversion methods can be found in del Toro Iniesta & Ruiz Cobo (2016).

These two approaches offer valuable insight into the physical conditions and processes that govern the solar atmosphere. It is essential to have a realistic understanding of the physical conditions of the solar atmosphere, that is, of the assumptions that are already known to be necessary to reproduce the observables as accurately as possible. In forward modeling, these assumptions are derived from previous observations and theoretical considerations of the problem, such as hydrostatic equilibrium, ambipolar diffusion (Khomenko et al. 2014;

Martínez-Sykora et al. 2020), or multifluid interactions (Wargnier et al. 2023). For inversions, the success of the results depends, among other factors, on the initial guess model used to start the inversion and—as with forward modeling—on the theoretical considerations made while solving the RT problem, such as the atomic model used to synthesize the spectral lines, statistical equilibrium, nonlocal thermodynamic equilibrium, or partial frequency redistribution. It should be noted that both forward modeling and inversion methods require solving the RT problem, and both may be limited by computational resources when certain assumptions are made, such as considering 3D forward modeling from numerical simulations or partial frequency redistribution for inversions. Thus, for instance, including the effect of the ambipolar diffusion and nonequilibrium ionization (Martínez-Sykora et al. 2023) or considering 3D RT (Sukhorukov & Leenaarts 2017; Bjørgen et al. 2019) have a direct impact on the shape of the Mg II k line. Until these realistic considerations are implemented to efficiently solve the RT in forward modeling and inversions, we shall work with the available state-of-the-art codes, accepting and acknowledging their limitations.

In this paper, we provide a collection of thermodynamic, nonmagnetic model atmospheres for active regions (ARs). We consider an active region as the nonexclusive combination of umbra, pore or pore-like structure, penumbra, plage, and the neighboring quiet Sun. There are many solar atmosphere models available:[6] some of them, mostly in the photosphere, are devoted to a particular solar feature, e.g., the HSRA quiet Sun model (Gingerich et al. 1971), the hot sunspot model (Collados et al. 1994), and the penumbra model (del Toro Iniesta et al. 1994); others consider a general behavior of some features in active regions (quiet Sun cell center, quiet Sun network, active network, penumbra, umbra, faculae, and plage) both in the photosphere and the chromosphere (Vernazza et al. 1981; Fontenla et al. 1999, 2006). Some of these models are considered as the standard model for their corresponding features. In the case of inversions, these models are mostly used as initial guess models. The IRIS^{2+} database provides a set of 40,320 representative model atmosphere (RMA) and their corresponding RPs. This database provides the most diverse set of profiles and models with a simultaneously constrained stratification in optical from the top of the chromosphere to the midphotosphere for the most relevant features of the ARs. Thus, each RP–RMA pair is labeled with the solar feature where it was observed (e.g., "plage"; see Section 2.2), its location on the solar disk, and the recording date.

The method followed to build the IRIS^{2+} database is conceptually the same as for IRIS^2 (Sainz Dalda et al. 2019). However, the RPs are now more evenly distributed among solar features and consider six lines in the chromosphere and six in the photosphere, while the models are more accurate in terms of optical depth sampling. To create IRIS^{2+}, we have enforced a more uniform distribution of clusters per solar feature detected in a data set, while in IRIS^2 the number of clusters was given for the entire data set. As a result, in IRIS^2, the number of RPs associated with the quiet Sun after clustering the data set was much larger than the number associated with the umbra in some cases. Moreover, the selection of spectral lines sensitive to the thermodynamics at different layers of the chromosphere and photosphere allows us to obtain more accurate RMAs. We note that here "accurate" does not refer to the uncertainty associated with the RMA obtained after the fit between an observed profile and the closest RP in the database. By "accurate in terms of optical depth sampling," we refer to the quality of thermodynamic sampling along the optical depth, i.e., the thermodynamics values are simultaneously inferred at a larger, better sampled optical depth range. Therefore, the inferred models are better constrained in optical depth in IRIS^{2+} than in IRIS^2, although the latter considers a larger variety of solar features.

In Section 2, we detail the main characteristics of the IRIS data that we have used in our study, the spectral lines that we have selected for the inversions, the formation height of these lines, and the response of them to changes in the solar atmosphere. We explain how we have selected the different areas for the features in the AR data, i.e., umbra, pore-like, penumbra, plage, and surrounding quiet Sun. We justify the need to cluster these data and how we treated the spectral lines to be simultaneously inverted. In Section 3, we show some of the inferred/inverted model atmospheres and their associated synthetic spectra. As an example of the utilization of the IRIS^{2+} database, we show the available RMAs in an interval of a heleiocentric angle for each solar feature in the ARs considered in this study. In Section 4, we discuss the reliability and usability of the IRIS^{2+} database.

2. Data

One of the most critical parts of this project is to select a set of data that properly represent the variety of active regions in the Sun. As we have already mentioned, by "active region" we mean the nonexclusive combination of umbra, pore, penumbra, plage, and quiet Sun (which may or may not be affected by the neighboring structures). This relaxed definition does not diminish the goal of this study: to provide a comprehensive collection of RPs and RMAs that characterizes the thermodynamics of the main elements of the ARs. To this aim, we have selected 126 IRIS data sets located at different positions on the solar disk, observed with different exposure times, and containing different active regions observed in the time range from 2013 July (the beginning of the IRIS mission) to 2021. Figure 1 shows the distribution of the selected ARs with respect to the observing date (left panel), the location on the solar disk given by $\mu = \cos(\theta)$, with θ the heliocentric angle (center panel), and the exposure time (right panel). The size of the field of view was 128×130 arcsec^2 for 104 data sets, 128×175 arcsec^2 for 14 data sets, 140×175 arcsec^2 for three data sets, 112×124 arcsec^2 for three data sets, 64×175 arcsec^2 for one data set, and 64×124 arcsec^2 for one data set. Thus, this selection of data sets is a good representation of the data observed by IRIS on the solar disk, with different signal-to-noise ratios, stages of the solar cycle, and pixel sizes (and thus numbers of profiles) per data set.

2.1. Selection of Spectral Lines

Another important issue is the selection of the photospheric spectral lines. We have tried different combinations of photospheric lines located in the near-ultraviolet (NUV) spectral range observed by IRIS. The bottom panel of Figure 2 shows the full spectrum taken by IRIS in its NUV channel. In the top panel of this figure, an average profile of the

[6] An excellent set of solar atmosphere models has been collected by Dr. Basilio Ruiz Cobo. They are available at https://github.com/BasilioRuiz/SIR-code/tree/master/models.

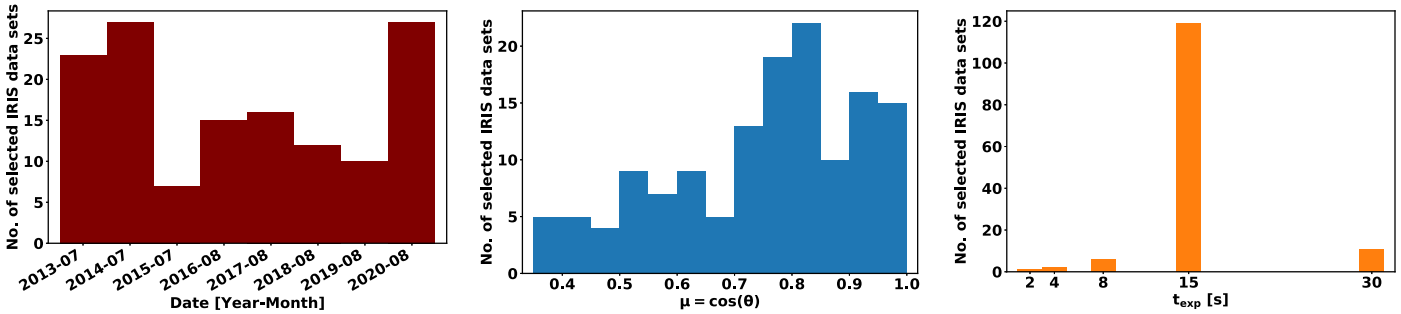

Figure 1. Distribution of the selected data in the IRIS^{2+} database with respect the observation date (left), $\mu = \cos(\theta)$—with θ being the heliocentric angle (middle), and the exposure time (right).

quiet Sun observed at the center of the solar disk is shown. The selected photospheric lines are indicated by orange vertical lines in the bottom panel and orange labels in the top panel, while the chromospheric lines, i.e., the Mg II h and k lines and the Mg II UV triplet lines, are similarly indicated in violet. The spectrum in the bottom panel of this figure is intended to aid in comparisons with the IRIS *linelist* images,[7] which illustrate the combination of spectral ranges selected in an IRIS observation. Table 1 shows which of the selected lines are available in the IRIS *linelists*. In addition to the lines observed in the NUV channel, we also considered the C II 1334 and 1335 Å chromospheric lines recorded in the far-ultraviolet (FUV) channel.

The selection of photospheric lines is intended to recover as accurately as possible the stratification of the thermodynamic variables from the bottom to the top of the photosphere. This region, in optical depth, is derived from the *response function* (RF; Mein 1971; Landi Degl'Innocenti & Landi Degl'Innocenti 1981) of the intensity profiles to a perturbation in a given physical parameter. For a given RP–RMA pair, their corresponding RFs are calculated by STiC. The left panel of Figure 3 shows the RF of the intensity (for a multiline RP in the database) to a perturbation in the temperature. We note that the optical depth range where a spectral line is sensitive to a change in a physical parameter is slightly different for different solar features (e.g., umbra, penumbra, filament, and plage), and it is also different for the various physical parameters (temperature, electron density, LOS velocity, and microturbulence). The right panel of Figure 3 shows the average optical depth ranges where the selected lines in IRIS^{2+} are sensitive to perturbations in the temperature for most of the solar features observed in the database.

2.2. Multiline Representative Profiles

We now prepare the selected spectral lines for clustering. The need for clustering our data is due to the computational resources needed to recover the physical information encoded in the spectral lines, specifically when we are working with the C II 1334 and 1335 Å and the Mg II h and k lines. The C II 1334 and 1335 Å lines were inverted taking into account non-LTE and complete frequency redistribution of the scattered photons, while non-LTE and partial frequency redistribution were taken into consideration for the Mg II h and k lines. On average, the inversion of a Mg II h and k profile with STiC under these conditions requires 1.5 CPU hours. Therefore, we follow the same approach introduced by Sainz Dalda et al. (2019), which was extended by Woods et al. (2021) for the multiline case of C II 1334 and 1335 Å lines and Mg II h and k (including Mg II UV2 and 3 lines), and more recently by Sainz Dalda & De Pontieu (2023) for the multiline case of C II 1334 and 1335 Å lines, Mg II h and k, and all Mg II UV triplet lines. Now, in addition to the set of chromospheric lines used by the latter authors, we include six photospheric lines. Therefore, identical chromospheric profiles can now be associated with photospheric profiles that are quite different, or vice versa. We must consider this degeneracy between the chromospheric and photospheric sets of lines when clustering them. Thus, compared to inversions that are based on fewer lines, we need to consider a larger number of clusters and RPs to properly capture that degeneracy. Using the elbow method, we have determined that 320 clusters per data set are capable of adequately representing most of the data sets considered in our study. This number of clusters, i.e., RPs, is also an optimal number of RPs to be inverted in a midsize 320 core server, without the need to get access to larger supercomputer facilities. We note that inverting a multiline profile, such as the one mentioned above, requires 6 CPU hours. In this investigation, the average number of profiles per data set is 50,000. Inverting all profiles would require $50{,}000 \times 6$ CPU hours $\times$ 126 data sets, that is, ≈38 million CPU hours or ≈14 yr on a 320 core server. Therefore, we have followed the approach of IRIS^2: we invert just 320 RPs per data set, obtaining the corresponding 320 RMAs per data set in 6 hr on a 320 core server, that is, in ≈31 days for the 126 data sets.

In this investigation, because we are interested in obtaining thermodynamic models for the ARs, we have stratified the data before clustering. This means that we have first identified those locations where the profiles belong to an umbra (or a pore), a penumbra, a plage, or the quiet Sun. These four regions are determined using different intensity thresholds in the reconstructed intensity map at 2810.58 Å of the data set to be clustered. The first three panels in the top row of Figure 4 show the intensity spectroheliogram or map for an AR at Mg II k, Ti II 2785.46 Å and at 2832.04 Å (photospheric continuum), respectively. The fourth panel shows the areas identified as umbra, penumbra, plage, and quiet Sun. Once these areas are determined, we impose the number of clusters for each region to be clustered. The number of clusters for each region is based on the average area covered by these regions and on the variability of different physical conditions in these areas. For instance, the quiet Sun occupies more area in most of the data selected compared to the penumbra. On the other hand, the physical conditions are likely more variable in the latter, although this is not necessarily true in the chromosphere. We have attempted to find a trade-off between the hardware limitations (to consider at most 320 RPs per data set) and

[7] Once an IRIS observation has been selected at https://iris.lmsal.com/search/, the linelist is shown in the "raster" column. The linelist is used for telemetry reasons, i.e., to minimize the volume of data downloaded from IRIS to the ground base facility.

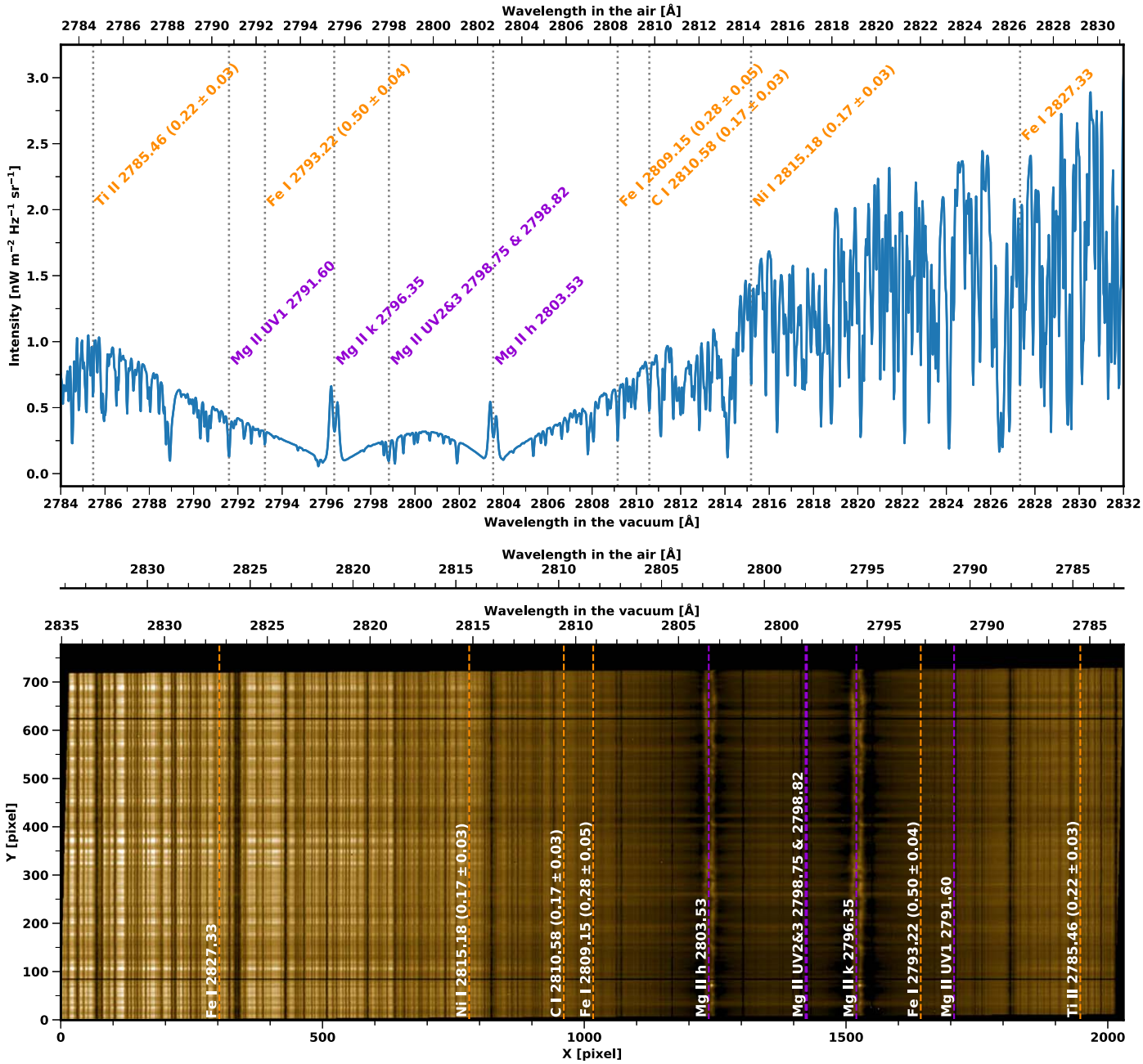

Figure 2. Top: Average profile of the quiet Sun at the disk center showing the spectrum around the Mg II h and k lines observed by IRIS in its near-ultraviolet camera. The core of the photospheric and chromospheric lines selected in this investigation are marked by dashed lines. The photospheric lines are labeled in orange, while the chromospheric are in violet. The values in parentheses indicate the approximate height (in megameters) above the optical depth at 500 nm equal to 1 sampled by the velocity line shifts, and the uncertainty associated with the fact that the formation occurs in a corrugated layer, as obtained by Pereira et al. (2013). Bottom: The spectrum as seen on the detector is shown for a better comparison with the IRIS linelist files, e.g., v36_01, which show all the spectral ranges selected for a given observation. It should be noted that the wavelength increases toward the left in this panel.

achieving a meaningful representation of the physics of the selected areas in each data set. The number of clusters for the umbra, penumbra, plage, and quiet Sun are 30, 50, 80, and 160, respectively. If there is no umbra in a given data set, the features detected are usually pores, orphan penumbras, or naked sunspots. In this case, the numbers of clusters for the penumbra are incremented up to 80, and the associated RPs are labeled as "pore-like." If there is neither umbra nor penumbra, the numbers of clusters for the plage and quiet Sun are 144 and 176 respectively. In total, the IRIS^{2+} database has 2280 RPs associated with umbra, 800 RPs with pore-like features, 3800 RPs with penumbra, 12,640 with plage, and 20,800 RPs with quiet Sun. To cluster the original IRIS profiles, we created a *joint profile* by cropping the selected lines and concatenating them together. The joint profiles are now scaled: at a given wavelength, the intensity value is scaled between 0 and 1 to the minimum and maximum of all the intensity values at that given wavelength. This is done for all the wavelengths in the joint profile. For example, if we do not adjust the C II 1334 and 1335 Å lines at the same order of magnitude as Mg II h and k before clustering, the information from C II 1334 and 1335 Å would be mostly ignored when the Euclidean distance is used to cluster the data, as the C II 1334 and 1335 Å lines are, in

Table 1
Availability of the Spectral Lines (Wavelengths in Angstroms) Considered in IRIS^{2+} for Each Observation IRIS Linelist

IRIS Linelist	Fe I 2827.33	Ni II 2815.18	C I 2810.58	Fe I 2809.15	Mg II h 2803.53	Mg II UV3 and 2 2798.85&.72	Mg II k 2796.35	Fe I 2793.22	Mg II UV1 2791.60	Ti II 2785.46
v36_00 and v38_00		✓		✓	✓	✓	✓	✓	✓	
v36_01 and v38_01		✓			✓	✓	✓			
v36_02 and v38_02		✓				✓	✓			
v36_03 and v38_03	✓	✓			✓	✓	✓	✓	✓	
v36_04 and v38_04	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
v40_00					✓	✓	✓			✓
v40_02					✓		✓			
v40_09	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note. The C II 1334 and 1335 Å lines, not included in the table, are always included and thus considered by IRIS^{2+}.

physical units, several orders of magnitude smaller than the Mg II h and k lines. As a result, we would be mainly clustering the Mg II h and k lines and mostly neglecting the C II 1334 and 1335 Å lines. After this, the cropped, scaled joint profiles are clustered in 320 clusters. Once the clusters are defined, that is, all profiles are associated with a cluster label, we calculate the actual RPs (centroids) as the mean of the original profiles belonging to a cluster. It should be noted that we are now using the original observed profiles, which are not scaled, that is, they preserve the physical units (specific intensity given in erg cm^{-2} s^{-1} sr^{-1} Å^{-1}). At this point, the RPs preserve the original spectral sampling, the original spectral range of the selected lines (not just the cropped lines of interest), and the physical units. They are ready for the inversion.

2.3. Inversion of the Multiline Representative Profiles

We have used the multiatom STockholm inversion Code (STiC; de la Cruz Rodríguez et al. 2016, 2019) to invert the RPs of our selected data. STiC is a Message Passing Interface–parallel non-LTE inversion code that utilizes a modified version of RH (Uitenbroek 2001) to solve the atomic population densities assuming statistical equilibrium and plane-parallel geometry, and it allows including partial frequency redistribution effects of scattered photons (Leenaarts et al. 2012). The RT equation is solved using cubic Bezier solvers (de la Cruz Rodríguez & Piskunov 2013). The inversion engine of STiC includes an equation of state extracted from the SME code (Piskunov & Valenti 2017).[8]

Accounting for the significant intensity difference between spectral lines is critical for a successful multiline inversion. This is particularly important for the C II 1334 and 1335 Å lines, because their intensity is much lower than the rest of the lines of this study. Therefore, during the inversion, the lines are weighted to account for this difference when the loss function (χ^2; see Equation (1)) is evaluated. We have also considered this weighting for the Mg II UV triplet and photospheric lines with respect to the Mg II h and k lines. To determine the weights, we calculate the most frequent value of the intensity in a spectroheliogram map in the Mg II h and k k_3, and we weight the C II 1334 and 1335 Å lines to that value. That means we *adjust* the C II 1334 and 1335 Å lines to the same relative intensity level as the Mg II h and k lines. For the rest of the lines, we found that the relative intensity with respect to that of the Mg II h and k lines for different data sets has a small standard deviation; therefore, we applied the same combination of weights for all data sets. The final relative weights are $w_{\mathrm{MgII\ UV\ triplet}}$: $w_{\mathrm{MgII\ h\ and\ k}}$: $w_{\mathrm{photos}} = 3{:}1{:}2$.

For each data set, STiC only accepts a single wavelength-dependent weight and noise value for all profiles. For simplicity, a noise value (standard deviation in a spectral range, $\sigma_{I(\lambda)}$) is considered for the FUV data (C II 1334 and 1335 Å lines, with $\sigma_{I(\lambda)}$ calculated between $1333.0 \leqslant \lambda[\text{Å}] \leqslant 1333.5$), while the NUV data (Mg II h and k, Mg II UV triplet lines, and the photospheric lines, with $\sigma_{I(\lambda)}$ calculated between $2800.20 \leqslant \lambda[\text{Å}] \leqslant 2800.55$) also have their noise. We note that any other investigation considering different lines from the FUV channel should calculate the noise for those lines independently.

The merit function used to quantify the quality of the fit between the observed $I(\lambda_i)^{\mathrm{obs}}$ and the inverted $I(\lambda_i, \boldsymbol{M})^{\mathrm{syn}}$ profile is

$$\chi^2 = \frac{1}{\nu}\sum_{i=1}^{q}(I(\lambda_i)^{\mathrm{obs}} - I(\lambda_i, \boldsymbol{M})^{\mathrm{syn}})^2\frac{w_i^2}{\sigma_i^2}, \quad (1)$$

with $i = 1,\ldots,q$ the sampled wavelengths, w_i their weights, σ_i uncertainties of the observation, and ν the number of degrees of freedom, that is, the difference between the observables (q) and the free parameters in the model M or nodes during the inversion. We use a unique value of w_i/σ_i per data set. As it has been mentioned above, different weights are given for different spectral ranges, and we assume that σ_i is the same for all i at each spectral channel. That is, we have used $w_{\mathrm{CII}}/\sigma_{\mathrm{CII}}$, $w_{\mathrm{MgII}}/\sigma_{\mathrm{NUV}}$, $w_{\mathrm{MgII}}/\sigma_{\mathrm{NUV}}$, and $w_{\mathrm{photos}}/\sigma_{\mathrm{NUV}}$. We have used four cycles with different numbers of nodes for the physical variables in the model. These values are detailed in Table 2. In addition, another free parameter is considered to *adjust* the boundary gas pressure. Thus, the inversion code can increase the gas pressure in the top boundary to higher values, which allows the observed line core of the chromospheric lines to be reproduced properly without affecting the photospheric lines (see Section 3.2 in de la Cruz Rodríguez et al. 2019 for more details). This particular combination of nodes has proven to be quite satisfactory in most inverted profiles. Naturally, a larger number of nodes would provide a better fit for the profiles, but at the expense of overfitting and generating nonrealistic model atmospheres. We note that the stratified-in-optical-depth electron density (n_e) and gas pressure (p_g) are derived assuming hydrostatic equilibrium, and therefore, they do not contribute to the total number of nodes used during the inversion. We used the FALC quiet-Sun model (Fontenla et al. 1993) as the initial guess model for all the inversions. Ideally, we should invert

[8] STiC is available at https://github.com/jaimedelacruz/stic/.

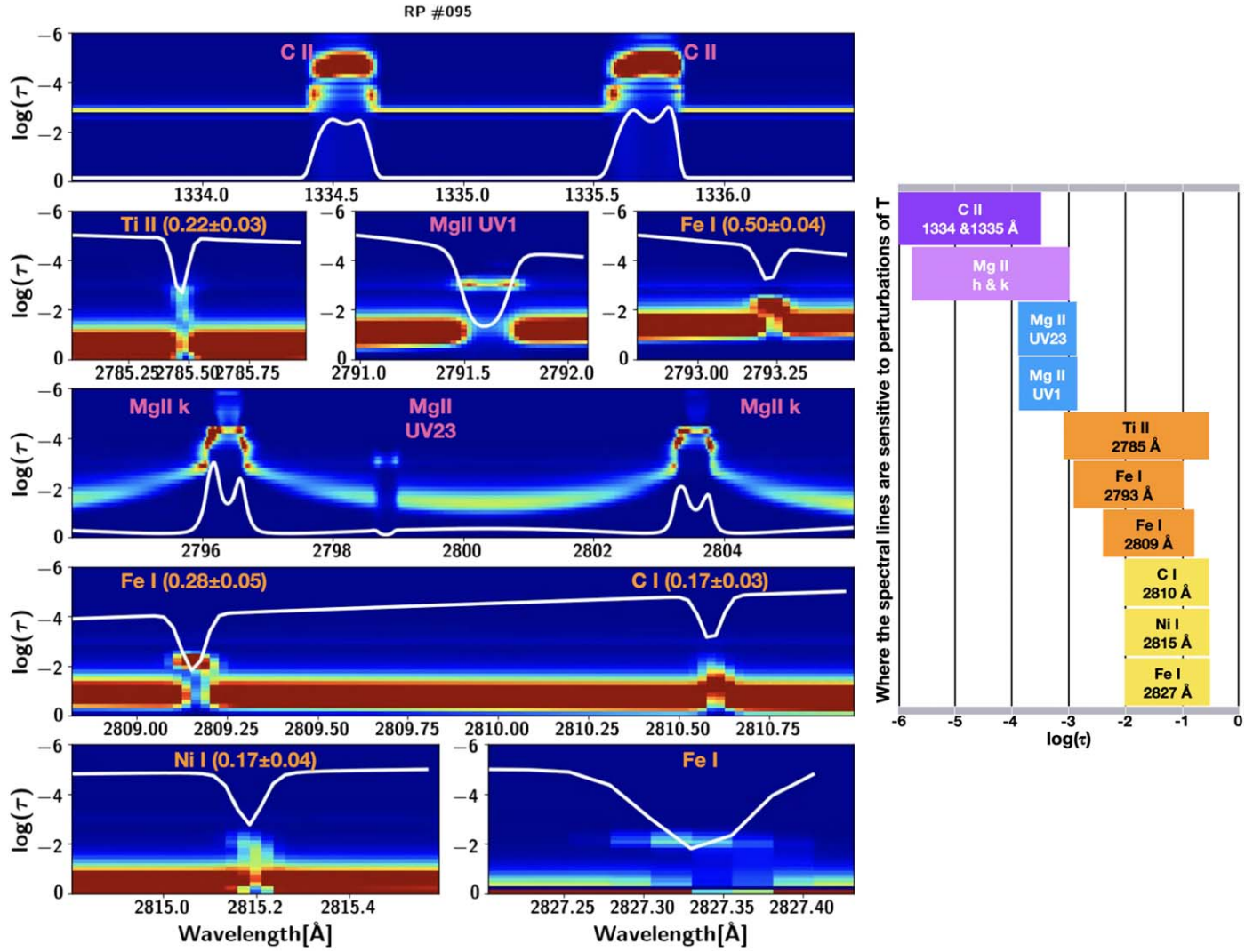

Figure 3. Left: Response function (RF) of the intensity to perturbations in T of the selected lines in an RP of the IRIS^{2+} database. The intensity profiles (white) are shown as a reference. The meanings of the values in parentheses are given in the caption of Figure 2. Right: Optical depth ranges where these lines are on average sensitive to a perturbation in T for the solar features included in the IRIS^{2+} database. These ranges may slightly vary for each solar feature and for the other physical parameters in the database.

each profile considering several initial guess models and choose the solution that best fits the observation. This approach has recently become possible thanks to the optimization of inversion codes; for example, the DeSIRe code (Ruiz Cobo et al. 2022) has this capability. Because the Mg II h and k lines need to account for the PRD effects, inverting them requires a significant amount of computing time; therefore, this inversion strategy is not feasible with the currently available inversion codes and computational resources. For this reason, we only consider one initial guess model.

The C II 1334 and 1335 Å lines and the Mg II UV triplet lines were inverted considering non-LTE conditions and complete frequency redistribution, while for the Mg II h and k lines, non-LTE and partial frequency redistribution were considered. All photospheric lines were inverted considering LTE conditions. We note that the lines of Fe I and Fe II are affected by the overionization due to the near-UV radiation (Athay & Lites 1972). As a consequence, the Fe I are underpopulated and a few Fe II levels are overexcited. Shchukina & Trujillo Bueno (2001) found that, for the low-excitation Fe I, the errors in the kinetic temperature in the granules are significant if these lines are not considered under non-LTE conditions. More recently, Smitha et al. (2020) found that inverting the Fe I 6301 Å and 6302 Å lines under 1D LTE conditions, i.e., neglecting the non-LTE, can introduce errors up to 13% in temperature and as high as 50% both in the $v_{\rm los}$ and B. Smitha et al. (2021) also found that the horizontal RT, i.e., considering 3D non-LTE, also has an impact in the temperature errors, although this impact is more localized in specific regions of the atmosphere. As Shchukina & Trujillo Bueno (2001) pointed out, the inversion fits of the Fe I lnes are apparently equally good in both cases, LTE and non-LTE, and this is due to the capability of the inversion code to compensate the non-LTE effects by changing the derived iron abundance in the LTE inversion. Therefore, although the inversion fits for the Fe I in the IRIS^2 database are rather good, we should be aware that they may introduce an error in the photospheric atmosphere. However, it would be possible to consider the Fe I lines in 1D non-LTE during the inversion and to investigate the differences with the results presented in this paper. That study will be done in future developments of IRIS^2.

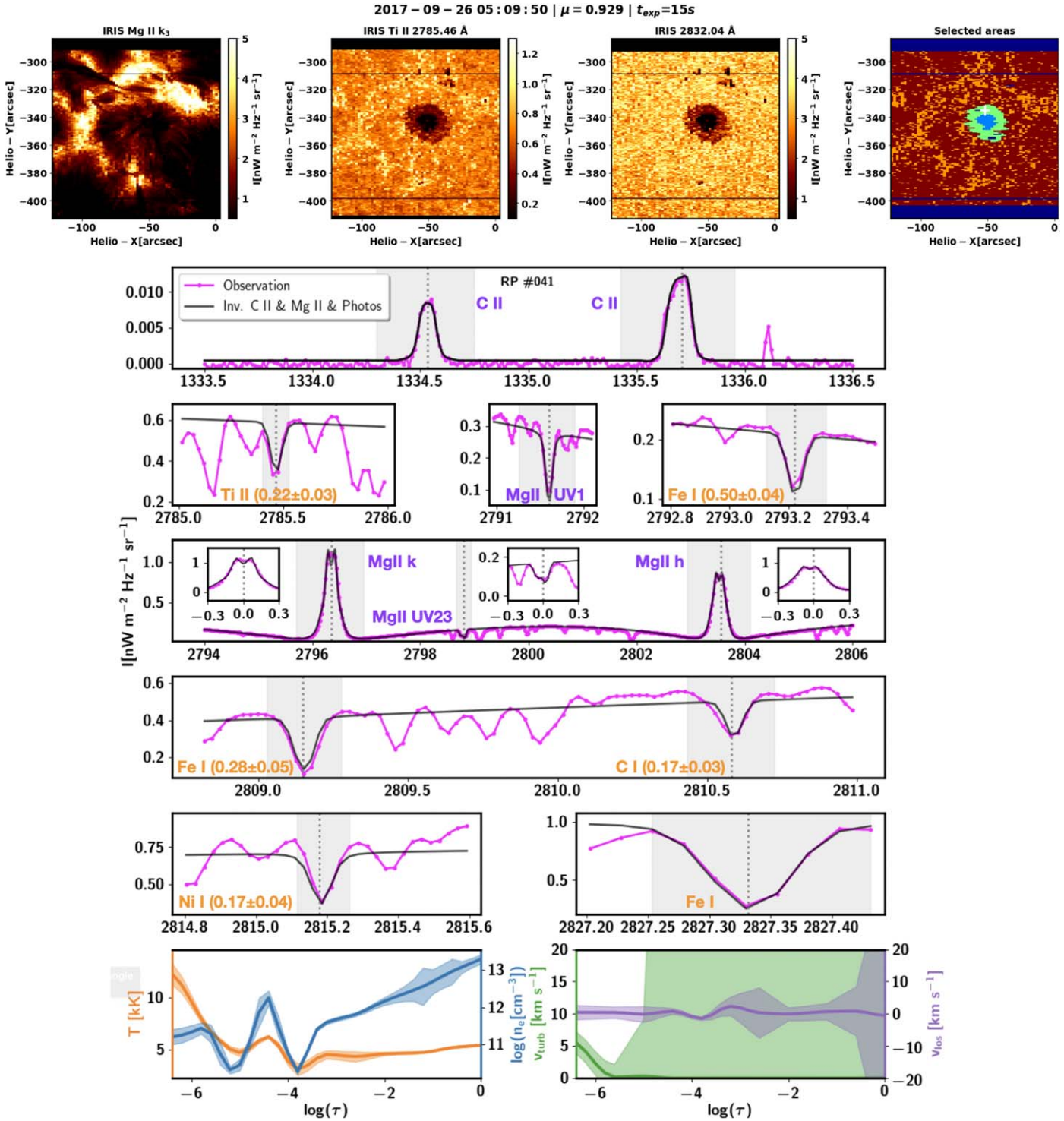

Figure 4. First row: Spectroheliogram map for Mg II h and k k_3, Ti I 2785.46 Å, the photospheric continuum at 2832.04 Å, and a map of the various types of features that were identified: umbra (light blue), penumbra (green), plage (orange), and quiet Sun (brown). The other rows are the results of an inversion for the location marked with a white cross on the map "Selected areas" to the right of the top row. From the second to the sixth row, from left to right, the observed (solid dotted in fuchsia) and the inverted profile (black) are shown. The value in parentheses is the average height (in megameters) at optical depth equal to 1. The gray shaded areas around the inverted lines show the spectral ranges considered for the calculation of the uncertainties. The bottom row shows the temperature (T, orange), the logarithm of the electron density (n_e, blue), velocity of turbulent motions or microturbulent (v_{turb}, green), and the line of sight (v_{los}, violet). The colored shaded area in these panels represents the uncertainty of each variable as described in the text (see Equation (2)).

3. Results

Figures 4 and 5 show the inversion of a multiline RP in the penumbra and the outer part of the plage of the NOAA 12681, respectively, both on 2017 September 26 at 05:09:50UT. In addition to the observed RP (in fuchsia) and its corresponding inverted, synthetic RP (best fit obtained by STiC, in black), the temperature (T), line-of-sight velocity (v_{los}), the velocity of turbulent motions or microturbulent velocity (v_{turb}), and the logarithm of the electron density (n_e) are shown in the last row of these figures. The rest position of the lines is indicated with a

Table 2
Number of Nodes in Each Cycle for the Thermodynamics Variables Considered during the Inversions

No. Cycles	1	2	3	4
T	4	7	9	13
$v_{\rm turb}$	2	4	8	13
$v_{\rm los}$	2	4	8	13

vertical dashed line. The inset plots in the fourth row show the core of the Mg II h and k and Mg II UV2 and 3 lines.

As a result of the strategy presented in this article, we have an indexed and labeled database. Thus, each inverted synthetic RP has an associated representative model atmosphere (RMA), as well as the following metadata: the IRIS filename and the solar feature ("umbra," "penumbra," "pore-like," "plage," or "quiet Sun") where the RP was clustered, the heliocentric angle of the original observed profiles associated with the RP's cluster, the exposure time, and the observation date and time. Because information is attached to the RP–RMA pair in IRIS^{2+}, one can search for RMAs for a given solar feature (e.g., "plage"), at a given time in the solar cycle (e.g., "2014-01-01" < *date_obs* < "2014-05-31"), and at given heliocentric angle on the Sun. Figure 6 shows an example of a search of the RMAs for all the solar features located at $0.65 < \mu < 0.70$. As one would expect, the RMAs show, for a given physical parameter, some variation both within the same solar feature (e.g., the $v_{\rm turb}$ for the RMAs in the plage), and between solar features (e.g., umbra versus penumbra). In this example search, only two umbras (30 RMA/umbra) and two penumbras (50 RMA/penumbra) were observed and included in the database. In fact, these two umbras and two penumbras observations belong to the same AR, and for that reason the dispersion between the RMAs for the different physical variables is rather small in these solar features. Several plages and quiet Sun areas belonging to different ARs are, however, found in this search, and for that reason their associated RMAs show a larger variability. We remind the reader that the term "quiet Sun" here refers to regions within an active region that are not plage, umbra, pores, or penumbra. It is not evident that such regions are, in fact, identical to quiet-Sun regions that are not part of an active region. This is because the overlying large-scale magnetic field topology and the associated canopy can affect the chromosphere of these regions, which are perhaps more like the quiet Sun at the photosphere.

The quality of the inversion is not always easy to evaluate from the value of χ^2 in the profiles of IRIS^{2+}. This is due to the behavior of the Euclidean distance when applied to high-dimensional samples such as our profiles. In many cases, a bad fit in the core of the Mg II h and k lines penalizes the value of χ^2 too much (see Section 3 of Sainz Dalda & De Pontieu 2023 for a more detailed discussion). Also, we now use multiple lines. This increases the possibility of having a bad χ^2 value when one or a few lines are poorly fitted, but the rest are well fitted. Of course, when all lines are well fitted, the χ^2 is good (i.e., low). We have verified that the large majority of profiles are well fitted, and for the minority of RPs with some badly fitted lines, several of the other lines are still well fitted.

We have calculated the uncertainties associated with the inverted profile and its corresponding model atmosphere. For each physical variable in the bottom panels of Figures 4 and 5, the uncertainty is displayed as a shaded area. The uncertainties were calculated using the following expression:

$$\sigma_p^2 = \frac{2}{{\rm nm} + r} \frac{\sum_{i=1}^{q} [I^{\rm obs}(\lambda_i) - I^{\rm syn}(\lambda_i; \boldsymbol{M})]^2 \frac{w_i^2}{\sigma_i^2}}{\sum_{i=1}^{q} R_p^2(\lambda_i) \frac{w_i^2}{\sigma_i^2}}, \quad (2)$$

where ${\rm nm} + r$ is the total number of free parameters used during the inversion, corresponding to n nodes along the optical depth in m variables (3×13), with r being the number of variables or free parameters considered constant along the optical depth ($r = 1$ corresponds to the upper boundary gas pressure). $Rp(\lambda_i)$ is the response function of the intensity to a perturbation in the physical parameter p of the model atmosphere $\boldsymbol{M}$. The formal derivation of this expression can be obtained by using the equations of Section 2.3 in Sánchez Almeida (1997) and of Sections 6.2 and 6.3 in Bellot Rubio et al. (1998). The meanings of the other symbols in Equation (2) are the same as those in Equation (1).

The gray shaded areas in Figures 4 and 5 indicate the spectral range considered in the calculation of uncertainties in the physical parameters. This range is slightly different from the one used for the inversions. The weights used for the inversions include some spectral positions between the lines, to take into account the general shape of the profile—for example, the bump between the Mg II k and Mg II h lines, or between the lines Fe I 2809.15 Å, C I 2810.58 Å.

4. Discussion

In the example of the IRIS^{2+}database shown in Figure 4, the inverted profiles fit the observed profiles quite well. Therefore, we can be confident with the RMA obtained. Regarding the temperature and the electron density, the uncertainties are relatively small. As for $v_{\rm los}$, the uncertainty is mostly acceptable at $\log(\tau) < -1$. However, $v_{\rm turb}$ has a very large uncertainty at any value of $\log(\tau) > -5.8$. In Figure 5, the profiles of the C II 1334 and 1335 Å, Mg II h and k, and Mg II UV triplet lines are relatively broad. The fit of these lines is good, although some features are missing in the C II 1334 and 1335 Å lines and in the Mg II UV triplet lines. The fit of the photospheric lines is also acceptable, but not as good as the chromospheric lines. The uncertainties of temperature and $n_{\rm e}$ are within acceptable limits. However, for $v_{\rm los}$, it has large uncertainties at some optical depths, and for $v_{\rm turb}$, it again has very large uncertainties for $\log(\tau) > -2$, but acceptable ones for $\log(\tau) < -4$. We cannot offer a plausible explanation for these very high uncertainties, other than the possibility that the uncertainties in $v_{\rm los}$ and $v_{\rm turb}$ are overstated due to the poor sensitivity of the lines to these variables at such heights. As suggested by Sainz Dalda & De Pontieu (2023), to infer the uncertainties of the model, we can also run a Monte Carlo experiment on the synthetic inverted profile by adding random noise and inverting the noisy synthetic profile 25 times (five times for five different noisy profiles). The uncertainty associated with the model is the standard deviation of the model resulting from the 25 inversions. These authors used this method for the entire IRIS^2 database, considering several exposure times. In total, approximately 1.25 million inversions were made to provide the uncertainties associated with RMAs. A similar approach would require approximately 7.5 million CPU hours for the

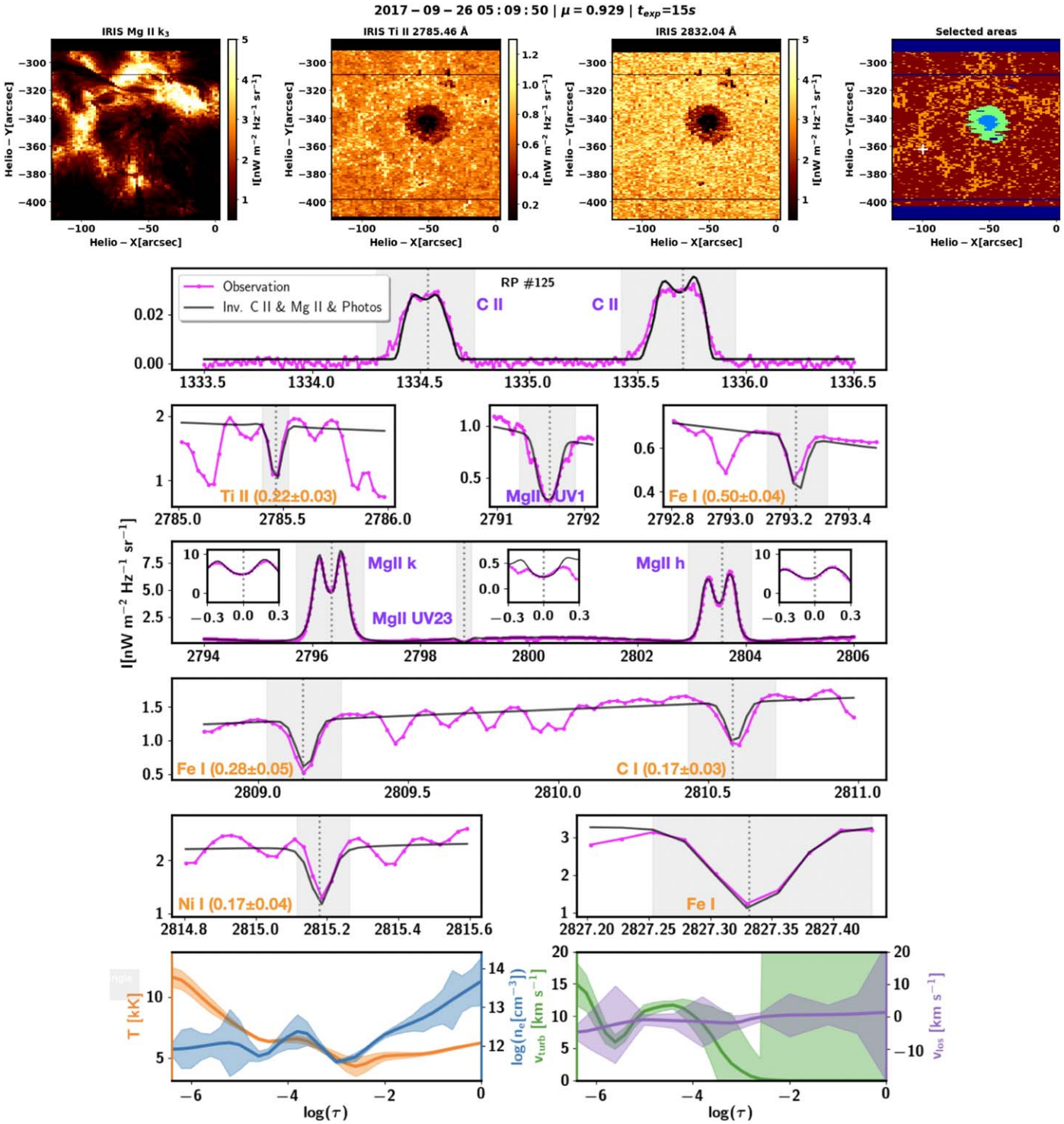

Figure 5. Same as Figure 4 for a multiline RP and RMA in the outer plage (marked with a white cross in the "Selected areas" panel). Both the inverted profile and the representative model atmosphere (RMA) are included in the IRIS^{2+}, and they are both labeled as "`plage`" in the database.

IRIS^{2+} database. Due to this computationally expensive process, we were unable to provide the associated uncertainties with the RMAs of the IRIS^{2+} database at the time of publication of this paper.

We note that the simultaneous inversion of spectral lines sensitive to different conditions in the solar atmosphere represent a numerical challenge in a high-dimensional space. In this case, Equation (1) for the evaluation of the goodness of fit has some flaws, since a large value of χ^2 can be associated with a fit having some lines fitted quite well and others not. Graphically, we can see this in Figure 5. There, the lines Fe I 2793.22 Å, Fe I 2809.15 Å, C I 2810.58 Å, and Ni I 2815.18 Å show some mismatch, while the other lines fit quite well. In this case, we should consider the values at optical depths $-3 < \log(\tau) < -1$ with some caution, because only the photospheric Fe I 2827.33 Å fits the observed profile quite well, in addition to the chromospheric lines. Therefore, we

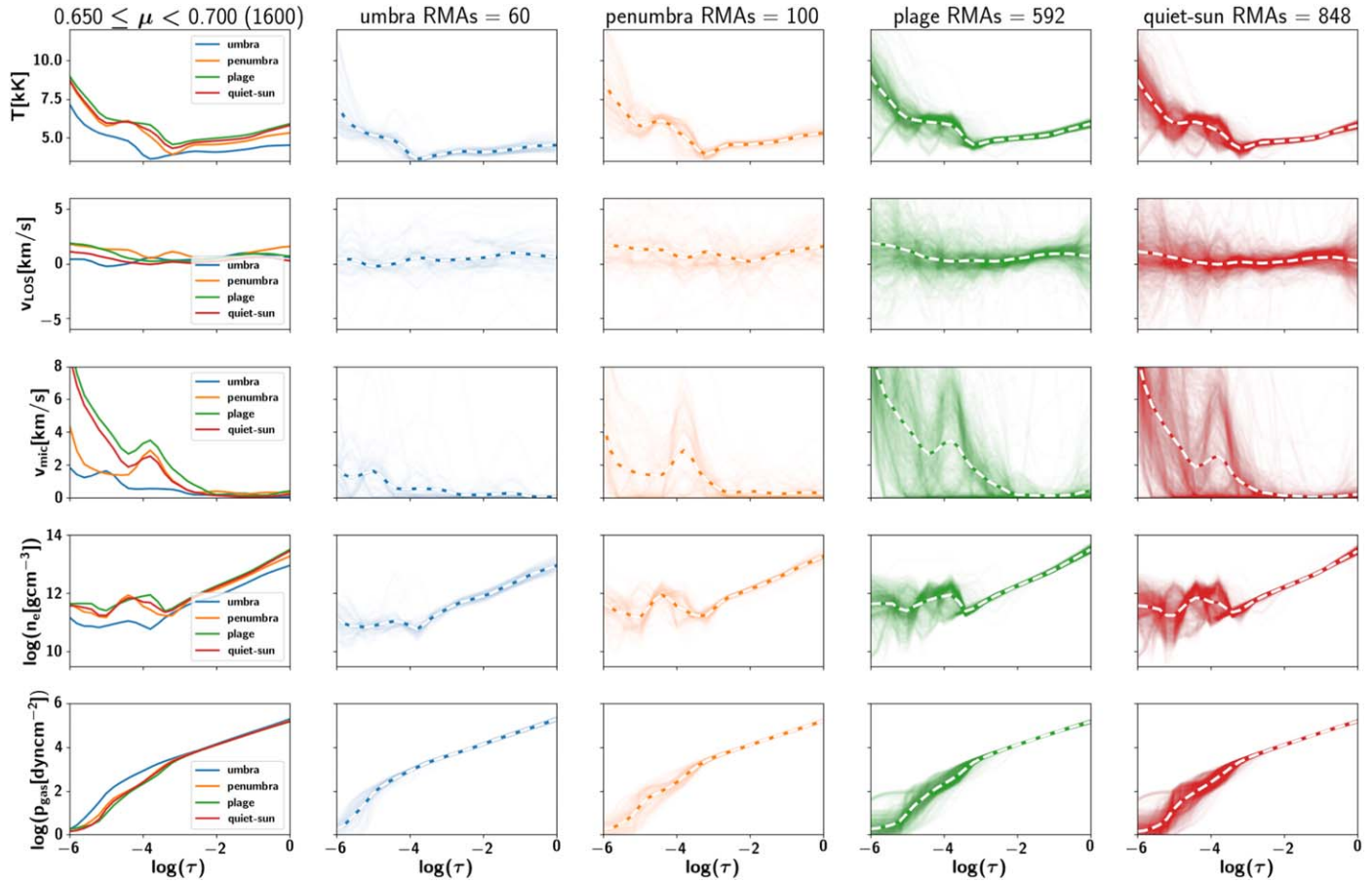

Figure 6. Using IRIS^{2+} database to find and visualize the different models of umbra, penumbra, plage, and quiet Sun at a given interval of $\mu = \cos(\theta)$, with θ being the heliocentric angle.

suggest maintaining a critical attitude when analyzing the inversions. On the other hand, how does the fit between the observed and the inverted affect the IRIS^{2+}database? Strictly speaking, it does not. The IRIS^{2+}database is a relational database between model atmospheres and their synthetic profiles. The physics encoded in the latter is strictly due to the values in the model atmosphere and the considerations made to solve the RT problem. Practically speaking, it has an impact if the fit is consistently bad, which is not the case. If so, that would mean that the synthetic inverted RPs and the RMAs are not representing well the observations, i.e., the solar lower atmosphere.

IRIS^{2+} can be used in different ways. The RMAs can be used as a set of reference model atmospheres for the various types of features within ARs. They can be used to synthesize photospheric and chromospheric spectral lines that can be compared either with observed lines or with the ones obtained from the synthesis of numerical simulations of these features. Although the stratifications of the atmosphere in the numerical simulations are rather different from those obtained by the inversions, with the former showing very variable behavior and the latter a smoother one, the values obtained from the inversions may help to constrain the mean behavior in some regions in the stratification of the numerical models. The IRIS^{2+} database can be used as a look-up table to invert the multiline IRIS data that contain any combination of the lines used in this article (see Table 1). A new database of synthetic RPs of chromospheric or photospheric lines can be created to invert observations made in the lines of interest. In both cases, the inversion time using the RP-RMAs database is considerably smaller compared to the time required to invert these lines by iteratively solving the RT problem considering either LTE or non-LTE. If the computational time required to invert the lines of interest is not too high (a few CPU minutes or less), the solution of the inversion proposed here can be used as an initial guess for the classical inversion of each profile of the observation.

5. Conclusions

The extraordinary multiline capabilities of IRIS has allowed us to create a database of 40,320 synthetic representative profiles and their corresponding representative model atmospheres that convey the essential thermodynamic information of active regions from the bottom of the photosphere to the top of the chromosphere.

Thanks to the state-of-the-art multiline and multiatom STiC inversion code, we were able to recover the stratification of physical parameters encoded in six chromospheric lines and six photospheric lines. Using the k-means technique, we have clustered multiline IRIS spectral data. This helps us to overcome the task of inverting all the profiles considered in this work, which would require very significant computational resources (≈38 million CPU hours).

Achieving an accurate representation of thermodynamics in the lower solar atmosphere is of great importance. The accuracy of the fits directly correlates with the quality of the

representation. However, synthetic profiles that display poor fits with respect to the observed RP may still exist in the solar atmosphere. Despite this, the IRIS^{2+} database is probably the most accurate and comprehensive set of stratified depth models and profiles in the chromosphere and the photosphere, as a vast majority of inversions calculated to build the database exhibit rather good fits.

IRIS^{2+} is a unique database that combines a large number of spectral lines that are sensitive to thermodynamics in the lower layers of the solar atmosphere. We are constantly working to improve the observational capabilities of IRIS by considering more lines, applying better inversion methods, and utilizing advanced machine-learning techniques. IRIS^{2+} has high potential to be used in various ways to gain knowledge about the low solar atmosphere. Both the IRIS^{2+} and IRIS^{2} databases are available to the public in different formats, readable both with IDL (Landsman 1993) and Python (Van Rossum & Drake 1995). We encourage the community to take advantage of IRIS^{2+}, to help answer open questions about the low solar atmosphere.

Acknowledgments

The authors thank the anonymous reviewer for providing corrections, suggestions, and comments that have contributed significantly to improving the content of this article. IRIS is a NASA small explorer mission developed and operated by LMSAL with mission operations executed at NASA Ames Research Center and major contributions to downlink communications funded by ESA and the Norwegian Space Agency. This work was supported by NASA contract NNG09FA40C (IRIS). A.A. was supported by the Lockheed Martin–Palo Alto Unified School District internship program. The authors are grateful to Mats Carlsson and Tiago Pereira for providing the C atom model, and to Jaime de la Cruz Rodríguez for his helpful comments on the inversion of photospheric lines. This investigation used basic Python packages including `scikit-learn` (Pedregosa et al. 2011), `Matplotlib` (Hunter 2007), and `NumPy` (Harris et al. 2020).

Software: IRIS^{2} (https://iris.lmsal.com/iris2/), IRIS^{2+}.

ORCID iDs

Alberto Sainz Dalda https://orcid.org/0000-0002-3234-3070
Bart De Pontieu https://orcid.org/0000-0002-8370-952X
Milan Gošić https://orcid.org/0000-0002-5879-4371

References

Athay, R. G., & Lites, B. W. 1972, ApJ, 176, 809
Bellot Rubio, L. R., Ruiz Cobo, B., & Collados, M. 1998, ApJ, 506, 805
Bjørgen, J. P., Leenaarts, J., Rempel, M., et al. 2019, A&A, 631, A33
Collados, M., Martinez Pillet, V., Ruiz Cobo, B., del Toro Iniesta, J. C., & Vazquez, M. 1994, A&A, 291, 622
de la Cruz Rodríguez, J., Leenaarts, J., & Asensio Ramos, A. 2016, ApJL, 830, L30
de la Cruz Rodríguez, J., Leenaarts, J., Danilovic, S., & Uitenbroek, H. 2019, A&A, 623, A74
de la Cruz Rodríguez, J., & Piskunov, N. 2013, ApJ, 764, 33
del Toro Iniesta, J. C., & Ruiz Cobo, B. 2016, LRSP, 13, 4
del Toro Iniesta, J. C., Tarbell, T. D., & Ruiz Cobo, B. 1994, ApJ, 436, 400
Fontenla, J., White, O. R., Fox, P. A., Avrett, E. H., & Kurucz, R. L. 1999, ApJ, 518, 480
Fontenla, J. M., Avrett, E., Thuillier, G., & Harder, J. 2006, ApJ, 639, 441
Fontenla, J. M., Avrett, E. H., & Loeser, R. 1993, ApJ, 406, 319
Gingerich, O., Noyes, R. W., Kalkofen, W., & Cuny, Y. 1971, SoPh, 18, 347
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Natur, 585, 357
Hunter, J. D. 2007, CSE, 9, 90
Khomenko, E., Collados, M., Díaz, A., & Vitas, N. 2014, PhPl, 21, 092901
Landi Degl'Innocenti, E., & Landi Degl'Innocenti, M. 1981, NCimB, 62B, 1
Landsman, W. B. 1993, in ASP Conf. Ser. 52, Astronomical Data Analysis Software and Systems II, ed. R. J. Hanisch, R. J. V. Brissenden, & J. Barnes (San Francisco, CA: ASP), 246
Leenaarts, J., Pereira, T., & Uitenbroek, H. 2012, A&A, 543, A109
Martínez-Sykora, J., de la Cruz Rodríguez, J., Gošić, M., et al. 2023, ApJL, 943, L14
Martínez-Sykora, J., Leenaarts, J., De Pontieu, B., et al. 2020, ApJ, 889, 95
Mein, P. 1971, SoPh, 20, 3
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, JMLR, 12, 2825
Pereira, T. M. D., Leenaarts, J., De Pontieu, B., Carlsson, M., & Uitenbroek, H. 2013, ApJ, 778, 143
Piskunov, N., & Valenti, J. A. 2017, A&A, 597, A16
Quintero Noda, C., Khomenko, E., Collados, M., et al. 2023, A&A, 675, A93
Ruiz Cobo, B., Quintero Noda, C., Gafeira, R., et al. 2022, A&A, 660, A37
Sainz Dalda, A., & De Pontieu, B. 2023, FrASS, 10, 1133429
Sainz Dalda, A., de la Cruz Rodríguez, J., De Pontieu, B., & Gošić, M. 2019, ApJL, 875, L18
Sánchez Almeida, J. 1997, ApJ, 491, 993
Shchukina, N., & Trujillo Bueno, J. 2001, ApJ, 550, 970
Smitha, H. N., Holzreuter, R., van Noort, M., & Solanki, S. K. 2020, A&A, 633, A157
Smitha, H. N., Holzreuter, R., van Noort, M., & Solanki, S. K. 2021, A&A, 647, A46
Sukhorukov, A. V., & Leenaarts, J. 2017, A&A, 597, A46
Uitenbroek, H. 2001, ApJ, 557, 389
Van Rossum, G., & Drake, F. L., Jr 1995, Python 3 Reference Manual (Scotts Valley, CA: CreateSpace)
Vernazza, J. E., Avrett, E. H., & Loeser, R. 1981, ApJS, 45, 635
Wargnier, Q. M., Martinez-Sykora, J., Hansteen, V. H., & De Pontieu, B. 2023, ApJ, 846, 115
Woods, M. M., Sainz Dalda, A., & De Pontieu, B. 2021, ApJ, 922, 137