Development of the Model of the Ionosphere

Carl Luetzelschwab K9LA February 2016 (with Feb 2017 Update)

In my various articles over the years I have made it a point to state that our understanding of the ionosphere is statistical in nature, not deterministic. This is due to the significant day-to-day variability of the ionosphere, coupled with our lack of a full understanding of <u>all</u> the parameters that affect the day-to-day variability. As a result, we do not have daily propagation predictions – we have predictions that are statistical (probabilities) over a month's time frame. The present correlation is between monthly median ionospheric parameters and a smoothed solar index.

So how did we end up in this situation? Simply put, that's where the data led us. Thus the purpose of this month's Monthly Feature is to review how we got to where we are. Right up front I'd like to acknowledge the efforts of Bill NQ6Z in helping with this historical research. He found many key old documents that allowed this story to be put together.

You might think coming up with a model of the ionosphere for propagation prediction purposes would have been a pretty simple endeavor. Scientists had data on what the Sun was doing and they had data on what the ionosphere was doing. But when they tried to correlate those two parameters on a daily basis, the correlation was extremely poor. Thus they had to look at the data over a longer-term. I'm sure they asked themselves "Should we average the values? If so, over what time frame should we average? Or should we change to something other than averages?" Those are interesting questions, so let's get started at the beginning.

As radio emerged as a viable long-distance method of communications, one issue was predicting what frequency and time of day would allow propagation from Point A to Point B. In order to make these predictions, a model of the ionosphere was needed. And in order to develop a model, data was needed. The piece of equipment responsible for taking ionospheric data was the ionosonde.

An ionosonde is a swept-frequency radar looking straight up. It measures the time of flight from when the pulse is transmitted until it is received back on the ground (assuming it comes back!). From this data, scientists could determine the critical frequencies of the ionosphere (and ultimately the electron density profile). As a side note, the ionosphere was initially called the Kennelly-Heaviside layer after the two scientists who in 1902 independently postulated the existence of an electrically conducting region in the atmosphere (note A). Another side note – the term "ionosphere" was introduced in 1932.

The concept of the ionosonde was demonstrated in 1925, and in early 1930 the critical frequencies of the E, F1 and F2 regions were being measured manually once each week in the vicinity of Washington, D.C. Beginning in May of 1933, automatic multi-frequency records were made hourly. Weekly values of the noon F2 region critical frequency were averaged by months for September 1930 to December 1935, and this data is shown in Figure 1 (reference 1) along with monthly average sunspot numbers (note B).

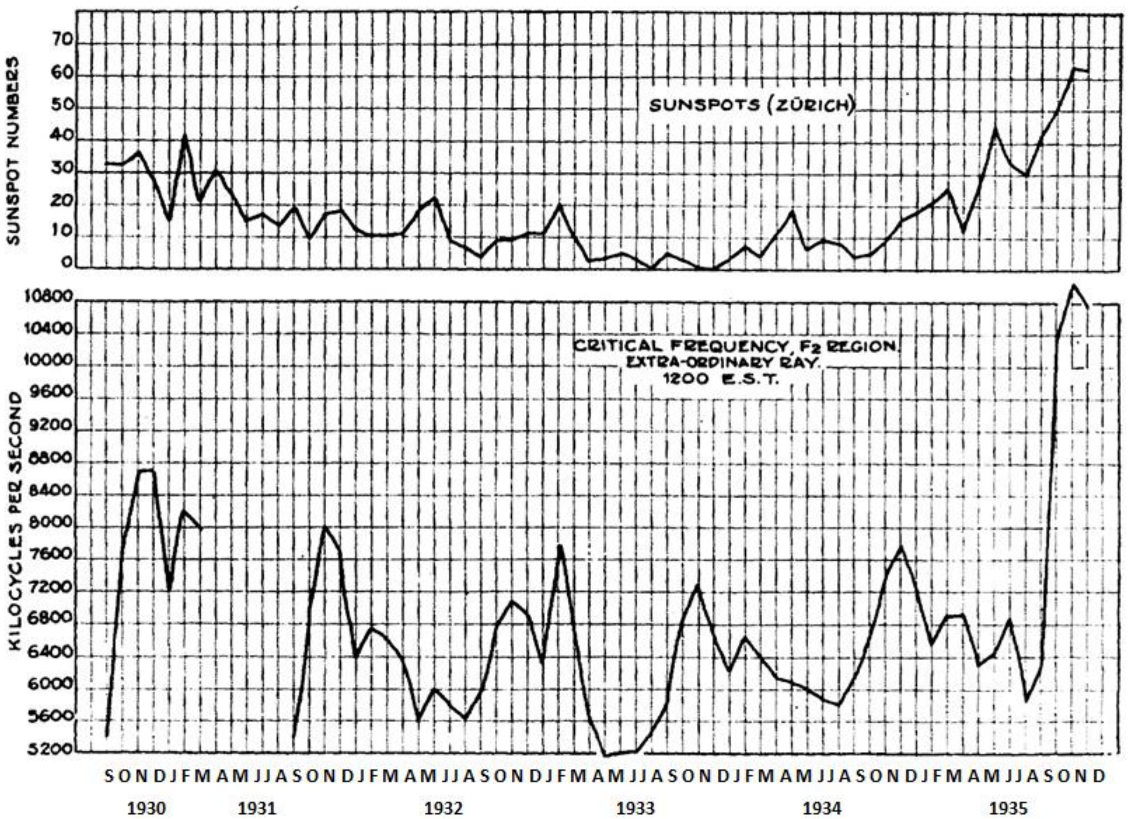

Figure 1 – Weekly Averages by Month of Critical Frequencies

The sunspot data in conjunction with the dates on the horizontal axis tell us that the measurement period was solar minimum between Cycles 16 and 17.

As for the critical frequency data, three important observations were made. First, the ionosphere was more highly ionized in the winter months than in the summer months. Second, and more in line with the title of this Monthly Feature, the author of the referenced paper made the statement "*a comparison of the critical frequency curve with the average sunspot curve shows no certain correlation between the two phenomena for corresponding months*". In other words, there didn't appear to be any short-term correlation between what the Sun was doing and what the ionosphere was doing.

Third, and also in line with the title of this Monthly Feature, there is a general trend in critical frequency and sunspot number when considering the envelopes of the two curves. In other words, there's a hint at a long-term correlation. You can see this by eyeballing a trend line onto the critical frequency curve. The trend line would be decreasing from 1930 through 1933, and then it would start back up – which is what the monthly average sunspot number is doing.

Now let's move forward to 1938. Figure 2 shows twelve-month running averages of critical frequencies and sunspot numbers (reference 2) from solar minimum between Cycles 16 and 17 to the peak of Cycle 17.

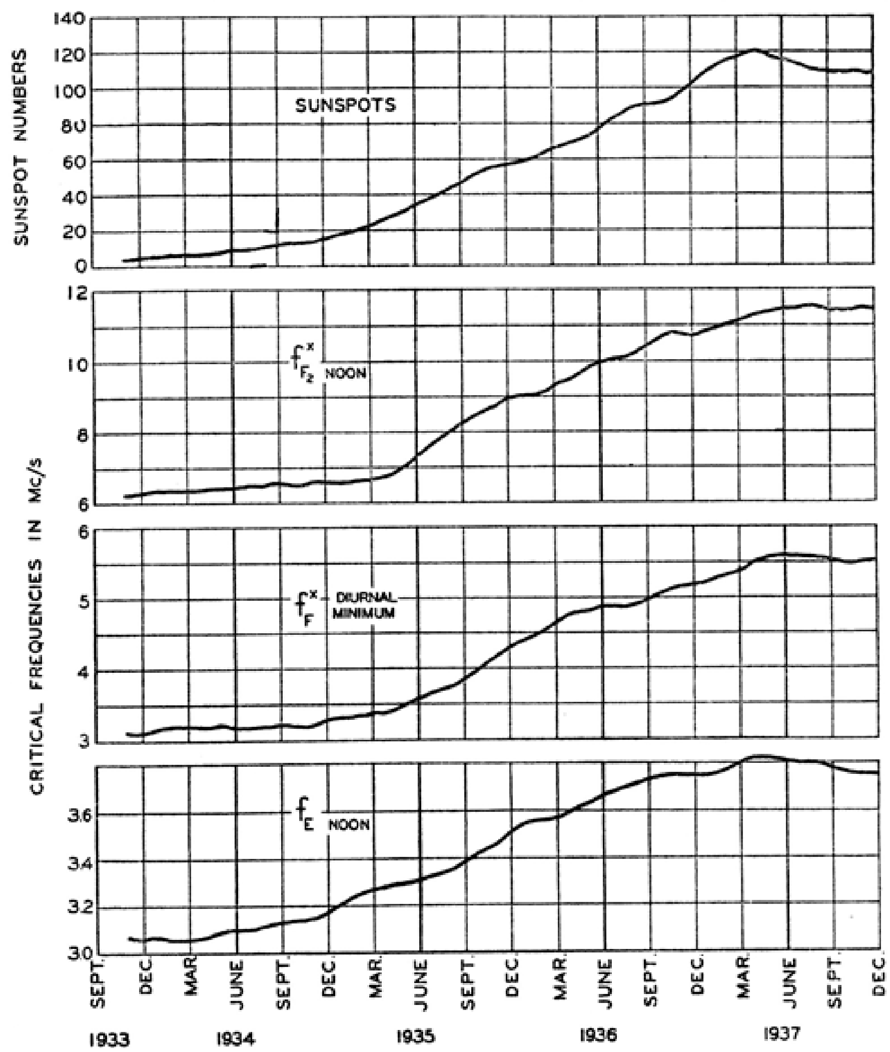

Figure 2 – Twelve-Month Running Averages of Critical Frequencies

It should be quite obvious that the correlation between all the critical frequencies and the sunspot number is very good when viewed over an extremely heavily averaged period (note C).

But there's a problem. Compare the F2 region critical frequency curve of Figure 2 (second from the top) to the F2 region critical frequency of Figure 1. The twelve month running average of the F2 region critical frequency removes the seasonal variation of the F2 region that is seen in Figure 1. We now don't see the fact that the ionization in the F2 region in the northern hemisphere in the winter is much greater than in the summer.

To remedy this, scientists began presenting data as monthly average critical frequencies to assure that the seasonal characteristics of the ionosphere were maintained. The Proceedings of the I.R.E ran monthly articles titled Characteristics of the Ionosphere at Washington D.C. As an example, Figure 3A (reference 3) shows monthly average values of critical frequencies for undisturbed days for May 1940.

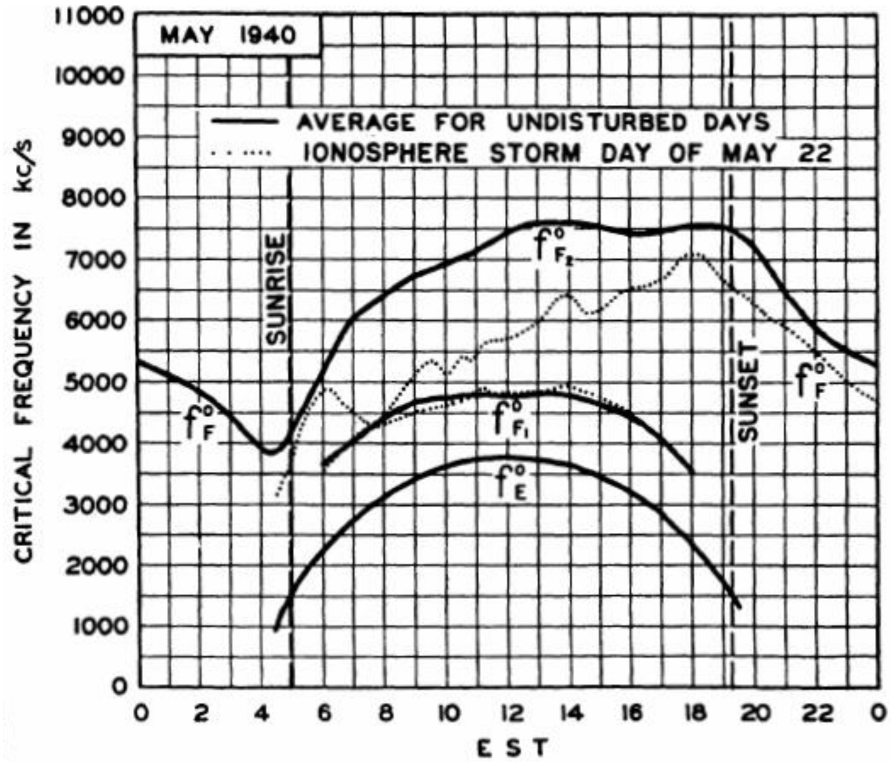

Figure 3A – Monthly Average Critical Frequencies

The characterization of the ionosphere was local in nature, and the horizontal axis was in Eastern Standard Time. Additionally, the distribution of all the hourly F2 region values about the May 1940 monthly average was given. This is Figure 3B for the data in Figure 3A. Thus we're starting to see a statistical model emerge for ionospheric parameters.

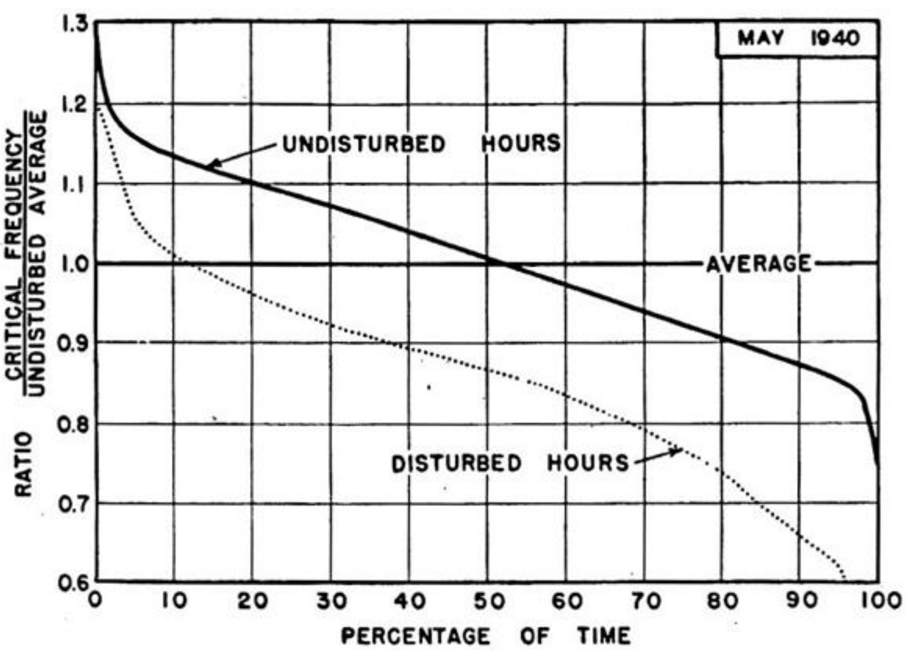

Figure 3B – Distribution of the F2 Region Critical Frequency about the Average

The next several years saw more and more data being gathered as the thrust moved from a local characterization of the ionosphere to a global characterization of the ionosphere. Figure 4 shows a worldwide map from early 1947 dividing the ionosphere into zones (reference 4): West, East and Intermediate. Note that the West zone is when the magnetic equator is most south, the East zone is when the magnetic equator is most north and the Intermediate zones are when the magnetic equator is transitioning between most south and most north.

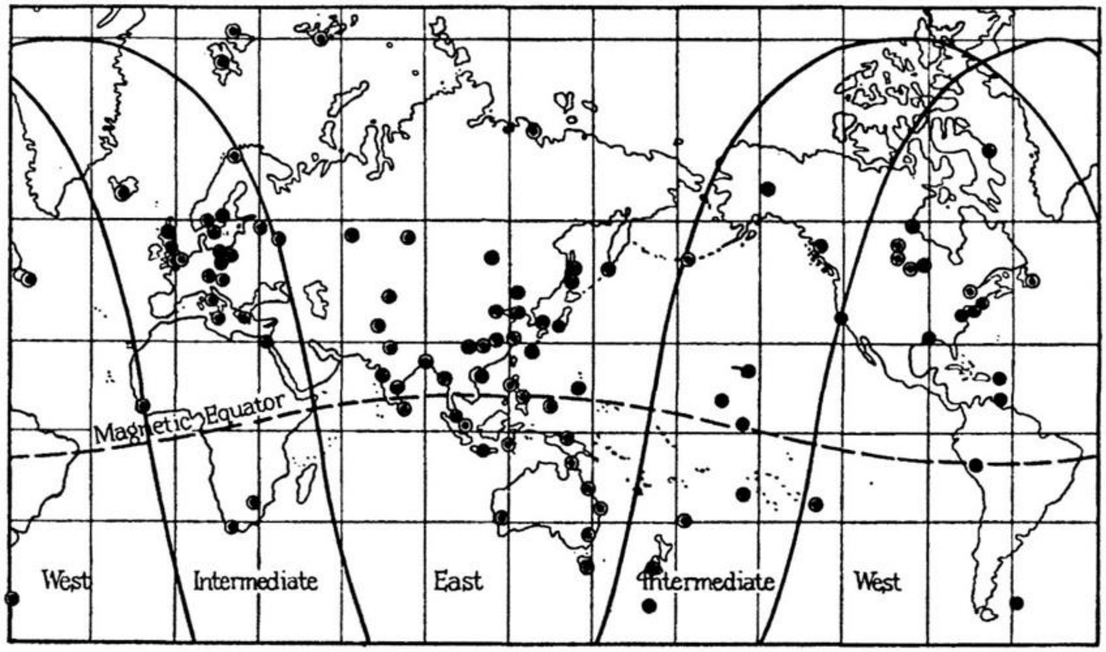

Figure 4 – Zones of the World for Characterization of the Ionosphere

Why was the world divided into zones? Scientists must have recognized that electrons, being charged particles, are affected by the Earth's magnetic field. As the geomagnetic equator meanders from about 10^{o} North geographic latitude (East zone) to about 10^{o} South geographic latitude (West zone), the electrons will follow. Thus characterizing the ionosphere with zones was the best way to handle this as more and more worldwide data was taken.

The black dots are ionosondes – as you can see, the worldwide characterization of the ionosphere was indeed underway in 1947. The authors of this paper also discussed whether monthly average values or monthly median values of the critical frequency were better. I believe this was driven by the fact that the data was statistical in nature, and the question was which parameter was best for the observed distribution. It was noted that the difference in monthly average and monthly median critical frequencies was not negligible (that is certainly true), but the difference at the time was deemed inconsequential in terms of ionospheric predictions.

In September 1947 we begin to see critical frequency data presented in terms of median values for sporadic E (reference 5). It's not clear what the F2 region and normal E region data were – monthly average values or monthly median values. And the world was still divided into zones.

In December 1947 (reference 6) data for the F2 region was presented in terms of monthly median values. This article appears to be one of the first that has data in the format of what our model of the ionosphere looks like now – a correlation between a smoothed sunspot number and monthly median values of ionospheric parameters. Figure 5 shows this important data.

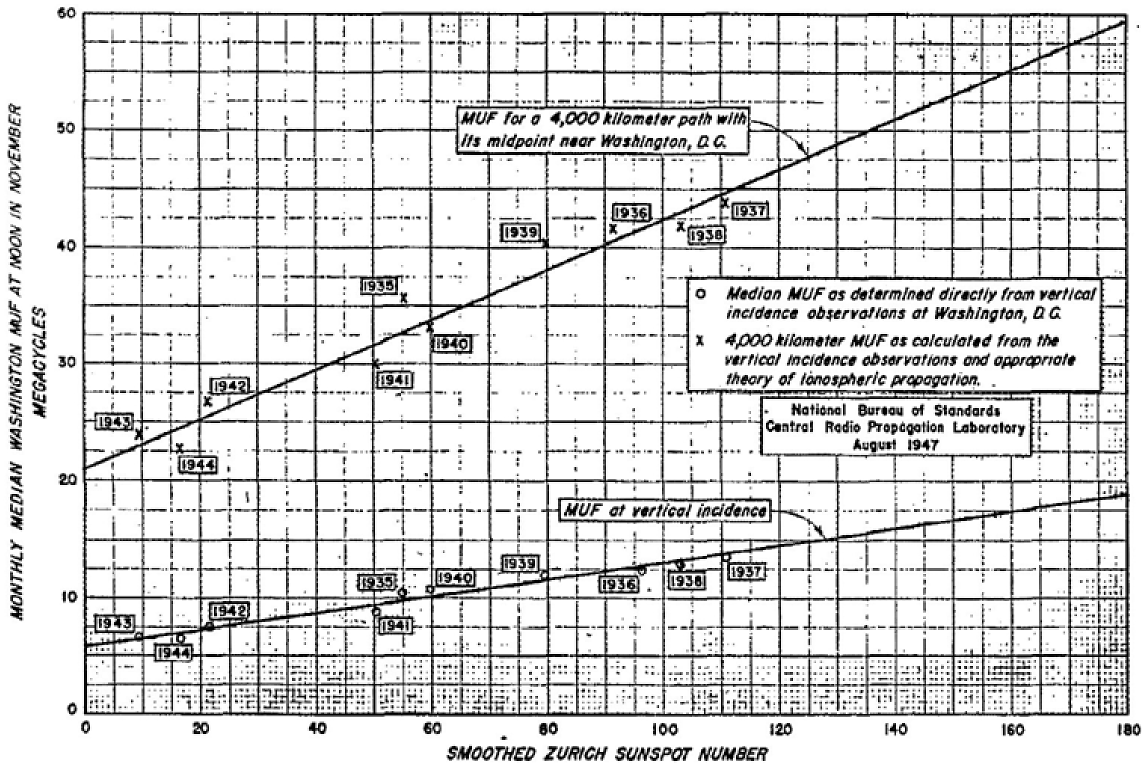

Figure 5 –Monthly Median Parameters vs Smoothed Sunspot Number

Note that the correlation for both the critical frequency and the maximum useable frequency (MUF) is very good. In other words, the data is minimally scattered about the two black linear trend lines.

Thus it looks like 1947 was the year when most of the model of the ionosphere was defined (monthly median ionospheric parameters versus a smoothed solar index). It would have been nice to have found an official document that said "the use of monthly median critical frequencies is preferred over average values for the following reasons." NQ6Z and I never found such a document (see February 2017 Update after the Notes at the end). My guess is that an average value is better suited for use with a normal distribution, while a median value is better suited for a non-normal distribution (as are critical frequencies – note D). So one half of the puzzle is in place – monthly median ionosphere parameters

There still appeared to be an issue with what format of the sunspot number to use – a monthly average value or a smoothed value. Figure 5 was in terms of a smoothed sunspot number, but the monthly average sunspot number also gave good correlation – and was still in the running. This issue was discussed in a 1948 paper (reference 7). This paper had plots of the monthly median critical frequency of the F2 region at Washington D.C. at two times versus a monthly average sunspot number and versus a smoothed sunspot number. Figure 6 shows the data for the noon local time for the years 1934-1946.

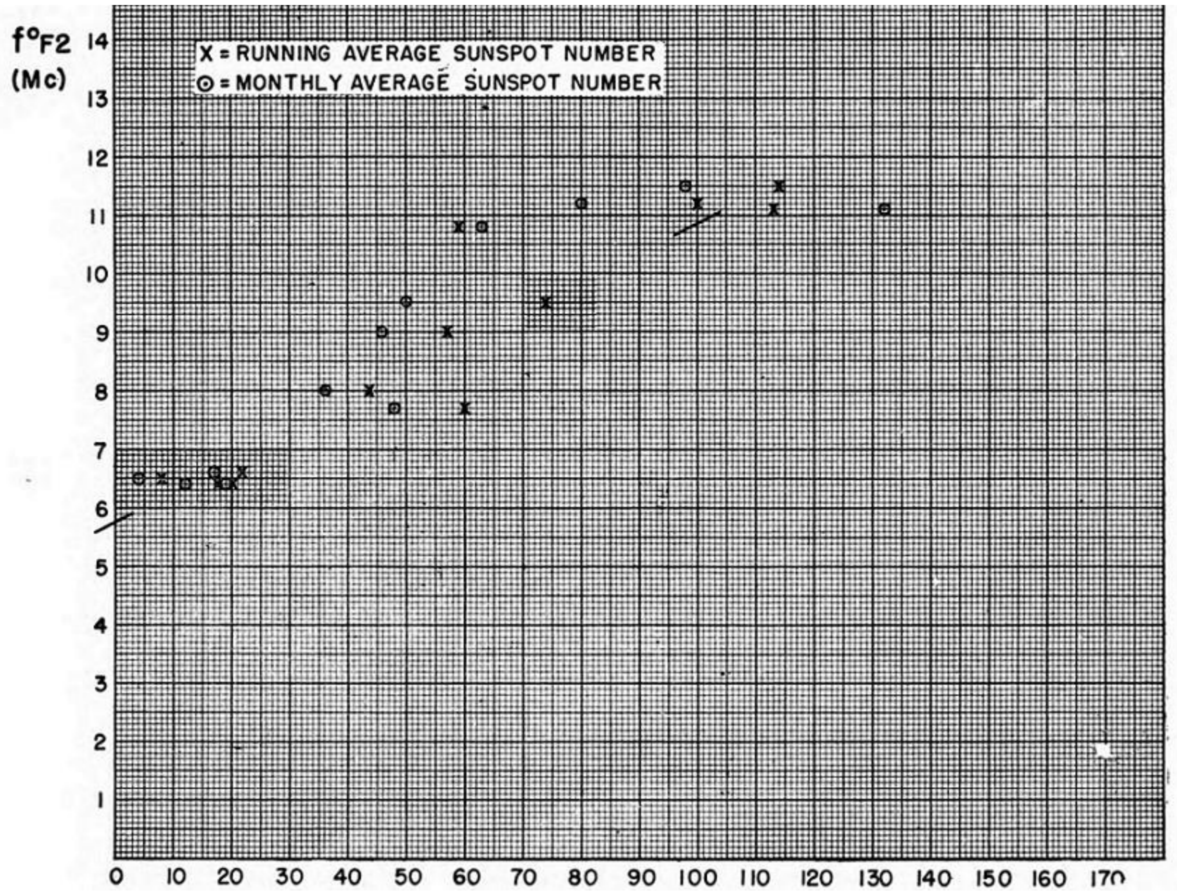

Figure 6 – Monthly Median Values versus Two Formats for Sunspots

The text in this reference states that "in general, points for the 12-month running average sunspot number seem to deviate less from a straight line than points for the monthly average sunspot number". I'll be the first to admit that this conclusion might be tough to see from Figure 6 due to some difficulty in distinguishing between the X's and O's in this copy of the plot and the lack of a continuous trend line. If you enlarge Figure 6, you can more easily distinguish between the two sets of data. And if you add a linear trend line between the two short line segments at x=0 / y=5.8 and at x=100 / y=10.8, you come to the same conclusion as the authors.

Fortunately, a book published in 1960 (reference 8) gave similar data (monthly median critical frequencies versus monthly average sunspot numbers and versus smoothed sunspot numbers) in tabular format, which allows one to better confirm this conclusion. Figure 7 shows the correlation of the F2 region critical frequency with the running average sunspot number (RASSN – what we now call the smoothed sunspot number) and of the F2 region critical frequency with the monthly sunspot number (MSSN – monthly average sunspot number).

Station	Zone	Month	Number of observations (n_i)		Standard deviation(s) about regression on RASSN (Mc)		Correlation Coefficient (r)			
							RASSN	MSSN	RASSN	MSSN
Hour (LT)			0000	1200	0000	1200	0000		1200	
Tromso	Auroral = A	June	5	6	0.204	0.100	0.914	0.872	0.991	0.966
		Sept.	3	6	0.141	0.294	0.997	0.999	0.985	0.954
		Dec.	4	6	0.696	0.373	0.908	0.926	0.993	0.995
Fairbanks	A	June	14	14	0.299	0.262	0.932	0.935	0.943	0.935
		Sept.	14	14	0.548	0.377	0.874	0.908	0.950	0.914
		Dec.	14	14	0.680	0.490	0.657	0.738	0.982	0.954
Ft. Monmouth	Medium Latitude = M	June	6	6	0.546	0.653	0.908	0.900	0.637	0.687
		Sept.	6	6	0.325	0.416	0.979	0.980	0.983	0.978
		Dec.	6	6	0.373	0.331	0.971	0.982	0.994	0.981
Washington, D.C.	M	Dec.		14		0.441			0.985	0.962
Trinidad	Lower Latitude = L	June	7	7	0.958	0.601	0.902	0.867	0.926	0.905
		Sept.	7	7	0.637	0.816	0.973	0.956	0.919	0.882
		Dec.	7	7	0.637	0.655	0.942	0.956	0.954	0.949
Guam	Equatorial = E	June	10	10	0.943	0.510	0.955	0.925	0.973	0.955
		Sept.	10	10	0.994	0.313	0.965	0.938	0.734	0.645
		Dec.	11	10	0.410	0.400	0.989	0.969	0.975	0.943
Leyte	E	June	4	4	0.272	0.170	0.993	0.952	0.997	0.968
		Sept.	3	3	0.794	0.315	0.953	0.885	0.996	0.965
		Dec.	3	3	0.113	0.153	0.999	0.978	0.997	0.950
Nairobi	L	Dec.	3		0.933		0.969	0.969		
Johannesburg	L	June	10	10	0.127	0.464	0.704	0.734	0.982	0.983
		Sept.	11	11	0.424	0.543	0.924	0.960	0.974	0.980
		Dec.	11	11	0.373	0.580	0.963	0.923	0.929	0.886
Hobart	M	June	11	10	0.307	0.772	0.928	0.942	0.938	0.879
		Sept.	11	11	0.530	0.780	0.949	0.973	0.951	0.961
		Dec.	11	11	0.414	0.330	0.949	0.944	0.910	0.924

Figure 7 – Smoothed Sunspot Number vs Monthly Average Sunspot Number

Note that this evaluation used ten ionosondes around the world (a more worldwide view than the data in Figure 6) at various latitudes with data mostly for three months and at two times. The last four columns of data are the correlation coefficients between the critical frequency and the sunspot number format. As a reminder, 1.000 indicates a perfect correlation. Although there were instances of the MSSN having a higher correlation than the RASSN, overall the RASSN value performed better. Thus the final piece of the puzzle was in place – we use a smoothed solar index (note E) for our propagation predictions.

We're just about done with this story. The last item to address is the concept of the zones of the ionosphere in Figure 4. Worldwide maps of the ionosphere were initially published in this format. They were for each of the zones in terms of local time throughout the day.

But if you're going to publish ionospheric maps of the world for the purpose of making worldwide predictions, you should take the format out of the "local" category using bands of longitude with local time and into the "worldwide" category showing how the ionosphere is ordered about geomagnetic coordinates in universal time. This transition to worldwide data began in January 1963 in the publication of ionospheric parameters (reference 9) from the Central Radio Propagation Laboratory in Boulder, CO. The new maps showed the worldwide ionosphere versus longitude for every two hours of universal time. Figure 8A shows a representative "old" map of the F2 region for the West zone, while Figure 8B shows a representative "new" map of the F2 region for the entire world.

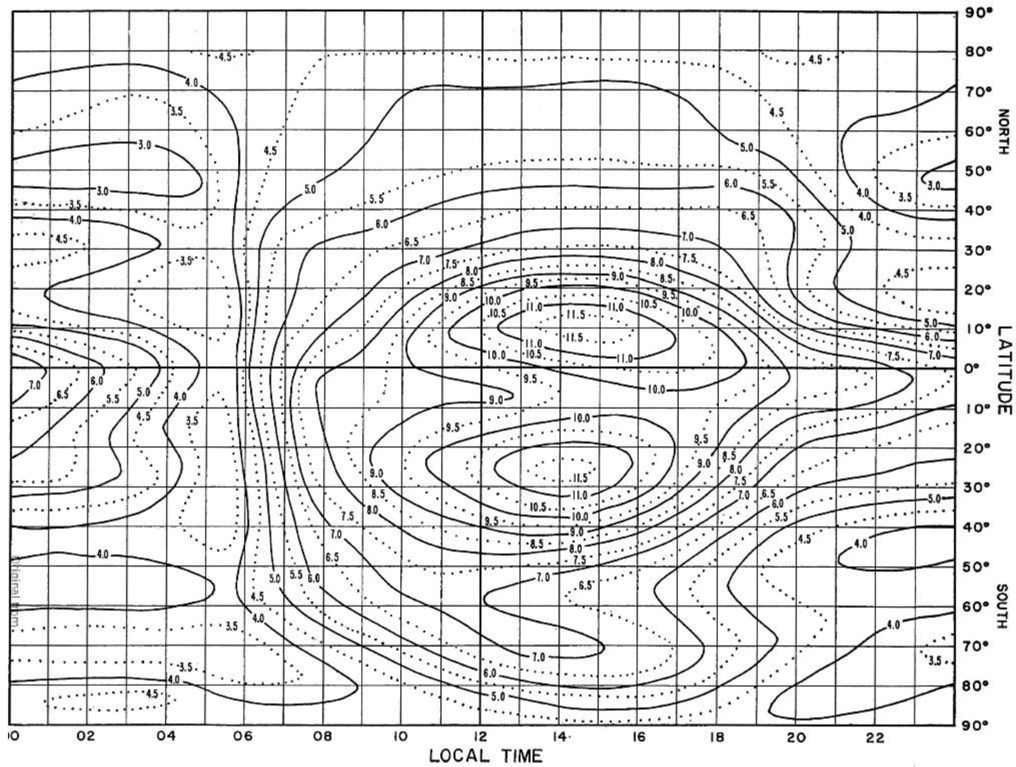

Figure 8A – F2 Region Critical Frequencies the "Old" Way in the West Zone

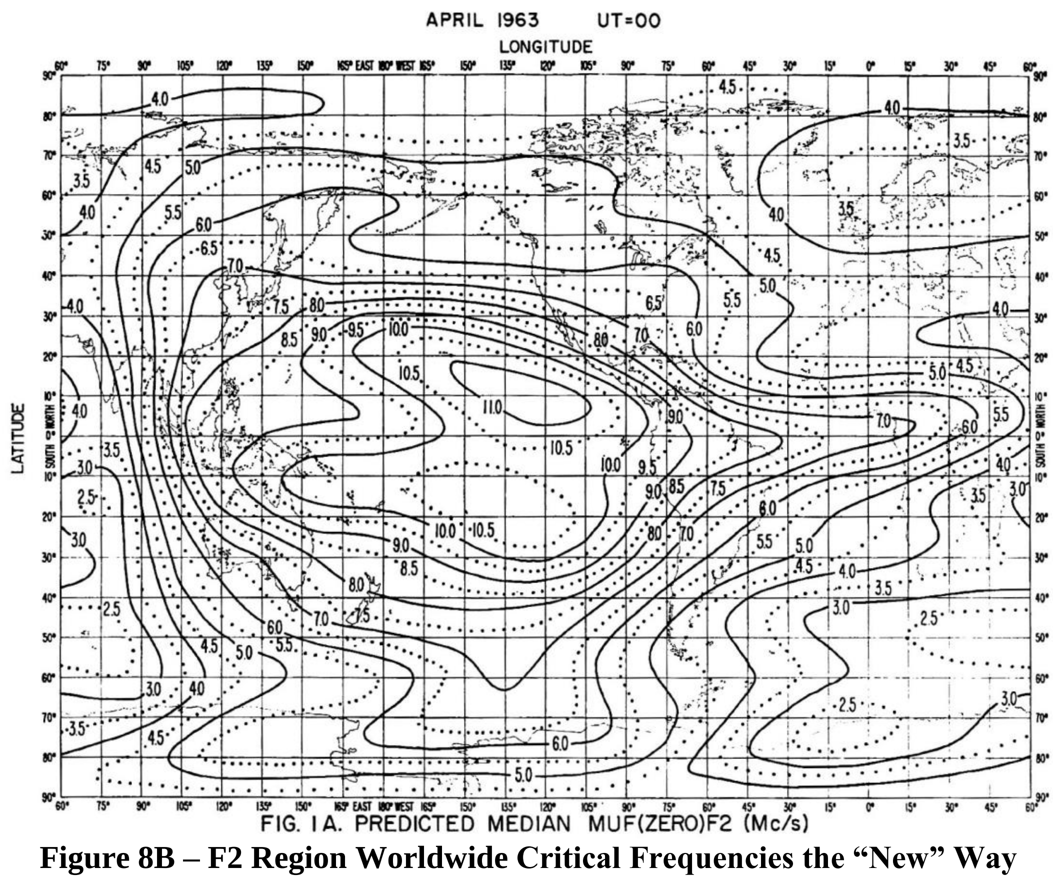

Figure 8B – F2 Region Worldwide Critical Frequencies the "New" Way

We have now reviewed history from the start to where we are now.

Now I don't want to leave you with the idea that the model of the ionosphere has been stagnant since 1963. In terms of the correlation being monthly median ionospheric parameters to a smoothed solar index, indeed nothing has changed. But the International Reference Ionosphere (IRI) is updated every several years with data that better characterizes the ionosphere. So we still use monthly median ionospheric values in the model, but the most recent IRI data is more representative of the true characteristics of the ionosphere.

Looking to the future, efforts are underway to better understand the day-to-day variability of the F2 region. Solar radiation indeed instigates ionization, but geomagnetic field activity and events in the lower atmosphere coupling up to the ionosphere also contribute to the ionization at any given point on Earth. One of these days we may have a true daily model of the ionosphere using a daily solar index

To wrap up this month's Monthly Feature (note F), Figure 9 is a timeline of the critical dates in the development of the present model of the ionosphere.

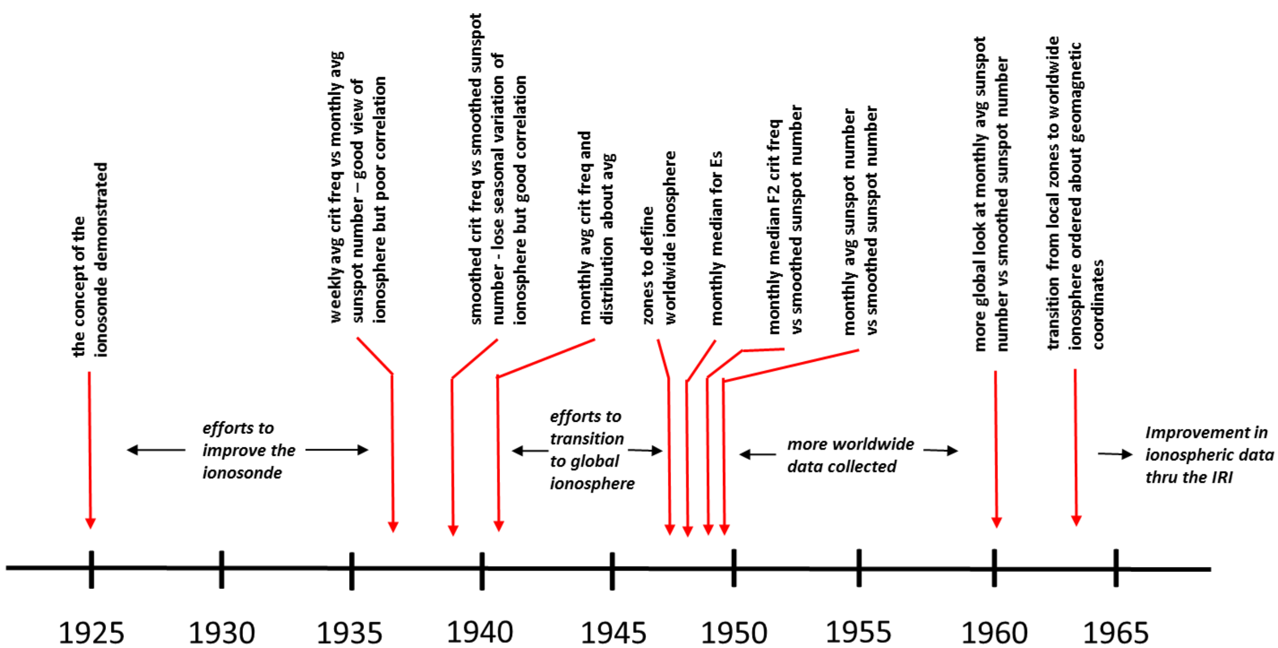

Figure 9 – Timeline of the Development of the Model of the Ionosphere

References

1. Elbert B. Judson, *Comparison of Data on the Ionosphere, Sunspots, and Terrestrial Magnetism*, **Proceedings of the Institute of Radio Engineers**, Volume 25, Number 1, Part 1, January 1937.

2. Newburn Smith, Theodore R. Gilliland, and Samuel S. Kirby, *Trends of Characteristics of the Ionosphere for Half a Sunspot Cycle*, **Journal of Research of the National Bureau of Standards**, Research Paper RP1159, Volume 21, December 1938.

3. T.R. Gilliland, S.S. Kirby, and N. Smith, *Characteristics of the Ionosphere at Washington D.C., May 1940, with Predictions for August 1940*, **Proceedings of the I.R.E**, July 1940.

4. K.W. Tremellen and J.W. Cox, *The Influence of Wave-Propagation on the Planning of Short-Wave Communication*, **Radiocommunication Convention**, January 1947.

5. Central Radio Propagation Laboratory, *Basic Radio Propagation Predictions for December 1947*, **CRPL Series D**, Number 37, September 1947.

6. Kenneth A. Norton, *Sunspots and Very-High-Frequency Radio Transmission*, **QST**, December 1947.

7. Multiple authors, *Ionospheric Radio Propagation*, **National Bureau of Standards Circular 462**, June 1948.

8. Edwin L. Crow and Donald H. Zacharisen, *The Error in Prediction of F2 Maximum Usable Frequencies by World Maps Based on Sunspot Number*, **Statistical Methods in Radio Wave Propagation**, Sec II, Radio Propagation Phenomology, Pergamon Press, New York, NY 1960.

9. *Ionospheric Predictions for April 1963*, National Bureau of Standards, Number 1, January 1963.

Notes

A. In addition to Kennelly and Heaviside, other names are tied to who came up with the idea of an electrically conducting region in the ionosphere. For example, in the October 24, 1925 issue of Nature, the authors said that about the same time when Kennelly and Heaviside came up with their idea, H. Poincaré, A. Blondel and C.E. Guillaume made similar hypotheses. Also, NQ6Z sent me a paper in the December 1925 issue of Scientific American saying that a Professor Schuster in England hypothesized this in 1887. I guess it's good that we call it the ionosphere rather than the Schuster-Kennelly-Heaviside-Poincaré-Blondel-Guillaume layer!

B. The only solar data we had back in the 1930s was sunspot data. The measurement of 10.7 cm (2800 MHz) solar flux did not begin until 1947.

C. The calculation of the twelve-month running average sunspot number was not explained in the paper. But it is safe to assume that it is akin to, if not exactly equal to, what we now call the smoothed sunspot number.

D. I've read that the distribution of the critical frequency is more like a log-normal distribution. I've also read it described as a chi-squared distribution.

E. The smoothed sunspot number and the smoothed 10.7 cm solar flux are interchangeable for prediction purposes. Use either one in your favorite program.

F. I would be remiss if I didn't mention that throughout all the referenced documents, F2 region critical frequency data was either presented for the extraordinary wave (early data) or for the ordinary wave (later data). The difference in the ordinary wave critical frequency and the extraordinary wave critical frequency depends on the latitude of the observing station. At the magnetic equator the difference is zero. At extreme high latitudes the difference is the electron gyro-frequency (generally in the neighborhood of 1.4 MHz). At mid-latitudes the difference is about half the gyro-frequency (around 0.7 MHz). When oblique propagation at low elevation angles is considered, the two oblique MUFs converge. So it appears to be a moot point with which one we use. For the record, we now generally use ordinary wave data.

February 2017 Update

On February 18, 2017 I received an e-mail from Bill NQ6Z. While looking for information on radar cove rage, he discovered the document "Report of Chief of IRPL" from May 1944. IRPL was an acronym for the Interservice Radio Propagation Laboratory during the World War II years. This group studied HF propagation due to the importance of understanding the ionosphere for long distance communications.

One of the paragraphs in this report stated that a study was being made of the use of the monthly median MUF instead of the monthly mean MUF. It was stated that first conclusions are that the median is useful on data from stations where the quantities to be scaled are sometimes outside the limits of the recorder. Thus there is a statistical reason to use the median, and not the mean – and the median is what is used now.

This old IRPL document is courtesy of the Carnegie Institution of Washington Department of Terrestrial Magnetism Archives.