



Local departures from IRI-2026 and observation-based correction at Hainan Fuke

Zheng Wang^{1,2}, Ziyue Wang¹, Jinhui Cai^{3,*}, Guojun Wang^{1,2}, Chu Qiu³, Jiankui Shi¹, Dabin Xue¹, Pengdong Gao³, and Xiao Wang^{1,2}

1 State Key Laboratory of Solar Activity and Space Weather, National Space Science Center, Chinese Academy of Sciences, Beijing, China

2 National Field Observation and Research Station (Hainan) for Space Weather, Hainan, China

3 Key Laboratory of Media Audio & Video (Communication University of China), Ministry of Education, Beijing, China

Correspondence: Jinhui Cai (caijinhui@cuc.edu.cn)

Abstract. Regional use of the International Reference Ionosphere (IRI) requires knowing how much historical calibration reduces local errors and which conditions remain poorly represented. We examine these questions at Hainan Fuke using 2014-2020 observations and a fixed IRI-2026 configuration. The targets are the F2-layer critical frequency (foF2), which reflects peak electron density, and peak height (hmF2). Corrections learned from 2015-2018 are evaluated in later years. In 2020, a simple month-hour correction reduced foF2 mean absolute error from 1.137 to 0.997 MHz, slightly below the error of the random forest of 1.013 MHz. For hmF2, the forest reduced error from 28.634 to 21.754 km, compared with 22.941 km for the simple correction. The forest's additional benefit therefore differs between the peak parameters. Across recorded spread-F classes, which describe broadened ionogram echoes, estimated percentage gains in foF2 exceeded those without spread F. Yet the corrected error during strong range spreading remained 3.622 MHz, compared with 1.037 MHz without spreading. Height showed its largest estimated gain without spread F. A separate retrospective update using the preceding 15-minute residual gave errors of 0.256 MHz and 10.851 km. For this Hainan record, method choice depends on the target and on whether average absolute error or mean bias matters most. Strong spreading retains substantial frequency errors despite larger relative gains, making the remaining errors for each parameter and state important for regional use. Recent observations further improve local estimates, although an operational forecast horizon remains untested.

1 Introduction

The ionospheric F2 layer contains a maximum in electron density that affects high-frequency radio propagation. Two quantities describe this peak: foF2 is the critical frequency associated with its density, and hmF2 is the altitude at which it occurs. A model can estimate one more accurately than the other. For regional studies and monitoring, it is therefore useful to know both the size of the local errors and how much observations can reduce them. At Hainan, a low-latitude station, this question also concerns spread F: the broadening of ionogram echoes that accompanies irregular ionospheric structure. Strong range spreading has been observed together with plasma bubbles and scintillation in this region (Wang et al., 2015).



35 The International Reference Ionosphere (IRI) provides a widely used reference for such comparisons. Developed through international cooperation under the Committee on Space Research (COSPAR) and the International Union of Radio Science (URSI), it combines ground-based and space-based measurements to describe electron density, plasma temperatures and ion composition (Bilitza et al., 2014). Its core describes monthly average conditions, including their regular changes with local time, season and solar activity
40 (Bilitza et al., 2022). It is thus a model of the usual ionosphere rather than a direct record of the state at every observation time. IRI also includes empirical storm adjustments, whose performance varies between regions (Miró Amarante et al., 2007). Local observations remain useful for identifying departures that the selected model settings do not reproduce.

Earlier work has already established that Hainan requires attention to its local ionospheric behavior.
45 Ratovsky et al. (2014) used manually scaled ionograms to construct local models of peak density and height, showing different responses to local time, season and solar activity. Wang et al. (2017) compared Hainan observations from 2002-2012 with IRI-2012 and found that the model generally underestimated foF2. Their height comparison used measured M(3000)F2, an ionosonde parameter that supplies information about the layer profile. Underestimation of foF2 has also been reported at Pu'er in the Chinese
50 equatorial ionization anomaly region (Zhu et al., 2020). These studies establish the regional background, but do not tell us how much a correction learned from earlier Hainan records improves estimates at later observation times, or whether a random forest correction improves on a simple seasonal and daily adjustment.

Observations have long been used to update IRI. Galkin et al. (2012) developed a framework that adjusts
55 its background maps using recent ionosonde measurements. Other approaches infer effective activity indices from ionosonde data (Pignalberi et al., 2018) or use vertical total electron content (VTEC) from Global Navigation Satellite System (GNSS) measurements (Pignalberi et al., 2019). VTEC measures electrons integrated through a vertical column, whereas foF2 and hmF2 describe one part of the profile. These observational constraints need not improve all parameters equally. In the South African GNSS
60 study, foF2 generally improved, but hmF2 became less accurate in the independent-year comparison (Pignalberi et al., 2019). A wider comparison of IRI with IRI Real-Time Assimilative Mapping (IRTAM) likewise found that improvements depended on the peak parameter and the observations used for evaluation (Pignalberi et al., 2021).

Machine learning provides another way to relate past measurements to model errors. In a five-station
65 comparison of daily foF2, Atıcı and Pala (2026) found that its benefit relative to IRI-2020 depended on station and activity conditions; their tables include cases in which IRI had the smaller error. This makes a simple local correction an important comparison. If an adjustment by month and hour performs similarly to a random forest, the simpler scheme may be sufficient for that target and accuracy measure. A separate comparison with the preceding observation then asks a different question: how much does knowledge of



70 the recent local state help? Short-term foF2 studies have used persistence for this purpose for many years (Chan and Cannon, 2002).

The Hainan record also allows the correction to be examined under different spread-F states. Frequency and range spreading occur under different nighttime and seasonal conditions at Chinese stations (Wang et al., 2018). A reduction in average model error may therefore conceal large remaining errors in some of
75 these states. Examining foF2 and hmF2 separately is useful because a change in peak density need not be accompanied by the same change in peak altitude. The question is whether the corrections describe both quantities well under the conditions actually present in the record.

We use the Hainan Fuke record from 2014-2020 to determine how much of the local IRI-2026 error can be reduced by historical calibration and where large errors remain. A correction by month and hour provides a
80 simple local reference; a random forest trained on the same years provides a second historical correction. We assess both peak parameters and then examine their errors across the recorded spread-F states. A separate update using the preceding 15-minute residual measures the benefit of recent local information, while GNSS VTEC supplies supporting column context. These comparisons inform the choice of a local reference and show which observed conditions require particular care when using it.

85 **2 Methods**

We compare two uses of observations. Historical calibration learns a correction from earlier years and applies it without a recent measurement. An update from a recent observation instead carries the last measured model error forward to the next time bin. We call the observed value minus the IRI estimate the residual. This definition lets us correct the same IRI background in each comparison. For each parameter
90 and evaluation year, all methods are assessed at the same available times.

2.1 Station and observations

Hainan Fuke is a low-latitude ionosonde-GNSS station. Its antenna reference point, estimated from RINEX records, is 19.527 deg N, 109.134 deg E, at 92 m WGS84 ellipsoidal height. The historical station record gives 19.5 deg N, 109.1 deg E and 89 m. IRI calculations used the fixed rounded coordinates of 19.5 deg N
95 and 109.1 deg E. At the study midpoint, 2 July 2017, IGRF-14 gives a centered-dipole geomagnetic latitude of 10.0 deg N, a downward inclination of 28.0 deg and an inclination-derived dip latitude of 14.9 deg N (Beggan et al., 2026). The two magnetic latitudes use different definitions. Model times were UTC. Where required as a predictor, local solar time was calculated by adding 109.1/15 h to UTC and taking the result modulo 24 h.

100 We analyzed available records from January 2014 to December 2020. An ionogram records radio echo delay as a function of sounding frequency. Experienced researchers inspected and edited traces processed



with ARTIST 5002, then exported foF2 and hmF2 using SAO Explorer 3.4.17. After trace editing, its NHPC inversion converts the delay-based trace into a profile against true altitude, from which hmF2 is obtained (UMLCAR, n.d.). For ambiguous spread-F traces, researchers used preceding and following
105 ionograms and continuous points in the current image, or followed the lower envelope of the ordinary-mode (O-mode) echo displayed in red. A separate model for classifying images assigned the SF labels (Wang et al., 2023). The classes are noSF (no spread F), Freq (frequency spreading), Range (range spreading), SRange (strong range spreading) and Mix (mixed spreading). We use these labels to group the errors after correction; the labels are not inputs to the correction models.

110 Native ionosonde timestamps were assigned to 15-minute bins by rounding down. Within each occupied bin, numeric parameters were aggregated by the median of available values and the five-class SF label by its mode; ties use the first label in sorted order. Missing numeric observations were retained as missing, without interpolation. Values outside the specified ranges were flagged but included in the correction analysis. Each parameter comparison requires finite observed and modeled values. Appendix A gives the
115 flag definitions and temporal matching details.

GNSS VTEC was taken from the Meridian Project vertical absolute-TEC product (ATEC). After removal of encoded invalid records, values were first aggregated across satellites by their median at each available observation epoch. The comparison value at ionosonde bin start t is the median of those epoch medians within $[t-7.5 \text{ min}, t+7.5 \text{ min}]$. This matches VTEC to approximately the same observation time in a
120 retrospective comparison. No additional cutoff in satellite elevation was applied. Scintillation indices are not used in this analysis.

The historical Hainan Fuke observations are catalogued under Danzhou Station in the Chinese Meridian Project data listings. The source products include ionogram images, ionospheric parameters, GNSS RINEX records and vertical absolute TEC (Chinese Meridian Project, n.d.-a, n.d.-b, n.d.-c, n.d.-d). The
125 peak parameters used here come from the manual tracing described above, rather than the unreviewed catalogue parameters.

2.2 IRI background and temporal partitions

We used 2015-2018 to train the corrections and calculate the medians used to fill missing predictors. The validation and evaluation years were 2019 and 2020, respectively. Records from 2014 contribute only to
130 the comparison over the full period with uncorrected IRI. IRI-2026 supplies the primary results; IRI-2020 provides a historical comparison using the same observations, times, predictors and fixed settings. The 2019 and 2020 observations had already been examined before the IRI-2026 comparison.

The IRI configuration was held fixed so that each correction started from the same reference. The selected frequency is the quiet-map foF2, which represents the background before the empirical storm adjustment.
135 We describe the settings below because the reported gains apply to this particular reference.



We used the official IRI-2026 Fortran distribution (IRI Working Group, 2026), with the historical solar and geomagnetic index files held fixed. The selected options were URSI foF2, Shubin-COSMIC hmF2, ABT-2009 bottomside parameters, IRI-COR2 topside, and the Ozhogin plasmasphere option without a plasmopause. The TEC upper integration limit remained 1000 km. GNSS VTEC includes the electron
140 column up to satellite altitude, so its difference from this IRI column includes both error in the modeled profile and the contribution above the selected integration limit (Bilitza et al., 2022; Pignalberi et al., 2025). For comparison with the historical calculation, the foF2 baseline uses the quiet-map value,

reconstructed as $\sqrt{\frac{N_{mF2}}{1.24 \times 10^{10}}}$, where NmF2 is the returned storm-adjusted peak density in m^{-3} and s is
s

the returned foF2 storm multiplier. The returned storm-adjusted NmF2 was retained separately as a
145 predictor. All reported foF2 gains refer to this quiet-map baseline, not to every available IRI configuration. Electron density profiles used 80-1000 km at 10-km spacing; native TEC was computed separately over 0-1000 km with 0.1-km integration steps. Native TEC in electrons per square metre was divided by 10^{16} to obtain TECU. Appendix A lists the complete switch configuration.

2.3 Historical correction and the preceding residual

150 The random forest averages predictions from many decision trees. Here it learns how the residual varies with the IRI estimates and the time of year and day. The month-hour scheme uses median residuals, whereas the forest uses additional IRI predictors and a squared-error fitting objective. Their comparison evaluates these two schemes as implemented. We fit frequency and height separately.

For each target y , we fitted the residual $r(t) = y_{obs}(t) - y_{IRI}(t)$ and obtained the corrected estimate
155 $\hat{y}(t) = y_{IRI}(t) + \hat{r}(t)$. Here the hat denotes an estimated residual. The random forest uses four IRI

quantities (foF2, hmF2, VTEC, and NmF2), sine and cosine of $\frac{2\pi DOY}{365.25}$, sine and cosine of $\frac{2\pi LST}{24}$,

and $\frac{year - 2015}{6}$. DOY is day of year and LST is local solar time. Missing features were filled using

medians from the training period. Separate forests for foF2 and hmF2 used 100 trees, maximum depth 15, minimum leaf size 10, squared-error splitting and random state 42. No further tuning was performed.
160 Historical ionosonde observations supply the training targets; inference uses IRI and calendar quantities without a recent observation sequence, current ionosonde value, SF label or GNSS disturbance feature.

The month-hour correction adds the median training residual for the same UTC month and hour to the IRI estimate. It asks whether a recurring seasonal and daily offset is enough to explain the benefit of calibration. The recent update uses the observed residual from exactly the preceding 15-minute bin.



165 Equivalently, $\hat{y}(t) = y(t - \Delta t) + [IRI(t) - IRI(t - \Delta t)]$, where $\Delta t = 15$ min and y is the observed parameter. Thus, it starts from the previous observation and adds the change predicted by IRI over the interval. If that preceding residual is missing, an older record is not substituted. To compare all four estimates, including uncorrected IRI, we retain only times with finite observed and modeled values and an available preceding residual. The recent update receives an additional observation that the historical
170 corrections do not use. Its 15-minute bin spacing is not an operational forecast lead, which would also depend on acquisition, tracing and delivery time.

2.4 Evaluation and uncertainty

We use mean absolute error (MAE) to report the average size of the error in MHz or km. Root mean square error (RMSE) gives greater weight to large errors. Bias shows whether the model is systematically
175 high or low: a positive bias (observed minus predicted) means underestimation. Relative improvement expresses the MAE reduction as a percentage of the uncorrected IRI error.

The coefficient of determination R^2 for the uncorrected IRI assessment uses absolute observed targets; the comparator tables calculate R^2 in residual space. Relative MAE reduction is
$$100 \times \frac{MAE_{IRI} - MAE_{corrected}}{MAE_{IRI}}.$$
 Its uncertainty was estimated using 2000 bootstrap replicates,
180 resampling complete dates with replacement and using a base seed 20260809 with deterministic offsets by target, split and stratum. The intervals quantify the correction gain within each class. Differences between classes, methods or model versions are descriptive unless a direct comparison interval is reported.

For clarity, the two coefficients of determination use different reference variances. With N paired samples, y denoting the observed target and r its residual, hats denoting estimates and overbars denoting sample
185 means, they are:

$$R^2_{absolute} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}, \quad R^2_{residual} = 1 - \frac{\sum_i (r_i - \hat{r}_i)^2}{\sum_i (r_i - \bar{r})^2} \quad (1)$$

Both sums run over $i = 1, \dots, N$. Within a common sample, adding the IRI background gives the same prediction error in either space, but the denominators differ. Table 1 uses the absolute-target definition; Table 2 uses the residual definition. Their R^2 values should therefore not be compared directly.

190 The correction comparisons concern foF2 and hmF2. GNSS VTEC describes the column difference from IRI and is not an input to the corrections evaluated here.



3 Results

3.1 Raw IRI departures

In 2020, the uncorrected IRI estimates were low for foF2 and high for hmF2 on average (Fig. 1 and Table 1). The biases (observed minus modeled) were 0.439 MHz and -16.065 km, and the corresponding MAEs were 1.130 MHz and 28.707 km. Thus the two peak parameters did not share the same direction of error. GNSS VTEC exceeded the IRI column by 0.910 TECU on average, with an MAE of 3.515 TECU. These values use all available pairs for each parameter. The correction comparisons below use the smaller set of times at which a preceding residual is also available.

Across 2014-2020, the corresponding MAEs for uncorrected IRI were 1.254 MHz for foF2 (222,206 pairs on 2442 dates), 27.747 km for hmF2 (217,686 pairs on 2442 dates), and 4.614 TECU for VTEC (157,915 pairs on 1697 dates). Mean biases were 0.651 MHz, -7.257 km and 3.085 TECU, respectively. Appendix B gives RMSE and R² over the full period, together with baseline errors by SF state and date partition.

Table 1. IRI-2026 baseline performance in the 2020 evaluation year. Bias is observed minus IRI; R² uses absolute-target values. The foF2 baseline is the quiet-map value defined in Sect. 2.2. Each target uses all available pairs.

Target	Unit	N	Dates	Bias	MAE	RMSE	R ² for observed parameters
foF2	MHz	31,586	355	0.439	1.130	1.506	0.756
hmF2	km	30,833	355	-16.065	28.707	36.143	0.049
VTEC	TECU	32,603	355	0.910	3.515	4.751	0.772

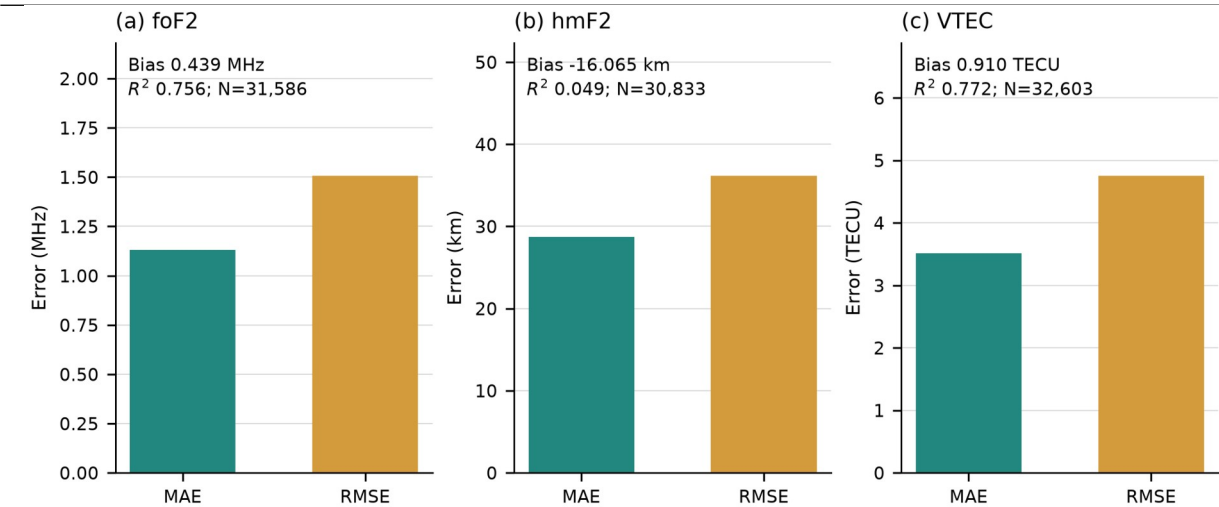


Figure 1. IRI-2026 baseline errors in the 2020 evaluation year. Panels separate foF2, hmF2 and VTEC units. Bars show MAE and RMSE; annotations give bias (observed minus modeled), R² for the observed parameters and paired sample size. The foF2 baseline uses the quiet-map value; each target uses its available pairs.



210 3.2 Historical calibration and recent observations

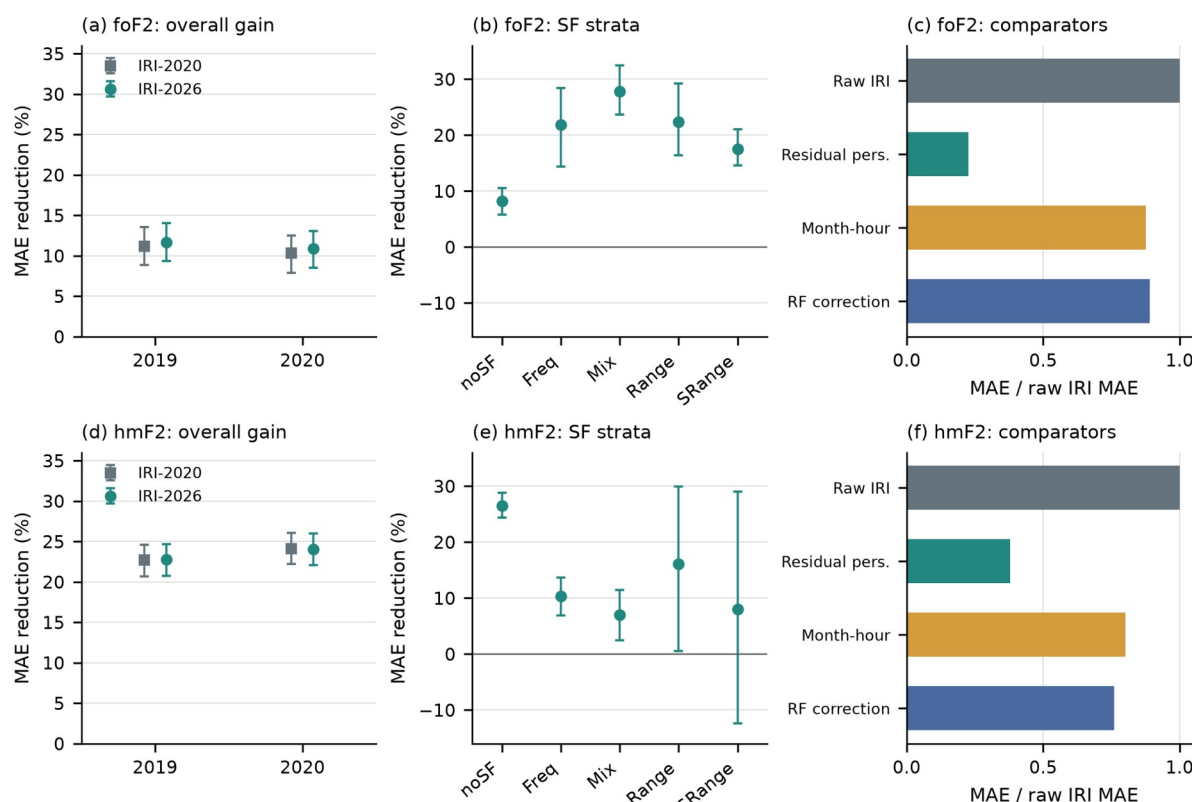
Historical calibration reduced the errors in both years, with a larger percentage reduction for height than for frequency (Table 2 and Fig. 2). In 2020, the forest reduced foF2 MAE from 1.137 to 1.013 MHz, a reduction of 0.124 MHz or 10.9% (95% interval 8.5-13.1%; 31,003 samples on 355 dates). The hmF2 MAE fell from 28.634 to 21.754 km, a reduction of 6.880 km or 24.0% (22.1-26.0%; 29,686 samples on 355 dates). The 2019 comparison showed the same direction: foF2 MAE fell from 1.068 to 0.943 MHz (11.7%, 9.3-14.1%; 30,158 samples on 354 dates), and hmF2 MAE from 27.602 to 21.319 km (22.8%, 20.8-24.7%; 29,579 samples on 354 dates).

The forest offered little additional reduction beyond the simple month-hour correction. For foF2, the simple correction gave an MAE of 0.997 MHz, slightly below the forest's 1.013 MHz. For hmF2, its MAE was 22.941 km, compared with 21.754 km for the forest. The differences between methods were therefore 0.016 MHz in favor of the simple correction and 1.187 km in favor of the forest. These are descriptive differences; the intervals reported above quantify the forest's gain over uncorrected IRI.

The update using the preceding residual had the lowest MAE: 0.256 MHz for foF2 and 10.851 km for hmF2. It uses an additional recent measurement, so this result shows the value of recent local information rather than a gain from a more complex historical fit. Appendix C gives the differences between methods and the corresponding MAE and RMSE plots.

Table 2. Correction performance on the same samples in 2020. The IRI-2026 baseline uses the fixed options in Sect. 2.2, including quiet-map foF2. Errors are in MHz for foF2 (N=31,003; 355 dates) and km for hmF2 (N=29,686; 355 dates). Bias is observed minus predicted; R² is in residual space. Persistence carries the exact preceding 15-min residual.

Target	Model	Bias	MAE	RMSE	R ² in residual space
foF2	Raw IRI	0.454	1.137	1.514	-0.099
foF2	Residual persistence	0.003	0.256	0.347	0.942
foF2	Month-hour	-0.244	0.997	1.363	0.110
foF2	Random forest	0.153	1.013	1.393	0.070
hmF2	Raw IRI	-16.484	28.634	36.011	-0.265
hmF2	Residual persistence	-0.488	10.851	15.360	0.770
hmF2	Month-hour	-10.193	22.941	29.262	0.165
hmF2	Random forest	-1.598	21.754	28.053	0.232



230

Figure 2. Local correction of foF2 (top row) and hmF2 (bottom row). Both IRI versions use the fixed options and quiet-map foF2 defined in Sect. 2.2. Panels (a,d) show the MAE reduction from the random forest relative to the corresponding raw IRI version in 2019 and 2020. IRI-2026 provides the primary results; IRI-2020 is a historical comparison. Panels (b,e) show IRI-2026 gains by SF state in 2020. In plotted order (noSF, Freq, Mix, Range, SRange), foF2 date counts are 355, 295, 244, 58 and 21; hmF2 date counts are 355, 293, 239, 58 and 20. Error bars are 95% intervals from 2000 bootstrap replicates that resample complete dates; they describe gains within each model version and SF class. Panels (c,f) show comparator MAE divided by raw-IRI MAE on the same samples. Persistence uses the residual from the exact preceding 15-minute bin. Appendix B provides class sample counts and absolute errors.

3.3 Correction across recorded spread-F states

Frequency correction had a larger estimated percentage benefit in every recorded SF class than in noSF (Fig. 2b). The reduction was 8.2% in noSF (95% interval 5.8-10.6%), 21.9% in frequency spreading and 27.8% in mixed spreading. Yet a larger percentage improvement could leave a large error. In strong range spreading, foF2 MAE fell from 4.392 to 3.622 MHz, whereas the corrected noSF error was 1.037 MHz. Strongly spread echoes therefore remained a difficult setting for the local frequency estimate.

Height showed a different pattern (Fig. 2e). Its largest estimated gain was 26.5% in noSF (24.3-28.8%), compared with 10.3% in frequency spreading and 7.0% in mixed spreading. For strong range spreading,



the estimated gain was 8.0%, but the interval ranged from -12.4% to 29.1%, based on 70 samples on 20 dates. The data therefore leave the benefit of height correction uncertain in this small subset. Appendix B gives all class errors and sample counts. The intervals describe gains within each class; they are not direct tests of differences between classes.

4 Discussion

4.1 Recurring local departures from the IRI background

The recurring foF2 departure at Hainan can be reduced using observations from earlier years. Ratovsky et al. (2014) described local daily and seasonal patterns in peak density and height, and Wang et al. (2017) found that IRI-2012 generally underestimated Hainan foF2. The positive bias (observed minus modeled) in our IRI-2026 comparison has the same direction. The present calibration results show how much error can be removed in later observations and whether the tested forest improves on the month-hour correction.

For frequency, a simple correction by month and hour performs at least as well as the tested forest in terms of MAE (0.997 versus 1.013 MHz). Earlier Hainan studies identified strong daily and seasonal organization, so a correction using these two coordinates has a clear physical setting (Ratovsky et al., 2014; Wang et al., 2017). The forest has no MAE advantage for frequency in this comparison. Atıcı and Pala (2026) also found performance that varied with station and activity in daily foF2 estimation, including conditions in which IRI outperformed the tested machine learning models. For the present Hainan frequency reference, the simple correction is therefore a useful starting point.

For height, the forest lowers MAE by a further 1.187 km relative to the month-hour correction. This modest difference occurs in a parameter whose uncorrected estimate has a substantial mean offset. Incoherent scatter radar (ISR) and ionosonde comparisons at Millstone Hill showed that the error in IRI height estimates depends on the selected height formula, season and reference observations (Mengist et al., 2020). Their overall preference for the Shubin option does not mean that it removes local errors everywhere. At Hainan, historical calibration still helps this selected background. Wang et al. (2017) used measured M(3000)F2 in their height comparison, whereas our reference uses the Shubin-COSMIC option. The studies therefore address different forms of observational constraint, rather than measuring a change in the accuracy of the height model under identical conditions.

4.2 Recent observations and departures from the usual local state

The update using the preceding residual was much more accurate than either historical correction. This is expected when adjacent measurements share information about the local state that a long-term average cannot supply. Galkin et al. (2012) built this idea into an IRI mapping framework using recent ionosonde data. In a later IRTAM evaluation, the overall ionosonde comparison improved more for foF2 than for



hmF2, while comparisons with COSMIC satellite radio occultation measurements did not show the same
280 overall benefit (Pignalberi et al., 2021). At Hainan, the immediately preceding residual is useful for
describing both peak parameters on the same available samples.

The kind of observation used for an update helps explain why height results differ between studies.
Pignalberi et al. (2019) updated IRI with GNSS VTEC and found that height estimates worsened in their
independent-year comparison. VTEC measures the whole electron column, so different combinations of
285 density, height and profile shape can produce a similar value. An hmF2 value derived from ionosonde
measurements instead constrains the peak position directly. Their IRI UP comparison also changed the
height formula from Shubin to a formulation based on M(3000)F2. These differences offer concrete
reasons why our update of the same measured peak parameter can improve height even when an update
based on column content does not.

290 A small error over 15 minutes is also compatible with persistence of the ionosphere and of local
measurement offsets. Chan and Cannon (2002) used persistence as a baseline in short-term foF2
forecasting. Our update carries the previous observed value forward and adds the modeled change over the
interval (Sect. 2.3). Because pure observation persistence was not evaluated separately, the present
comparison measures the usefulness of an estimate containing recent information; it does not determine
295 how much the IRI change adds to that previous observation.

4.3 Possible reasons for larger foF2 gains during spread F

The larger percentage foF2 gains during spread F can be understood by examining which part of the error
is removed. In the Freq class, the raw MAE is actually smaller than in noSF (0.907 versus 1.130 MHz), yet
the percentage reduction is larger (21.9% versus 8.2%). The mean bias (observed minus IRI) falls from
300 0.672 to 0.050 MHz in Freq, compared with 0.348 to 0.110 MHz in noSF. Records with frequency
spreading therefore contain a stronger net underestimation relative to their average absolute error.
Removing much of that offset offers a plausible explanation for their larger relative gain.

The timing of spread F provides a second part of this explanation. Wang et al. (2018) found different
nighttime and seasonal occurrence patterns for frequency and range spreading, with frequency spreading
305 more frequent at lower foF2. Hainan studies also describe pronounced seasonal changes and changes from
evening to night in the peak parameters, plausibly linked to low-latitude plasma drifts (Ratovsky et al.,
2014; Wang et al., 2017). SF records can therefore occur preferentially when the global background has a
repeatable local error. A forest using time and IRI predictors can correct part of that error without receiving
an SF label. Low measured foF2 and a positive residual (observed minus IRI) can coexist if IRI
310 underestimates the local density still further.

Background correction still leaves substantial errors in the more strongly spread records. Wang et al.
(2015) observed strong range spreading with plasma bubbles and scintillation over Hainan. In our Range
and SRRange records, corrected foF2 MAEs remain 2.453 and 3.622 MHz. Irregular plasma structure and



changes that are not captured by the historical relationship may both contribute. Thus, removing a
315 recurring offset can yield a large percentage improvement while leaving a poor estimate of an individual
observation. The class comparisons combine their actual seasonal and local-time distributions, so this
interpretation concerns the sampled conditions rather than an isolated effect of SF.

4.4 Why peak height responds differently

For height, the largest estimated correction gain occurs without spread F: 26.5% in noSF, compared with
320 10.3% in Freq (Fig. 2e). The noSF bias changes from -19.185 to -2.179 km, showing that the correction
removes much of the mean overestimate. In Freq, the raw bias is already near zero (-0.193 km), even
though MAE is 25.587 km. Positive and negative errors cancel in the mean but remain large in individual
records. This difference in error structure helps explain the larger noSF gain: reducing a persistent offset is
useful there, whereas records with frequency spreading require a better description of variations around the
325 mean.

Density and height also respond to different aspects of the profile. Ratovsky et al. (2014) found different
daily patterns and responses to solar activity for Hainan peak density and height. Plasma transport can
move the layer and alter its density distribution, so the density maximum and its altitude need not change
together. Wang et al. (2018) related range spreading to bottomside virtual height $h'F$, a delay-based
330 quantity distinct from h_mF2 . These earlier results give a physical basis for evaluating frequency and height
separately. The uncertain SRange height gain, with an interval of -12.4% to 29.1% , leaves the benefit of
the tested correction unresolved; it does not indicate an absence of height changes during strong spreading.

4.5 Implications for ionospheric studies and monitoring in Hainan

Table 2 shows that MAE and mean bias can favor different choices. The month-hour correction has
335 slightly lower foF2 MAE, but its mean bias is -0.244 MHz, compared with 0.153 MHz for the forest. For
 h_mF2 , the forest reduces the remaining mean overestimate from 10.193 to 1.598 km while reducing MAE
by a further 1.187 km. Method choice therefore also depends on whether the application prioritizes average
absolute error or systematic offset.

For regional research, the benefit of the month-hour correction shows that part of the departure from global
340 IRI follows a recurring local pattern. Accounting for this pattern provides a clearer reference for examining
individual events. Frequency and height should be assessed separately because their systematic errors and
their responses to correction differ. Residual departures can then be studied in relation to the event and its
background conditions.

When recent measurements are available, they give a much closer estimate of both parameters in this
345 record. Even so, relative improvement alone is an incomplete guide to performance: the corrected foF2
MAE during strong range spreading is 3.622 MHz, compared with 1.037 MHz in noSF. Applications



involving strongly spread echoes must account for this larger remaining error. Translating these peak-parameter errors into radio link performance requires a propagation assessment.

The supporting VTEC comparison describes a different part of the problem. Improving the density peak
350 does not ensure that the rest of the vertical profile is equally accurate. A comparison with preliminary Swarm measurements found differences between observed and modeled topside density trends under quiet and disturbed conditions (Pignalberi et al., 2016). GNSS VTEC also includes electrons above the 1000-km limit used for IRI integration here. The difference between the two columns therefore includes information about the upper profile as well as the unequal integration heights (Bilitza et al., 2022; Pignalberi et al.,
355 2025).

These findings apply to the available Hainan Fuke record and the years examined. SF groups reflect labels assigned by the classifier, available peak measurements and their seasonal and local-time sampling. The recent-residual comparison is retrospective, including cases where manual tracing used later ionograms. Operational use would require the preceding measurement to arrive in time. Historical correction and
360 recent updating were assessed separately; their combination was not evaluated.

5 Conclusions

At Hainan Fuke, historical calibration reduces IRI-2026 errors in the peak parameters, with different benefits from the two correction schemes. In 2020, the simple month-hour correction gives a slightly smaller foF2 MAE than the forest, whereas the forest lowers hmF2 MAE by a further 1.187 km. The forest
365 improves on uncorrected IRI by 10.9% for foF2 and 24.0% for hmF2, with the same direction of improvement in 2019. These comparisons support choosing a local correction for each parameter and judging both its MAE and its remaining mean bias.

The remaining error also depends on the recorded SF state. Estimated foF2 percentage gains are larger in every SF class than in noSF, but strong range spreading retains an MAE of 3.622 MHz, compared with
370 1.037 MHz in noSF. Height has its largest estimated gain in noSF, while the SRange benefit remains uncertain. Regional use therefore requires both the relative gain and the remaining error for each parameter and state. The separate update using the preceding residual gives MAEs of 0.256 MHz and 10.851 km, supporting the value of continuing observations for short-term local estimates. Its operational forecast horizon remains to be established.

375 Appendix A Processing details

The ionosonde processing flags foF2 outside 0.5-25 MHz, hmF2 outside 100-700 km, and ionosonde TEC outside 0-200 TECU. These annotations do not exclude finite values from the reported analysis. The stored



five-class label is the mode within the bin, with ties resolved by the first label in sorted order; the separate binary field marks any SF in the bin and does not replace the modal class in stratified comparisons.

380 The GNSS comparison uses the interval $[t-7.5 \text{ min}, t+7.5 \text{ min}]$. Values are first aggregated across available satellites at each epoch and then across epochs, using the median at both stages. This centered window can include observations after the target bin start, so it is used for the retrospective VTEC comparison.

For the IRI-2026 calculation, switches JF(4), JF(5), JF(6), JF(22), JF(23), JF(30), JF(33), JF(34), JF(35), JF(39), JF(40), and JF(47) were false; the other switches through JF(50) were true. The solar and
385 geomagnetic index files were identical to the historical calculation. The four IRI predictor quantities include both the reconstructed quiet foF2 and the returned storm-adjusted NmF2; when the storm multiplier differs from unity, their relationship also depends on that multiplier.

Random forest models used bootstrap samples of training rows, all predictors available at each split ($\text{max_features}=1.0$), minimum split size 2 and two computational workers. The uncertainty calculation
390 resamples complete dates rather than treating every quarter-hour observation as independent.

Appendix B Baseline and spread-F details

Figures B1 and B2 show errors in uncorrected IRI by recorded morphology and time period. These provide context for the correction results in Sect. 3. The results below include values for the full period, corrections for each class, sample counts and intervals obtained by resampling complete dates for each target.

395 Across all available foF2 pairs, the bias (observed minus IRI) was 0.651 MHz, MAE was 1.254 MHz, and RMSE was 1.675 MHz; R^2 for the observed parameters was 0.811 ($N=222,206$; 2442 dates).

Across all available hmF2 pairs, the bias (observed minus IRI) was -7.257 km, MAE was 27.747 km, and RMSE was 36.138 km; R^2 for the observed parameters was 0.392 ($N=217,686$; 2442 dates).

Across all available VTEC pairs, the bias (observed minus IRI) was 3.085 TECU, MAE was 4.614 TECU,
400 and RMSE was 6.565 TECU; R^2 for the observed parameters was 0.771 ($N=157,915$; 1697 dates).

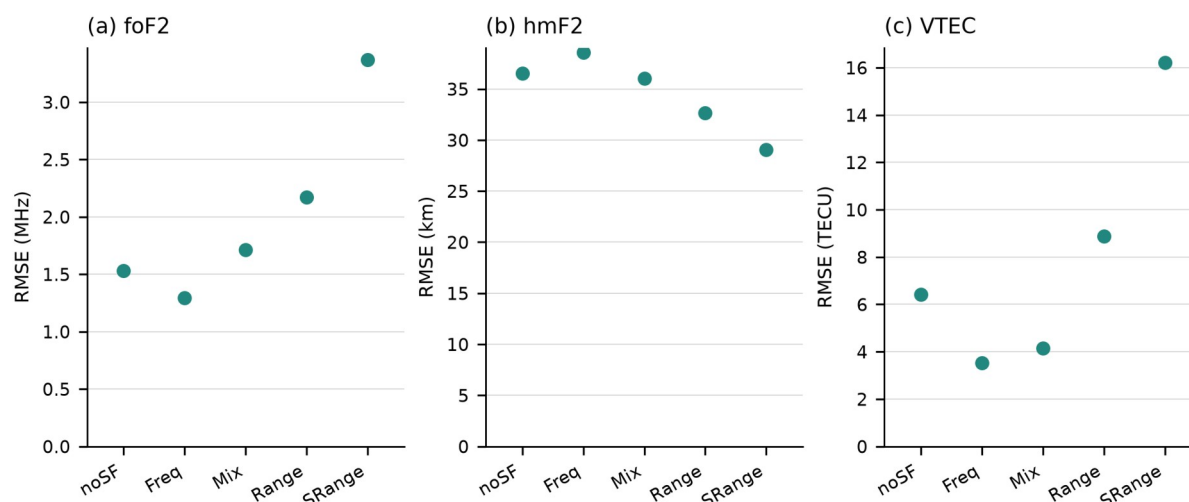


Figure B1. IRI-2026 baseline RMSE across the five recorded SF states, evaluated over all available years. Sample sizes for each parameter and class are given in the published aggregate data (Wang and Cai, 2026). Points are descriptive estimates.

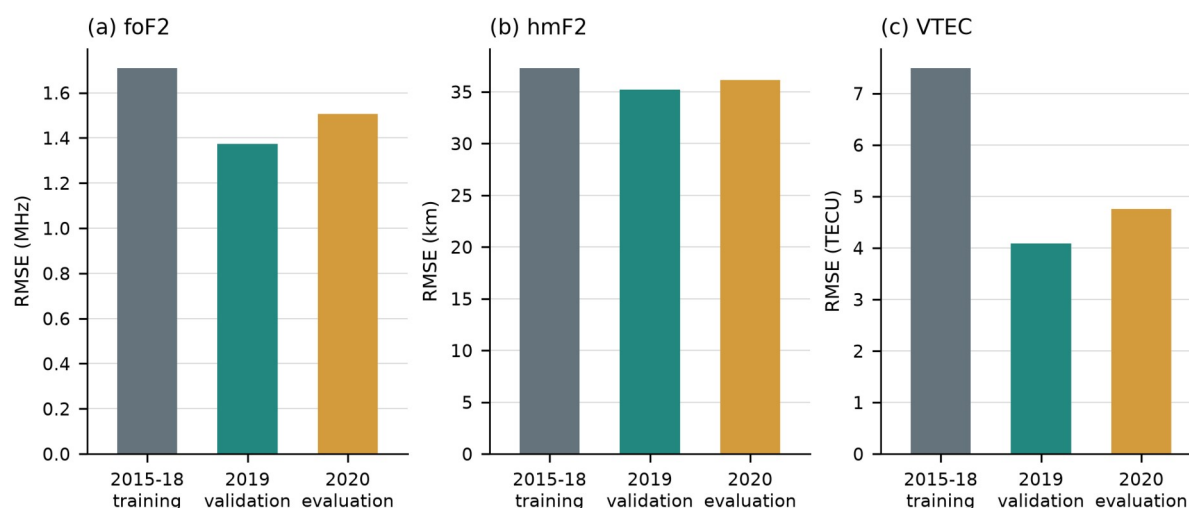


Figure B2. IRI-2026 baseline RMSE in the training, validation and evaluation periods. Each target retains its physical unit and available observations. The partitions contain different observations and do not represent paired temporal changes.

The foF2 correction differed across the recorded spread-F strata: Freq: 0.907 to 0.709 MHz (21.9% reduction; 95% interval 14.4 to 28.5%; N=3,232; 295 dates); Mix: 1.398 to 1.009 MHz (27.8% reduction; 95% interval 23.7 to 32.4%; N=1,718; 244 dates); Range: 3.159 to 2.453 MHz (22.4% reduction; 95% interval 16.4 to 29.2%; N=107; 58 dates); SRange: 4.392 to 3.622 MHz (17.5% reduction; 95% interval 14.6 to 21.1%; N=76; 21 dates); noSF: 1.130 to 1.037 MHz (8.2% reduction; 95% interval 5.8 to 10.6%; N=25,870; 355 dates).



The hmF2 correction differed across the recorded spread-F strata: Freq: 25.587 to 22.945 km (10.3% reduction; 95% interval 6.9 to 13.7%; N=3,096; 293 dates); Mix: 24.665 to 22.942 km (7.0% reduction; 95% interval 2.4 to 11.4%; N=1,627; 239 dates); Range: 25.081 to 21.035 km (16.1% reduction; 95% interval 0.5 to 30.0%; N=106; 58 dates); SRange: 23.125 to 21.284 km (8.0% reduction; 95% interval -12.4 to 29.1%; N=70; 20 dates); noSF: 29.305 to 21.531 km (26.5% reduction; 95% interval 24.3 to 28.8%; N=24,787; 355 dates).

Appendix C Comparator details

Figure C1 compares absolute MAE and RMSE for the methods in Table 2. Table C1 gives the corresponding differences between the random forest and each comparator.

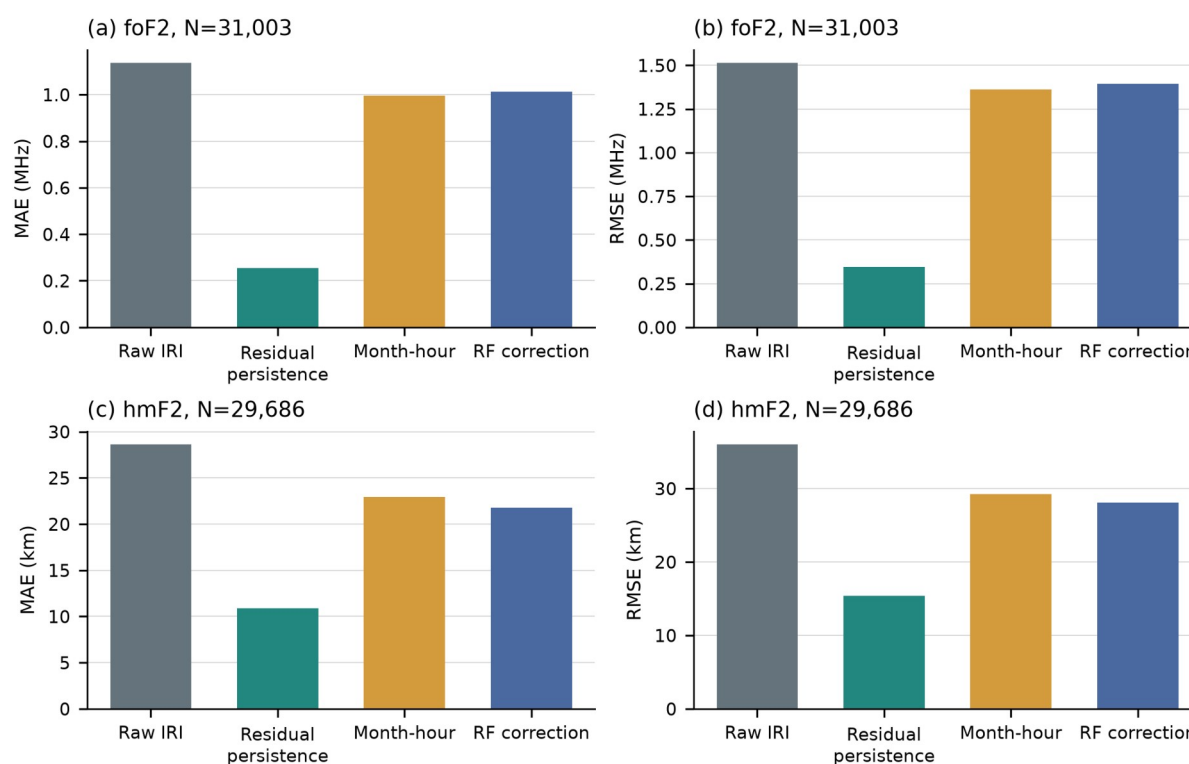


Figure C1. Errors on the same samples for uncorrected IRI-2026, persistence of the exact preceding residual, a month-hour correction based only on the training years, and a random forest correction. For each target, all methods use the same samples with finite target values and an available residual from the exact preceding bin. Panels show MAE and RMSE in physical units; R^2 in residual space is reported separately in Table 2. Values are descriptive point estimates.



Table C1. Random forest minus comparator on the same samples as Table 2. Negative MAE and RMSE differences favor random forest. Error units are MHz for foF2 and km for hmF2; R^2 differences are dimensionless. These pairwise differences are descriptive.

Target	Comparator	MAE difference	RMSE difference	ΔR^2 residual space	in
foF2	Raw IRI	-0.124	-0.121	0.169	
foF2	Residual persistence	0.757	1.046	-0.872	
foF2	Month-hour	0.016	0.030	-0.040	
hmF2	Raw IRI	-6.880	-7.958	0.497	
hmF2	Residual persistence	10.902	12.693	-0.538	
hmF2	Month-hour	-1.187	-1.209	0.068	

430 Code availability

The IRI-2026 Fortran distribution is available from <https://irimodel.org/IRI-2026/> (IRI Working Group, 2026). Code for reproducing the figures from the published aggregate values is available in ScienceDB, version V1, at <https://doi.org/10.57760/sciencedb.013ff> (Wang and Cai, 2026), under the MIT license. The processing, model fitting and evaluation code is not included in this public release.

435 Data availability

Source observations are available through the Chinese Meridian Project portal at <https://dcstatus.meridianproject.ac.cn/>, subject to registration and use conditions (Chinese Meridian Project, n.d.-a, n.d.-b, n.d.-c, n.d.-d). Aggregate tables and figure source data are openly available in ScienceDB, version V1, at <https://doi.org/10.57760/sciencedb.013ff> (Wang and Cai, 2026), under CC BY
440 4.0. The dataset supports inspection of the reported statistics and reproduction of the figures. It excludes raw observations, individual manual traces, peak parameter records and SF labels. The manually traced records used in this study are held by the research team and are not publicly available. The authors have limited the release to aggregate results to retain the individually curated records as a research resource; this is an author choice rather than an institutional or legal restriction.

445 Author contributions

Zheng Wang: Conceptualization, Methodology, Data curation and Validation (manual ionogram tracing and calibration), Formal analysis, Writing - original draft. Ziyue Wang: Formal analysis, Visualization. Jinhui Cai: Methodology, Formal analysis, Writing - review and editing. Guojun Wang: Conceptualization,



Data curation and Validation (manual ionogram tracing and calibration), Writing - review and editing. Chu
450 Qiu: Methodology, Formal analysis, Writing - review and editing. Jiankui Shi: Resources (instrument
management), Writing - review and editing. Dabin Xue: Writing - review and editing. Pengdong Gao:
Methodology, Formal analysis, Writing - review and editing. Xiao Wang: Resources (instrument
management), Formal analysis.

Competing interests

455 The authors declare that they have no conflict of interest.

Acknowledgements

We acknowledge the use of Ionogram Graph of Ionospheric Digital IonoSonde at Danzhou Station
(DOI:10.12176/01.06.00507-V01), Ionospheric Parameters of Ionospheric Digital IonoSonde at Danzhou
Station (DOI:10.12176/01.06.00508-V01), RINEX Format Parameters of GNSS Ionospheric TEC and
460 Scintillation Monitor at Danzhou Station (DOI:10.12176/01.06.00578-V01), Ionospheric Vertical Absolute
TEC of GNSS Ionospheric TEC and Scintillation Monitor at Danzhou Station
(DOI:10.12176/01.06.00588-V01) from the Chinese Meridian Project. The data can be accessed via
<https://dcstatus.meridianproject.ac.cn/>. We thank the staff of the Chinese Meridian Project
(<https://cstr.cn/31122.02.MSP>) for their technical support and assistance in data collection and analysis.

465 OpenAI ChatGPT and Codex assisted with language editing, manuscript restructuring and code-assisted
consistency checks during manuscript preparation and revision.

Financial support

This research was supported by the Youth Innovation Promotion Association CAS (2023000116), the
National Key R&D Program of China (2025YFF0510703), the Strategic Priority Research Program of the
470 Chinese Academy of Sciences (XDA0470301), and the Specialized Research Fund for the State Key
Laboratory of Solar Activity and Space Weather. The authors also acknowledge operational funding
support from the Chinese Meridian Project.

References

Atıcı, R. and Pala, Z.: Machine Learning Models for Ionospheric foF2 Prediction: A Comparative
475 Analysis, Radio Science, 61, e2025RS008336, <https://doi.org/10.1029/2025rs008336>, 2026.



- Beggan, C. D., Kloss, C., Amblard, P., Amezza, T., Amit, H., Aubert, J., Baerenzung, J., Bregnhøj Lauridsen, J., Brown, W., Califf, S., Campuzano, S. A., Chauvet, L., Chulliat, A., Claveau, R., Coisson, P., da Silva, M., Davies, C., Deborde, R., Dobrovolsky, M., Fillion, M., Finlay, C. C., Firsov, I., Fournier, A., Gailler, L., Gillet, N., Gomez Perez, N., Grayver, A., Gwartz, K., Hamoudi, M., Holme, R., Holschneider, 480 M., Hulot, G., Jager, T., Jiang, Y., Kada, M., Korte, M., Krasnoperov, R., Kuang, W., Kudin, D., Léger, J. M., Lemgharb, A., Lesur, V., Liu, P., Livermore, P. W., Madsen, F. D., Manda, M., Matsushima, M., Matzka, J., Michaelis, I., Minami, T., Nair, M., Nakano, S., Oehler, J.-F., Olsen, N., Pavón-Carrasco, F. J., Presnyakov, S., Qian, B., Rang, X., Rogers, H. F., Rother, M., Rouxel, D., Sabaka, T., Sato, S., Saynisch-Wagner, J., Schanner, M., Serrano, M., Shakespeare-Rees, N., Shi, Y., Smith, A., Soloviev, A., Terra- 485 Nova, F., Thebault, E., Tøffner-Clausen, L., Toh, H., Wang, J., Wardinski, I., Watson, C., Whaler, K., Xiong, C., Yan, Q., Yang, Y., Yao, H., Zeren, Z., Zhang, K., and Zhou, B.: International geomagnetic reference field: the fourteenth generation, *Earth, Planets and Space*, 78, 127, <https://doi.org/10.1186/s40623-025-02360-0>, 2026.
- Bilitza, D., Altadill, D., Zhang, Y., Mertens, C., Truhlik, V., Richards, P., McKinnell, L.-A., and Reinisch, 490 B.: The International Reference Ionosphere 2012 - a model of international collaboration, *Journal of Space Weather and Space Climate*, 4, A07, <https://doi.org/10.1051/swsc/2014004>, 2014.
- Bilitza, D., Pezzopane, M., Truhlik, V., Altadill, D., Reinisch, B. W., and Pignalberi, A.: The International Reference Ionosphere Model: A Review and Description of an Ionospheric Benchmark, *Reviews of Geophysics*, 60, e2022RG000792, <https://doi.org/10.1029/2022rg000792>, 2022.
- 495 Chan, A. H. Y. and Cannon, P. S.: Nonlinear forecasts of foF2: variation of model predictive accuracy over time, *Annales Geophysicae*, 20, 1031-1038, <https://doi.org/10.5194/angeo-20-1031-2002>, 2002.
- Chinese Meridian Project: Ionogram Graph of Ionospheric Digital IonoSonde at Danzhou Station [data set], <https://doi.org/10.12176/01.06.00507-V01>, n.d.-a.
- Chinese Meridian Project: Ionospheric Parameters of Ionospheric Digital IonoSonde at Danzhou Station 500 [data set], <https://doi.org/10.12176/01.06.00508-V01>, n.d.-b.
- Chinese Meridian Project: Ionospheric Vertical Absolute TEC of GNSS Ionospheric TEC and Scintillation Monitor at Danzhou Station [data set], <https://doi.org/10.12176/01.06.00588-V01>, n.d.-c.
- Chinese Meridian Project: RINEX Format Parameters of GNSS Ionospheric TEC and Scintillation Monitor at Danzhou Station [data set], <https://doi.org/10.12176/01.06.00578-V01>, n.d.-d.
- 505 Galkin, I. A., Reinisch, B. W., Huang, X., and Bilitza, D.: Assimilation of GIRO data into a real-time IRI, *Radio Science*, 47, RS0L07, <https://doi.org/10.1029/2011RS004952>, 2012.
- IRI Working Group: International Reference Ionosphere IRI-2026 Fortran software distribution [software], <https://irimodel.org/IRI-2026/> (last access: 8 September 2026), 2026.



- Mengist, C. K., Yadav, S., Kotulak, K., Bahar, A., Zhang, S.-R., and Seo, K.-H.: Validation of
510 International Reference Ionosphere model (IRI-2016) for F-region peak electron density height (hmF2):
Comparison with Incoherent Scatter Radar (ISR) and ionosonde measurements at Millstone Hill, *Advances
in Space Research*, 65, 2773-2781, <https://doi.org/10.1016/j.asr.2020.03.017>, 2020.
- Miró Amarante, G., Cueto Santamaría, M., Alazo, K., and Radicella, S. M.: Validation of the STORM
model used in IRI with ionosonde data, *Advances in Space Research*, 39, 681-686,
515 <https://doi.org/10.1016/j.asr.2007.01.072>, 2007.
- Pignalberi, A., Pezzopane, M., Tozzi, R., De Michelis, P., and Coco, I.: Comparison between IRI and
preliminary Swarm Langmuir probe measurements during the St. Patrick storm period, *Earth, Planets and
Space*, 68, 93, <https://doi.org/10.1186/s40623-016-0466-5>, 2016.
- Pignalberi, A., Pezzopane, M., Rizzi, R., and Galkin, I.: Effective Solar Indices for Ionospheric Modeling:
520 A Review and a Proposal for a Real-Time Regional IRI, *Surveys in Geophysics*, 39, 125-167,
<https://doi.org/10.1007/s10712-017-9438-y>, 2018.
- Pignalberi, A., Habarulema, J. B., Pezzopane, M., and Rizzi, R.: On the Development of a Method for
Updating an Empirical Climatological Ionospheric Model by Means of Assimilated vTEC Measurements
From a GNSS Receiver Network, *Space Weather*, 17, 1131-1164, <https://doi.org/10.1029/2019sw002185>,
525 2019.
- Pignalberi, A., Pietrella, M., and Pezzopane, M.: Towards a Real-Time Description of the Ionosphere: A
Comparison between International Reference Ionosphere (IRI) and IRI Real-Time Assimilative Mapping
(IRTAM) Models, *Atmosphere*, 12, 1003, <https://doi.org/10.3390/atmos12081003>, 2021.
- Pignalberi, A., Azpilicueta, F., Nava, B., Pezzopane, M., and Bilitza, D.: Validating the International
530 Reference Ionosphere model through altimeter-satellite vTEC observations: implications for the topside
ionosphere modeling, *Journal of Geodesy*, 99, 81, <https://doi.org/10.1007/s00190-025-02003-y>, 2025.
- Ratovsky, K. G., Shi, J. K., Oinats, A. V., and Romanova, E. B.: Comparison of Diurnal, Seasonal and
Solar Cycle Variations of High-latitude, Mid-latitude and Low-latitude Ionosphere, *Chinese Journal of
Space Science*, 34, 143-153, <https://doi.org/10.11728/cjss2014.02.143>, 2014.
- 535 UMLCAR: SAO Explorer [software], <https://ulcar.uml.edu/SAO-X/SAO-X.html> (last access: 19
September 2026), n.d.
- Wang, G. J., Shi, J. K., Reinisch, B. W., Wang, X., and Wang, Z.: Ionospheric plasma bubbles observed
concurrently by multi-instruments over low-latitude station Hainan, *Journal of Geophysical Research:
Space Physics*, 120, 2288-2298, <https://doi.org/10.1002/2014ja020245>, 2015.
- 540 Wang, G. J., Shi, J. K., Wang, Z., Wang, X., Romanova, E., Ratovsky, K., and Polekh, N. M.: Solar cycle
variation of ionospheric parameters over the low latitude station Hainan, China, during 2002-2012 and its



comparison with IRI-2012 model, *Advances in Space Research*, 60, 381-395,
<https://doi.org/10.1016/j.asr.2016.12.013>, 2017.

Wang, N., Guo, L., Zhao, Z., Ding, Z., and Lin, L.: Spread-F occurrences and relationships with foF2 and
545 h'F at low- and mid-latitudes in China, *Earth, Planets and Space*, 70, 59, <https://doi.org/10.1186/s40623-018-0821-9>, 2018.

Wang, Z. and Cai, J.: Aggregated source data for local departures from IRI-2026 and observation-based
correction at Hainan Fuke, V1, Science Data Bank [data set], <https://doi.org/10.57760/sciencedb.013ff>,
2026.

550 Wang, Z., Zhan, M., Gao, P., Wang, G., Qiu, C., Qi, Q., Shi, J., and Wang, X.: Automatic Detection and
Classification of Spread-F From Ionosonde at Hainan With Image-Based Deep Learning Method, *Space
Weather*, 21, e2023SW003498, <https://doi.org/10.1029/2023sw003498>, 2023.

Zhu, P., Xie, C., Jiang, C., Yang, G., Liu, J., Li, Z., and Zhao, Z.: Ionospheric Behavior of foF2 over
Chinese EIA Region and Its Comparison with IRI-2016, *Universe*, 6, 122,
555 <https://doi.org/10.3390/universe6080122>, 2020.