Review of Scott+ 2023 "Calibrating estimates of ionospheric long-term change" (egusphere-2023-2599)

Contents

General comments 1

Specific comments 2
- Attribution of effect to F1: tighten up composition link . . . 2
- Attribution of effect to F1: exclude E as a confounder . . . 2
- Enhance existing plots . . . 3
- Create supplementary material plots . . . 4
- Better justify attribution of scatter to foE floor at 0.4 MHz . . . 5
- Check/discuss error bars in Figs 2, 7 . . . 5
- Check unexpected striations in time series scatter plots . . . 6
- Flag up more clearly where your findings differ from others' . . . 6
- Minor points . . . 7

Technical corrections 7
- Improve equations: typesetting, M(3000)F2 consistency . . . 7
- Make various things more explicit . . . 8
- Possible typos . . . 16

General comments

Scott et al. consider the derivation of ionospheric F2 layer heights from historic scaled ionosonde parameters. These parameters have been recorded since the early 20th century, and in principle provide a very valuable long-term record in which to seek an expected long-term ~20 km decrease in this F2 layer arising from climate change. Similar expected decreases are harder to discern in other upper atmosphere layers, where records are also shorter. Consequently, many previous authors have examined this ionospheric data, trying to identify a secular long-term trend, but this has turned out to be difficult, with inconsistent, countervailing trends emerging from different ionospheric observations worldwide.

Scott et al. provide a valuable synthesis of the state of the art, and make a compelling, novel, and lucidly-written case that previous analyses have relied on semi-empirical equations which incorrectly assume any F1 layer present has an insignificant effect. Taking ionosonde data through such an equation, and comparing the resulting imputed F2 layer heights with more reliable estimates from an incoherent scatter radar, they show F1 layers can cause biases in height estimates a little larger than the climate trend being sought, on various timescales. Hence risking masking any climate trends, or causing misattributions. They attribute this F1 effect to thermospheric composition effects, driven by geomagnetic changes affecting energy deposition into the thermosphere, and resulting impacts on thermospheric heating and circulation.

The quality of the analysis and manuscript is generally very good, and nearly ready for publication. However to ensure the work reaches its full potential, becoming an extremely valuable reference for such efforts, there are a few areas which need addressing. Some where the authors seem to have gone a little fast in their analysis, and where their case needs to be evidenced a little better, to ensure their conclusions are as robust as possible. Or where the authors can make their reasoning more explicit. The remarks below are quite fussy, holding the authors to a very high standard. They are invited to push back as appropriate!

Specific comments

Note B&D abbreviation used throughout = Bradley & Dudeney.

Attribution of effect to F1: tighten up composition link

You've made a fairly compelling case that F1 can contaminate hmF2 values:

Pretty strong: evidence

- Fig 4, showing clear break in ISR-ionosonde hmF2 bias, ordered by xF = foF2/foF1
- Fig 5, showing you can use the trend from smaller xF in Fig 4 to successfully remove the summer daytime hmF2 bias
- Fig 8a, showing (modelled) hmF2 bias has a trend and timescales which could lead to confounding with climate signal

Slightly weaker however: foF1 <-> composition link

- Physical rationale (introduced ~L364, invoked ~L405) linking presence of foF1 to composition changes (may be a priori knowledge I lack - make this more explicit if so!)
- Qualitative support from composition proxy in 8b showing similar trend to 8a Am not familiar with this Wright & Conkright proxy, and I note comment CC1 queries it…
- Discussion of annual/semi-annual circulation-induced changes to composition in L430- L447 seems to suggest foF2 is more affected than foF1: L441 "foF2 is suppressed… when foF1 is at its peak"

So it would be good if the foF1 <-> composition link rationale could be tightened up a little

Attribution of effect to F1: exclude E as a confounder

While the F1 evidence *seems* good, I'm a bit concerned that the apparent "smoking gun" in the derived parameters means you may not have looked at the raw data, to see if there are other potential causes for the patterns you've discerned.

Apologies - perfectly possible you've done this already, discovered it's not a concern, and decided not to clutter paper with this. If this has happened though, I'd recommend you say so explicitly, along lines of "An examination of ABC (not shown) excludes XYZ"

Specific areas of concern:

- the domain of validity for B&D model is for xE (= foF2/foE) > 1.7
- both xF and xE share a common numerator, so I'd expect them to be somewhat correlated

Data-wise:

- What happens if you create an equivalent to Fig 4, but order the hmF2 bias by xE?
  - Is there a similar "break" partway along?
  - Shouldn't be, if the B&D model for hmF2(xE) is decent along its validity range…
  - but worth double-checking, in case of any nasty non-linearities?
- Likewise does any of your data have xE <= 1.7 (i.e. outside of B&D validity range)?
  - What happens to bias then?

Physically:

- Might composition changes affecting foF1 also affect foE?

I think it's worth being certain that the attribution to foF1 contamination is rock-solid.

Not least as a foE effect might be good news: given foE scaled more frequently than foF1, if any effect can be characterised, *might* give more data for looking at trends globally using already-manually scaled parameters, without data-rescue-based inversions?

See my suggestion below on a few supplementary material plots for probing foF1 attribution.

- there may be other ways of excluding potential confounders of foF1 attribution

- not least simply telling me if I'm wide of mark here!

Enhance existing plots

There are various places where the existing plots could be enhanced, to reveal details which are currently hard to see, or which aren't fully probed.

A few suggestions below. From your code, I think all should be ~easy to add.

Figure 1: flag 180 km limit, plot delta_hmF2

On ISR subpanel, worth adding a horizontal line / note in caption that there's a 180 km threshold getting imposed to avoid sporadic E layer?

- This to help readers spot this in subsequent plots
- And consider this when ISR is acting as source of truth

Add a 3rd subpanel below current two, plotting `delta_hmF2 = hmF2_ISR - hmF2_ionosonde`

This will help:

- act as a hook for you to define delta_hmf2 in text (not currently done - see pedantry below on why useful)
- make the bias (higher ionosonde hmF2) clearer
  - I can see this by concentrating hard, but much easier to spot on a difference plot
- allow upfront spotting of any temporal patterns in raw delta_hmf2 (later plots are processed)
- help readers reason about intercept in Fig 2, 7
- act as hook for decimated delta_hmF2 in Fig 4, modelled delta_hmF2 in 8a

Pedantic quibble - no need to change sense, just why IMO useful to define delta_hmF2:

- above seems to be the definition for delta_hmF2 you use in 4, 8a
- but in this paper, you treat hmF2 from ISR as the "truth"/reference value
- so definition is ~backward cf the usual definition of change = value - reference
- a little counterintuitive, so adds to why worth explicitly defining delta_hmF2 in text

*Note if you **did** change sense, delta_hmF2 values would be biased to positive values, so making the sign consistent with the positive offsets in Figs 2, 7*

Figure 2: break out the day/night/terminator populations

Recommend changing this to a 2x2 subpanel plot

- keep current plot as an "all populations" subpanel
- but use other 3 subpanels to plot each day/night/terminator population separately

I say this as you've gone to some trouble to stratify the three populations, and I think it's worth getting all the benefit out!

Current plot is good for general intercomparison of the three populations, and reasoning about the "all" fit. But overplotting means it's hard to see & assess any detailed features in each population.

E.g. there seem to be ~two distinct clumps in the nighttime population, above/below the main trendline, at low/high ISR hmF2. But as everything is overplotted, it's hard to assess if there's anything similar in the twilight / day populations, e.g.

- do they have similar clumps?
- if so, do their "upper" clumps also contain ISR hmF2s which hit the 180 km floor?

Likewise hard to see scatter cores - is anything interesting lurking in there?

Overall makes it harder to assess validity of the linear fits. These look OK, but good to reassure reader before these get used (Fig 5, 8a)

Figure 3: plot the delta_hmF2; bin count / 180 km count

These plots are great for probing the seasonal and local time behaviour, but I think they could be made even more revealing.

Again I'd recommend inflating this to a 2x2 plot. Keep the current two, but add:

- 3: subpanel of delta_hmF2
- 4: subpanel of number of year bins [0 - 35] which have contributed to the ISR mean value in current lower panel

3: to help readers spot the differences, and quantify these. Very hard to compare colours by eye! Might also reveal more subtle differences.

4: to check there's no concerns about quality of ISR averages in any bins I *hope* that the 48% occupancy you report for the 35 years means there's ~16-17 counts per ISR month-localtime bin. But I think worth exposing bin count to be sure:

- as you're treating ISR as ~truth-ier than ionosonde
- as ISR is a campaign-based instrument, and campaigns might clump in season-localtime space!
- as this ISR climatology seems to be showing a few interesting features (e.g. dips in summertime night hmF2) - worth checking these are robust!
  - are these summertime night dips likely to be real?
  - if not, could they be contributing to bias?

alternative 4: fraction of bin with ISR hmF2 = 180 km If the ISR bins all turn out to have OK-ish occupancy, flat enough that you don't think there's an issue, I think you can simply report this in text, and use the subpanel to look at other minor concern instead:

- The 180 km threshold on ISR will skew the bin mean value
- and in worst case (all "true" ISR hmF2 values in bin across years < 180 km) set it to 180

This might be affecting some of the blue-er bins. You could quantify this by calculating what fraction of the counts in each bin have values == 180 km.

Figure 5: plot delta_hmF2

The daytime summer peak has clearly gone, but it's hard to compare colours here to Fig 3b, to get anything more out of this.

Again, recommend adding a subpanel for delta_hmF2 - help readers compare with the previous delta_hmF2, see if there's anything more here.

Will also support your discussion of the nighttime hmF2 differences remaining

- in L365
- in L388 - let you be a bit more nuanced - your "generally lower" point may apply across all months, but doesn't apply in summer months.

Create supplementary material plots

The plots you present are examining the hmF2 or foF2/foF1 ratio. You don't present plots for individual components which go into these derived values.

I've suggested a few supplementary material plots below, to do a deeper dive into the underlying ionosonde data you've retrieved and processed, to:

- ensure that the attribution of hmF2 bias to F1 contamination is correct - check that none of the other parameters (e.g. foE) yield similar behaviour, i.e. act as confounding factors
- check that there aren't any compensating errors occurring

Seems worth extracting all you can out of this - may even give some extra insights!

Looking at your code, I think should be ~easy to generate such plots. Realise this deeper dive is extra work, and on margins of "out of scope for this paper"! Will leave you to judge merit & feasibility!

Specific plots I'm considering you create:

1. For long-term trends: stackplot of hourly monthly median timeseries, like Fig 1, but for the individual foE, foF1, foF2, and M(3000)F2 components. Should show if there's any striking long-term trends in m/any of these individually

2. For seasonal and diurnal patterns: 3x2 subpanels of month-localtime bins, meaned over ionosonde data years. So like 3a, but showing:

    - 2a: foE
    - 2b: foF1
    - 2c: foF2
    - 2d: M(3000)F2
    - 2e: xE i.e. foF2/foE
    - 2f: xF i.e. foF2/foF1

Will help test validity of foF1 attribution, and more, by giving more insights into:

- xF: L479 discusses diurnal & seasonal variation - don't think you've shown this?
    - with this plot, you would!
- xE > 1.7 domain of validity for B&D eqn 4 you're using
    - are there any times & seasons where this is breached
- probing any correlation between xE and xF
    - do both show a summer midday peak, or only xF?
- foE 0.5 MHz assumption for low foE
    - what local times / seasons does this affect most? Is this what you expect?
- M(3000)F2 not accounting for mag field
    - what are values here, how does it compare to Elias?
    - probing conclusions in L404
    - Even do you see Russell-McPheron type variation signatures?
- foF1: help illustrate climatology
    - quantify point in Fig 4 caption "for hours and months where such values exist"
    - similar for L365 "only visible during daylight hours"

Better justify attribution of scatter to foE floor at 0.4 MHz

In discussion of Fig 2 ~L345, you tentatively attribute the scatter to the imposed floor of foE = 0.4 MHz.

Later, in L368 & L389, you strengthen this conclusion - I'm concerned too much. Especially as in L390 you link scatter to ionosonde underestimating (? see typos) hmF2. I assume on basis of offset? Is this scatter though? Or gradient?

As you'll see in technical section, I played with B&D equation in Excel. I set "true" foE in range 0.001 to 0.399, set the M(3000)F2 and foF2 to fixed values, and compared resulting hmF2 to the hmF2 value I got with imposed foE = 0.4 MHz. The hmF2 values different by ~3 to 0 km.

Such small OoM ~1 km sizes, compared to the ~50 km (rough width of "core" scatter) in Fig 2 makes me query if this attribution holds.

But maybe I'm not using appropriate M(3000)F2 and foF2 values, or thinking about this in the right way.

Can you dig into this a little more please?

Check/discuss error bars in Figs 2, 7

Can you check the error bars are implemented correctly? And if correct, briefly discuss relative sizes.

By eye, looks like they're often much larger in vertical cf horizontal direction

- I may be mistaken, as for Fig 2 this is dominated by what I can see in outliers - subpanels would help!
- Looks slightly more equal in Fig 4

If true, seems slightly surprising to me, as your 48% figure above means ionosonde errors are nominally being constrained by considerably more data than your ISR errors. So if errors of equal magnitude, on standard error basis, I'd expect smaller (~70%) error bars for ionosonde.

But in Fig 2, bigger ionosonde error bars suggest ionosonde errors are inherently larger.

- But this doesn't seem to square up with Figure 1, where ISR looks noisier?

I've not looked at your code here, but worth a quick check that error bar axes are consistent with point axes, that things haven't been swapped accidentally?

Check unexpected striations in time series scatter plots

Can you double-check your plotting code in Fig 6 (likely applies to Fig 1 too)? And if all is coded OK, explain how the striations in panel a are arising?

The striations in panel a seem a bit weird - at this zoom level I'd not expect to see these.

I think you're plotting your binned results here, and if I've understood your binning code correctly, bin values for a given month would be would be 24 vertically-aligned points (all local time bins for that month), and a year 12 such vertical lines (all month bins).

So I'd expect to see striations if the plot was on timescales of a few years.

But on the multi-decades timescale, being examined here, you should have 120 lines / decade. Which shouldn't I think give rise to any visible stripes. Whereas I seem to see stripes on ~annual cadence.

I guess striations *could* arise if there's something in orig data / your post-processing decimating F2/F1 bin entries to NaNs preferentially in one half of year - say summer.

But from discussion so far, don't think there is anything like this? (Again the supplementary plots requested might help diagnose anything like this here!)

In absence of something like decimation, I wonder if there's a mistake in plotting code? e.g. month going into a date plotting routine in a day index position?

Wild guess - I've not checked code to see if there's anything like this!

Flag up more clearly where your findings differ from others'

There are a few places throughout intro sections where you flag up ~detailed points from previous literature - often I think as these run counter to what you find.

But you don't refer back to these when you present your results / conclusions.

Upshot is that it's not easy for the reader to spot these - requires multiple passes.

I'd recommend flagging any such disagreements up more explicitly somewhere: either in results as you go, or aggregating somewhere just before/in conclusions.

Ones I've spotted (may well be more!)

- L119: Xu not finding significant geomag effect (cf your Fig 8)
- L215: B&D claiming F1 insignificant. OK, "claim", and it's ~clear that your whole paper is contesting this general point! But if B&D / derivatives get used as much as L246 implies, I think worth being blunt somewhere "B&D's claim is wrong"!
- L487: the juxtaposition is a little too subtle for me - think you could make if clearer that that you're ~disagreeing with Jarvis. How about something like following?

```diff
@@ L488 @@
-explain this. Our results indicate that
+explain this. Our results indicate such mechanisms do not need to be invoked, rather that
```

Minor points

L212: "They noted that `xE > 1.7` is equivalent to about `xF ≈ 1.2`"

- Where are you getting 1.2 from? Looking at B&D I can't see this.

Closest match their sentence below, but I don't think this yields this equivalence. "(In the limit for x = 1.7, f1 = foF2 and we have an entirely linear FI/F2-layer.)"

L164: discussion of routine estimates of hmF2 from digisondes

Likely ~irrelevant/tangential cf the manually-scaled parameters considered in this paper (so not necessarily suggesting you cite it) but in case you missed it, worth sharing Themens+' interesting demonstration that ARTIST's confidence scores for auto-scaled features (incl hmF2) need to be treated carefully.

Possibly of more relevance to here: they also show ARTIST misses F1 layer quite often! Themens+ 2022, ARTIST Ionogram Autoscaling Confidence Scores: Best Practices, https://doi.org/10.46620/22-0001

Technical corrections

Improve equations: typesetting, M(3000)F2 consistency

I found maths in section 2 a little hard to follow due to:

- M(3000)F2 factor changing form a lot
- equations having inline fraction form

Great if you could tighten this up a little.

M(3000)F2

Not your fault this ratio's name has such a convoluted form. *Grr, not a single letter, no subscripts/superscripts, just one long set of inline chars! Guess a fine ionospheric tradition - hmF2 etc - and historic typesetting constraints.*

But this convoluted form doesn't help readers parse these equations, twig it's just a factor. Even harder as the form in this section changes quite a bit:

- Same thing: specific case: M(3000)F2, M3000F2
- Not the same thing: general case: M (looks like eqn 2 is for a general layer)

Realise that for sake of for sake of clarity/historic continuity, in specific case you'll have to use something like M(3000)F2. But there:

- can you use the M(3000)F2 form consistently please - not M3000F2
- great if there's anything you can do with typography to make it clearer this is just a factor!
  - `M_{3000}^{F2}` say?
  - ignore me if this won't wash!

Inline equations

Generally, there are various divisions going on in these equations. With the current inline fraction form it's much harder to follow them.

- for eqns 6 and 7 I ended up downloading Dudeney to check what was going on!

Could you change these to non-inline fractions please? Assuming you're using latex, with `\frac{numerator}{denominator}`

Few other minor things to correct:

- Eqn 6: +/- instead of ±
- Eqn 7:
  - MF and ΔM terms need to be better separated
  - ΔM term has a ˘ symbol - checking Dudeney, this should be a minus sign

Make various things more explicit

There are various places where things were quite implicit, and I had to work ~hard to understand. Might be just me, but am flagging these so you're aware. I think you could usefully make these more explicit, to:

- help convey (what I think are!) your messages
- help less-expert readers follow subtler points (likely ~obvious to you / other experts, but could help this important work get a wider readership / impact)

I've put some suggestions for text which would likely have helped me in most places.

- Very interfering / word-smith-y - naturally please adjust / ignore as needed!

Abstract

Make importance of this work clearer Clarify up-front that effect you've found is of same order as climate signal, so could mask latter / cause misattribution

```diff
@@ L1 @@
-Long-term change in
+Long-term reduction of ~20 km
```

```diff
@@ L9 @@
-±25km an altitude of 250km).
+±25 km at an altitude of 250 km, i.e. similar to the long-term climate signal change being sought).
```

Other: clarify this "shown" part of the point being made comes from previous work, not this one

```diff
@@ L12 @@
-have been
+have previously been
```

Introduction

Effects I lost you a little in the discussion of the effects (L44 - L58). As this is ~crucial for this paper, a few overall ideas on making this section easier for non-experts - some specific ones below.

Overall points:

- Multifaceted nature / spatiotemporal scales of processes: Once I realised importance of this list, I had to draw out diagrams for myself to distinguish the processes, how they work, and outcomes. I don't think a diagram required - but would be helpful to flag up a few things readers should consider when reading list. Specifically outlining at outset sentence that (I think!) effects cover a range of timescales. And that a specific effect can encompass a ~complex set of in/direct processes, which can be local or remote (e.g. advected in).

- Clarity on spatiotemporal scales of effects: Optional, but consider whether here / in a separate paragraph / even a table below, for each effect, it's worth briefly discussing timescales (storm -> solar cycle(s)) and spatial extent - just in auroral zone, or more global due to transport?

- Process impact on height: Have you put these effects in order of ~expected impact on height? Ignore this if you can't / haven't. But if so, indicate explicitly!

- List of effects here: yours / someone else's / common knowledge? Can you be more explicit (e.g. "therefore we expect that changes") if to your knowledge you're the first authors to give this enumeration of effects. Helps readers sense-check "is this everything / is something missing?". Else a citation if someone else has listed these before.
- Citations: If ~easy to do, worth putting in an overall reference / specific references to each effect. Your point 2. got me fishing out Schunk & Nagy, and checking block diagram of energy flow in Ch9!

Specific points:

- effect 1: demarcate better the loss rate change covered above, from the height change introduced here

```
@@ L46 @@
-thermosphere which increases the height
+thermosphere, as described above. This also increases the height
```

- effect 2: be bit more specific on solar irradiance process

```
@@ L48 @@
-to the atmosphere, increasing
+to the upper atmosphere (thermosphere, ionosphere, mesosphere), increasing
```

```
@@ L49 @@
-of the atmosphere
+of the upper atmosphere, and hence raising of pressure levels
```

- effect 3: Make the composition point clearer. This is where I was glad I'd drawn out diagrams for myself! Helped me see that all other processes were changing the *height* of layer features (e.g. hmF2), whereas this process seemed not to. As written, ~implies this is only directly changing the *density* (e.g. nmF2). I *think* this is your intent - rest of article showing that such changes to (say) F1 densities can contaminate hmF2 reconstructions. I've suggested an extra sentence at end of intro to discuss this point, and tease out this difference. But possibly worth flagging this up already here? E.g. via tweak below? Of course, if the composition changes *also* directly affects true feature height (not just density), then alter tweak! In that case, recommend being as explicit as you can.

```
@@ L49 @@
-greater altitudes
+greater altitudes. Profile changes mainly affect densities,
+but this can indirectly affect height estimates.
```

- effect 5: clarify that this effect acts exclusively on long-term timescales?
- L64: might need rewording for clarity 4 and 5? Think only 5 is discussing contraction. And isn't your point in following sentence and section that most authors considered residual to be 5 alone? Whereas only a few - Jarvis, Bremer, Mikhailov - have pointed out that at least 4 can affect long-term trend too?

If so, something like following?

```
@@ L63 @@
-due to the contraction of the thermosphere assumed to be the dominant part of (4) and (5).
+widely assumed to be dominated by the contraction of the thermosphere (5).
```

Composition and circulation, effect 3 Recommend you also consider adding/moving following to this section:

- moving your composition+circulation discussion from ~L431 - L439 to this section
- appending a brief discussion about effect 3

Composition+circulation discussion: I think you could usefully move this to introduction, say just before the effects discussion Highly relevant, and usefully bulks out wind pattern change point, as well as better motivating the seasonal timescale. Currently the intro doesn't give much indication about why results consider seasonal timescale carefully - current timescales are really only ~ diurnal, space weather event, climate. This would help fill out spectrum! Having this here would also let you refer back to this in discussion of Scott et al ~L240.

Brief discussion on 3: I think worth adding something to end of intro:

- while it's very easy to juxtapose with the other effects
- to re/plant seed in readers' minds
- to help introduce following section
- to explicitly tie this in with wider thrust of paper

Something along lines of following - playing around with true/measured as appropriate!

"Absent from most previous literature has been consideration of the ionisation profile changes in (3). This effect may seem irrelevant, as it mainly affects the density of given features in the profile, rather than their true height. However, such density changes can have an indirect effect on estimates of this true height using ionosonde measurements, as discussed in the following section. And hence, as we show in this paper, can contaminate efforts to reconstruct true height estimates to extract the target climate signal from (5)."

Measuring the ionosphere using ground-based radar

Scaling You briefly cover inverting (L84), but don't really cover scaling, despite referring to this later (L164). "Scaling" is quite jargon-y, so may not be accessible to all readers.

Once you got to B&D discussion of foF1, I had to do a quick refresher on ionosondes. I ended up needing to go quite deep into INAG links to understand what the issue was.

As you're (necessarily!) going into the weeds in this paper, possibly worth:

- adding a brief description of scaling process
- adding some citations to guides to both scaling and inverting ionosondes?

I ended up looking at the below for scaling. Might usefully supplement Rishbeth & Garriott (which I didn't check)?

Flagging resources like this helpful:

- for any readers wanting to do a deeper dive here
- esp to any you motivate to run with data-rescue inversion efforts!

~Old guides for scaling linked off https://www.sws.bom.gov.au/IPSHosted/INAG/ :

- Wakai, Ohyama and Koizumi, 1987, Manual of Ionogram Scaling, Radio Research Laboratory, Japan, 3rd edition, https://www.sws.bom.gov.au/IPSHosted/INAG/scaling/japanese_manual_v3.pdf
- Piggott & Rawer (eds.), 1972, U.R.S.I. handbook of ionogram interpretation and reduction, 2nd edition, UAG-23A, World Data Center A, https://www.sws.bom.gov.au/IPSHosted/INAG/uag_23a/UAG_23A_indexed.pdf (note high latitude supplement linked from top link too)

Using ionosonde data to estimate hmF2

Motivation I think you could usefully make it even more explicit here what the motivation for using ionospheric measurements is, and hence why so many authors have tried. A few suggestions:

```diff
@@ L97 @@
-effects are altitude dependent which will lead to differing results. Searching
+effects in the upper atmosphere are altitude-dependent, which will lead to
+unhelpfully differing results, depending on the altitudes of the specific
+observations in each study.
+By contrast, searching
```

```diff
@@ L99 @@
-The longevity of ionospheric data series
+Furthermore, of the upper atmosphere regions, the ionosphere is unique as it can
+be observed remotely with relative ease. Due to the ionosphere's importance for
+long-distance radio communication, such observations have been made routinely
+since the early 20th century. The resulting longevity of ionospheric data series
```

```diff
@@ L101 @@
-can be extracted from the data, allowing for all other effects.
+can be extracted from the data, with all other potentially-confounding effects compensated for.
```

```diff
@@ L102 @@
-The first published analysis
+This potential has motivated similar re-analysis of ionospheric data,
+seeking evidence of climate-driven trends.
+The first published analysis
```

Be more explicit on what relevance is:

```diff
@@ L115 @@
-Kokubunji in Japan with
+Kokubunji in Japan (the same station examined here) with
```

Ap and am L131: Ap index vs am index: Somewhere appropriate (here / a very short new section 2.3 / section 3.4) it would be good to have a brief discussion of this, making it explicit why you use am, rather than the Ap you flag has been used by most previous studies have used Ap!

Took me down a bit of a rabbit hole refreshing myself on am - had been a while! I think I eventually got my answer from Mienville & Berthelier, 1991, The K-derived planetary indices: Description and availability, https://doi.org/10.1029/91RG00994:

- end of their section 5 basically says that better sensitivity and distribution of Km network (cf Kp) makes am more suitable for statistical studies.

If this sort of reasoning governed your collective choice to use am, could you add a brief explicit sentence to this effect, citing Mienville & Berthelier.

- Or Lockwood+ 2019 which you cite already - looks like similar reasoning in there

Say this as if "accurate and painstaking" is required, favouring am over Ap seems very aligned: worth making your reasoning explicit, so others can follow suit!

Confounding factors Make it more explicit that the tempting elision that "remaining trend == climate change" is dangerous? And loop back to your different effects from introduction? E.g.

```diff
@@ L134 @@
-Mikhailov and Marin
+When analysing such trends in residuals however, it is important not to assume
+any local environment changes which may underly this can be attributed to
+greenhouse forcing alone.
+The previously-discussed effects altering thermospheric composition (3)
+and wind patterns (4) risk being confounding factors, as they can potentially
+also lead to trends on long timescales.
+Mikhailov and Marin
```

Clarify scope L160: worth being blunter, as think you've been nicely impartial in the actual analysis?

- if you adopt my "in section X" suggestion below, you could leave this as-is, and have summary outcome in there

```diff
@@ L160 @@
-to investigate the efficacy of
+is to demonstrate the potential pitfalls in
```

L161: current sentence ~implies the present work is looking into this reconciliation globally. Section 3.4 is a great start, and you've ~convinced me that global trends need to be reinterpreted in light of this paper, to see if reconciliation possible. But given this paper only examines Kokubunji, Chilton, and Stanley, this sentence may need watering down a little. E.g.

```diff
@@ L161 @@
-investigate the extent to which this can reconcile the difference
+demonstrate that such effects may have potential for reconciling the differences
```

Add an outline L162: you lost me a little in section 2, trying to keep key bits in my working memory, and interpreting the detail you're exposing in light of this, while we're (necessarily!) knee-deep in complex empirical equations!

I think you could usefully insert a brief outline here at L62, summarising what you'll do & key findings. A la "In Section 2, we do XYZ and show, In Section 3 …".

Would let readers see where you're going, so help follow the thread of your argument better once in there - and flick back to this if they lose thread!

Estimating hmF2 from empirical formulae

Existing parameters To make point clearer to those less familiar with ionosondes, something like following:

```
@@ L167 @@
-from existing scaled ionospheric parameters
+from existing standard ionospheric parameters (e.g. foF2, MUF).
+Determining these from an ionogram only required a scaling process (not inversion),
+hence these were routinely calculated from ionograms at the time of measurement.
```

Minor fluctuations L204: I'm not following. Can you try to interpret what B&D / their Paul ref means? Is the point that:

- these are true F1 layers, but hard to pin down, as noisy?
- this is effectively noise (ionospheric variability / instrumental) getting mislabelled as F1?

Good if you could also tweak "more ubiquitous" on L207 to whatever is more appropriate:

- more ubiquituously-recorded
- less contentious

Shape of profile L213: I don't follow - can you clarify what this "well above the limit" profile point means? I assume this supports B&D's argument that their approach means any F1 ionisation present can be ignored. But I'm unclear why. Especially if they were doing this off synthetic ionograms. Is this:

- a "how far are F2 and E separated along x-axis" measurement error type argument?(!)
- a physics-based argument on the retardation between E and F2 exceeding retardation from F1 features above this cutoff foF2/foE ratio?
- other?

Limitations Can you give a physical insight? E.g.

```
@@ L217 @@
-X_E < 1.7, which
+X_E < 1.7, due to a more ionised E layer, which
```

Possibly worth adding foF1? Belabours point from previous sentence, but think OK, as it's a key finding of this paper that this is wrong!

```
@@ L219 @@
-field.
+field. And by definition this assumes F1-related effects are negligible.
```

Which formula? L246: "refinements of the formula":

- a specific one? If yes, can you specify which one
- if general class, can you be clearer, e.g. "refinements of similar formulae"

What is impact of foE = 0.4 MHz assumption? L279: can you indicate what expected impact of this fixed foE assumption is on hmF2? (realise elsewhere you say noise in results - I'm less sure! Here's why)

I assume this acts as a floor - the true foE would likely be lower, but this can't be measured, so it gets floored (ceilinged really!) to 0.4 MHz?

If so, what is this going to do to the "true" hmF2?

From a quick play with B&D eqn in Excel, and guessing some values:

- foE = 0.001 - 0.4
- foF2 = 5 - 7
- M(3000)F2 = 3 I think this floor will impose a very small < 5 km shift downwards on the value of hmF2 which would be calculated with the "true" foE. And this would affect night time values, so the ~300 km mark.

So should have a negligible effect on your plots & analysis? But would be good to have you check this, and to convey any such message.

The Kokobunji ionosonde data

Clarify binning, ionosonde data retrieved L282: worth defining & justifying binning better. And enumerating all data retrieved.

Until I looked at code / thought hard about plots, I thought your monthly medians collapsed local time information. Also, it's only later that justification for medians appears. And that you also download F1 values.

How about something like following? Longer, but covers all data, and makes why & how for averaging more explicit up front.

"To estimate hmF2, scaled critical frequency parameters for the E, F1, and F2 layers (foE, foF1, and foF2), as well as the M(3000)F2 factor were downloaded. Monthly averages were then calculated for these data, to protect against outliers caused by short-lived space weather events that are not representative of the data on monthly timescales. Specifically, hourly monthly medians were used - medians calculated across corresponding hours (in local time) within a given month. For a given year, this yields 288 median values (bins of 12 months, and 24 local time hours). Such hourly monthly medians of foE, foF2 and M(3000)F2 were then used to estimate corresponding hourly values of hmF2, the true height of the F2 layer, using…"

The Middle and Upper Atmosphere (MU) Radar

Clarify averaging

```diff
@@ L318 @@
-height profiles were then averaged to
+height profiles were then averaged over the four antenna positions to
```

```diff
@@ L327 @@
-data, monthly means were
+data, corresponding hourly monthly means were
```

```diff
@@ L328 @@
-35 years of data used in this study, observations have
+35 years of ISR data used in this study, ISR observations have
```

```diff
@@ L329 @@
-10080 bins (monthly averages for each hour over 35 years).
+10,080 bins (monthly means, in bins for each local time hour and month, over 35 years).
```

I've suggested moving justification below to ionosonde section

```diff
@@ L329 @@
-parameters, this is to protect against outliers in ionospheric
-concentrations caused by short-lived space weather events that
```

```
-are not representative of the monthly data. For the ISR data,
+parameters, for the ISR data,
```

Seasonal and diurnal comparison

Fig 1 interpretation L339: recommend flagging up in text a few extra things which jump out at me from Fig 1:

- The ISR data are noisier.
  - I assume because the campaign nature of ISR measurements means fewer underlying datapoints are available to constrain each hourly monthly mean datapoint?
- At solar min, the ISR data hit a floor at 180 km due to your sporadic E contamination procedure (suggest flag in figure too?)

Fig 3 interpretation

```
@@ L351 @@
-averaging over the 35 years
+averaging the aforementioned 10,080 bins over the year axis, for the 35 years
```

L352: you don't explain the white lines - I had to look at your code to double-check interpretation. Can you:

- add an explanation that these are the sza = 90, 100 limits used to isolate the previously-discussed day/terminator/night populations in Fig 2
- make these lines a bit thicker - the 90 one is particularly hard to see
- add lines like this to all subpanels, and all Figures like this
- consider a very brief discussion of results with respect to these?

Re latter, when I hand-draw similar lines on ISR, I see that for both ISR and ionosonde, dawn hmF2 hugs this sza line across all months. But that sza sn't constrain dusk hmF2, esp in winter months. I expect a ~easy Appleton anomaly explanation for diurnal, something else for seasonal. Would be good to have this flagged, and briefly explained by you!

```
@@ L353 @@
-peaks in hmF2 at
+peaks in night-time hmF2 at
```

Flag up midday summer peak

- you concentrate on this difference for ~rest of paper
- so good to flag up here that this is the target you'll concentrate on

```
@@ L354 @@
-a stronger peak
+an unexpected strong peak
```

```
@@ L355 @@
-source of this bias
+source of this midday summer bias
```

Fig 4 interpretation L356: foF2/foF1: flag that this corresponds to xF from earlier eqn 4

L357: foF2/foF1: can you explicitly link back to discussion of eqn 4? Make point that B&D's range is nominally xF > 1.2 (note earlier query). But that here you're seeing it looks like it's more restrictive, that the 1.2 <= x.F <= 1.6 range is problematic?

Correcting for foF1 L360: can you be explicit on how you've done this?

Ideally break this out into an equation, as there are a couple of times later where you could usefully refer back to this.

Good to have this made explicit - currently I've got questions/guesses below.

I assume making the assumption that for

- `xF < 1.6`, `hmF2_iono_true ~= hmF2_ISR`
- `xF >= 1.6`, `hmF2_iono_true = hmF2_iono_raw`?

And so you filter ionosonde data to select 1st branch (`xF < 1.6`).

And for those data you apply linear model `hmF2_iono_true = (231*xF) + hmF2_iono_raw - 356`?

And you apply this to the binned hourly monthly median values (i.e. rather than to raw values, pre-binning)?

And for Fig 5, you do this after binning, but before collapsing over the year axis?

Long-term bias

L380: reword - this is unclear. My guess:

```diff
@@ L380 @@
-to correct affected daytime values within the hourly monthly median hmF2 values.
+to correct affected hmF2 values (where foF2/foF1 < 1.6)
```

- If you've followed my suggestion above to add an equation for your model, you could usefully refer to it here

```diff
@@ L381 @@
-When plotted
+When daytime values are plotted
```

The impact of foF1 on the long-term drift in estimates of hmF2

L411: if you followed my suggestion re model equation, here you could say "values, yielding similar models to eqn X, but for each year".

L413: add unit and make the interpretation clearer here!

```diff
@@ L413 @@
-of ±10% (±25km at 250)
+of up to ±10%(±25km at 250 km, i.e. a little larger than the
+~20 km decrease expected from climate change)
```

```diff
@@ L420 @@
-composition.
+composition, arising in turn from long-term changes in geomagnetic activity.
```

L440: can you make this point a little clearer?

I'm getting a bit lost on the temporal aspect - annual vs semi-annual. I think this is a red herring, not least as you're showing a clear difference in Fig 9 when you simply present annual averages!

I think the argument is ultimately altitudinal - what I'm (mis)understanding so far:

- near the poles, the neutral upwelling in summer disproportionately suppresses nmF2 (??? not sure why nmF1 less affected?)
- which pulls the yearly average of foF2/foF2 down?
- whereas further from poles, there isn't a strong altitudinal variation by season.

Conclusions

L496: latitude only? Does longitude need considering too cf circulation. Or is this ~purely meridional? Ask as ISRs which Bilitza 1979 used all have similar MLON (ie MLT) values.

Refs

Good if you can add missing URLs

- Bradley & Dudeney: https://doi.org/10.1016/0021-9169(73)90132-3
- Dudeney: https://nora.nerc.ac.uk/id/eprint/509197
- …

Possible typos

```
@@ L22 @@
-peak density of the F2 layer, foF2, would
+peak density of the F2 layer, nmF2, would
```

```
@@ L86 @@
-automatically identify and invert
+automatically scale and invert
```

L92: citepshangguan2019

L96: citepphilipona2018

L175: I think this should be inverted, no? M(3000)F2 = MUF/foF2 definition used by Elias 2017 (eqn 13), Japanese manual (p109), UAG (p23)

```
@@ L175 @@
-The ratio foF2/MUF
+The ratio MUF/foF2
```

L282: concentration vs critical frequency Think this should be critical frequencies, no, given foE and foF2, and formula definitions? So Note "concentration" is used quite often, and on a few other occasions is applied to a frequency, rather than a density/concentration. Realise these are ~equivalent (per eqn on line 68-69), but this doesn't help clarity. Could you do a search for "concentration", and try to rationalise this?

```
@@ L282 @@
-The peak electron concentrations of the E and F2 layers (foE and foF2)
+The critical frequencies of the E and F2 layers (foE and foF2)
```

L296: fragment?

```
@@ L296 @@
-integrated over. In the ionosphere, While the
+integrated over. While the
```

```
@@ L342 @@
-Daytime data (for which the solar zenith angle, sza > 100◦ )
+Daytime data (for which the solar zenith angle, sza < 90◦ )
```

```
@@ L342 @@
-twilight data (90◦ ≤sza≥ 100◦)
+twilight data (90◦ ≤ sza ≤ 100◦)
```

```
@@ L382 @@
-values (figure 7, the correction
+values (figure 7), the correction
```

```
@@ L383 @@
-the remaining offset is not 1:1.
+the remaining gradient is not 1:1.
```

L388: ionosonde values being "generally still lower" and "underestimate of the F2 layer height". Err, are you sure?

- Lower in some bins - equinox night
- But not in others - midsummer night
- cf Fig 2: the two clumps I've flagged above/below trendline

- No idea about "general" (average over all months), and how much there's in it!
- Think this is where my requested difference subpanel could help!

L469-L471: Comparing the before/after gradient values, I think you're currently a factor 10x too large in your reported improvements! Hence following changes

```
@@ L469 @@
-resulted in an improved relationship
+resulted in a slightly improved relationship
```

```
@@ L470 @@
-the revised gradient for daytime hmF2 increases by 0.3 to 0.89±0.01
+the revised gradient for daytime hmF2 increases by 0.03 to 0.89±0.01
```

```
@@ L471 @@
-the equivalent after correcting for the presence of foF1 increases by 0.3 to 0.92±0.01
+the equivalent after correcting for the presence of foF1 also increases by 0.03 to 0.92±0.01
```

Fig 2 caption

```
-gradient of 0.71pm0.01
+gradient of 0.71±0.01
```

Fig 4 caption

```
-(gradient 231±21, offset −356±31km
+(gradient 231±21, offset −356±31 km)
```

Fig 5 caption

```
-by the ISR (figure 5 lower panel
-).
+by the ISR (figure 4, lower panel).
```

Fig 6 caption

```
-(as identified in 4. It can
+(as identified in Figure 4). It can
```