

A Review of Statistical Techniques used for Ionospheric Model Validation

Sean Elvidge, Matthew J. Angling

Space Environment and Radio Engineering, University of Birmingham, UK

There are a wide range of ionospheric models in active use and development around the globe, utilizing a number of modelling techniques. To gain an insight into the success of the various model types at specifying the ionosphere, two things are required: (1) metric(s) to measure the success criterion of the models and (2) common testing scenarios.

A variety of metrics are used to validate and compare ionospheric models. It is difficult to quantify the overall success criterion for a model, and the metrics are often applied to specific parameters instead. These metrics regularly depend directly upon certain statistics or the representation of those statistics, such as the mean, standard deviation or correlation. Commonly the statistics are calculated from a time series of the parameter. For example, for a given test study, the foF2 is recorded every 15 minutes. Then the mean and standard deviation of foF2 can be found, and using truth data, the correlation and error statistics can be found.

In this paper the underlying assumptions of common statistical techniques used for ionospheric model validation are reviewed. If the initial assumptions are invalid, then the conclusions cannot be trusted. The first section of this paper reviews the stationarity of time series' and the use of confidence intervals. A stationary time series is one in which its statistics, such as the mean and variance, do not change over time. The mean, standard deviation and correlation of a non-stationary time series need to be worked with carefully to ensure valid conclusions are drawn. This paper will review the techniques and assumptions of these oft used statistical parameters.

The second part of the paper will discuss the need for common testing scenarios. For fair model comparison the same test scenario should be used for all of the models. A good test scenario needs to be of sufficient length to reduce confidence intervals, have high quality data with a few gaps and independent observations for verification. As well as common scenarios, how the model output is used and processed (such as the removal of trends to make the time series stationary) for the comparison needs to be consistent. Such a testing scenario is suggested in the paper, and the data made publically available.