How Well Do We Forecast the Aurora?

Michaela K. Mooney[1]
Colin Forsyth[1], Mike Marsh[2], Michael Sharpe[2], Teresa Hughes[2], Suzy Bingham[2] and David Jackson[2]
[1]Mullard Space Science Laboratory, University College London, Dorking, UK
[2]Met Office, Exeter, UK

From the Sun to the Aurora

From Earth, the Sun may seem a to be quiet and distant neighbour. However, the Sun is constantly accelerating a stream of charged plasma particles out into the solar system. This plasma travels radially away from the Sun carrying with it the embedded solar magnetic field, known as the interplanetary magnetic field (IMF). For most of us on Earth, we are blissfully unaware of this constant solar wind, thanks to Earth's protective magnetic field. The Earth's magnetic field expands out into space and forms a region known as the magnetosphere and shields the Earth from this constant bombardment of charged particles. Despite the name, the magnetosphere is not spherical but has a shape distorted by its interaction with the solar wind. The pressure of the incoming solar wind compresses the magnetic field on the Sun-facing side of the Earth and stretches out a long magnetotail on the nightside. When the solar wind encounters the Earth's magnetic field it is mostly deflected however, the two magnetic fields can interact with each other through a process known as *magnetic reconnection*. During reconnection, energy and plasma from the solar wind can be transferred into the Earth's magnetic field.

Charged plasma particles which are trapped in the magnetosphere are funnelled down the magnetic field lines around the North and South poles until they reach the partially ionised region of Earth's upper atmosphere known as the ionosphere. Here, the incoming particles collide with other partially ionised or neutral particles, releasing a shower of lower energy electrons. These collide with further ions and neutral atoms until the incoming electrons have insufficient energy to ionise the atmospheric particles, instead exciting the bound electrons which, as they de-energise, emit light that we see as the aurora. The auroral emission forms a ring around the magnetic poles in both the northern and southern hemispheres, known as the auroral oval.

The arrival of the plasma particles in the ionosphere varies the conductivity of the ionosphere which in turn affects the propagation of long-range radio communications. The location and intensity of the auroral oval is therefore of interest to many industrial sectors such as aviation and defence where continual radio communication with ground stations is imperative. The constant driving of near-Earth space by internal and external drivers of the magnetosphere maps down to an auroral oval which is constantly changing in shape and intensity in response to this activity. This makes the auroral oval a particularly challenging area of space weather to model and forecast. The OVATON-Prime 2013 (OP-2013; Newell et al., 2014) auroral forecast model aims to do just that and has been implemented in the operations centre of both the Space Weather Prediction Centre (SWPC) in the U.S. and the Met Office Space Weather Operations Centre (MOSWOC), in the U.K. The operational version of OP-2013 at MOSWOC produces a forecast of the auroral oval 30 minutes ahead based on the upstream solar wind conditions.

To understand how useful a model is and to determine where further development is required, it is important to validate the model against observations, in particular using datasets not used in the original model development. Validation studies provide a quantitative analysis of the model performance which can be used to assess the benefit of new upgrades to the model and also allow a fair comparison between similar models.

Previous validation studies of the predecessor model, OVATION-Prime (OP) (Newell et al., 2002) compared the forecasts from OP with data from the ultra-violet imager (UVI) onboard the Polar satellite (Newell et al., 2010; Machol et al., 2012). Newell et al. (2010) found that the hourly averaged total hemispheric power predicted by the OP model correlated with that calculated from the Polar UVI observations with a Pearson's correlation coefficient of 56%. While correlation coefficients may provide a general summary of how model predictions and observations compare, they don't provide any information about the strengths or weaknesses of the model.

More recently, the space weather community has been looking towards terrestrial weather forecast validation to provide appropriate methods and techniques to evaluate space weather models (e.g. Machol et al., 2012; Murray et al., 2017 and Sharpe et al., 2017). Machol et al. (2012) use binary event analysis to compare the OP forecasts and Polar UVI observations. From the comparisons of each corresponding OP forecast - UVI observation pair, they build up a contingency table of the number of true positive forecasts (where the aurora is forecast and latterly observed), false alarms (the aurora is forecast but does not occur in the observations), missed forecasts (aurora is not forecast by the model but subsequently observed) and correct rejections (the aurora is not forecast and subsequently is not observed). An example contingency table is shown in Table 1. From this analysis, Machol et al. (2012) found that the OP model correctly forecast the nightside aurora with an overall accuracy (the number of correct forecasts out of the total number of forecasts) of 77% and a false alarm rate of 14%.

While the analysis of Machol et al. (2012) pushes the validation of the OP model further than previous studies, there are additional techniques from terrestrial weather forecast validation that we can apply to space weather forecasts. In this study, we apply terrestrial weather forecast validation techniques to quantify the performance of the OP 2013 model against a set of satellite observations of the auroral oval.

Table 1
An example contingency table which results from comparing model forecasts and corresponding observations using binary event analysis.

Forecasts	Observations	
	Yes	**No**
Yes	Correct Positives	False Alarms
No	False Negatives	Correct Negatives

Auroral Observations

One of the main difficulties for space weather model validation is having a robust dataset of observations to compare with the model forecasts which span a long enough time frame for a fair comparison with the model. The IMAGE satellite was a polar orbiting satellite which was in operation between 2000-2005. Between 2000-2002, the IMAGE satellite was in a prime position to make observations of the northern hemisphere auroral oval. On-board the satellite, IMAGE had far-ultraviolet (FUV) imaging cameras which took observations of the entire northern auroral oval every 2 minutes. From this data, Longden et al. (2010) systematically identified the inner (poleward) and outer (equatorward) boundaries of the northern auroral oval in each local time sector. In our validation study, we use these defined boundaries as our *ground truth* data to compare with the auroral emission forecast by the OP-2013 model.
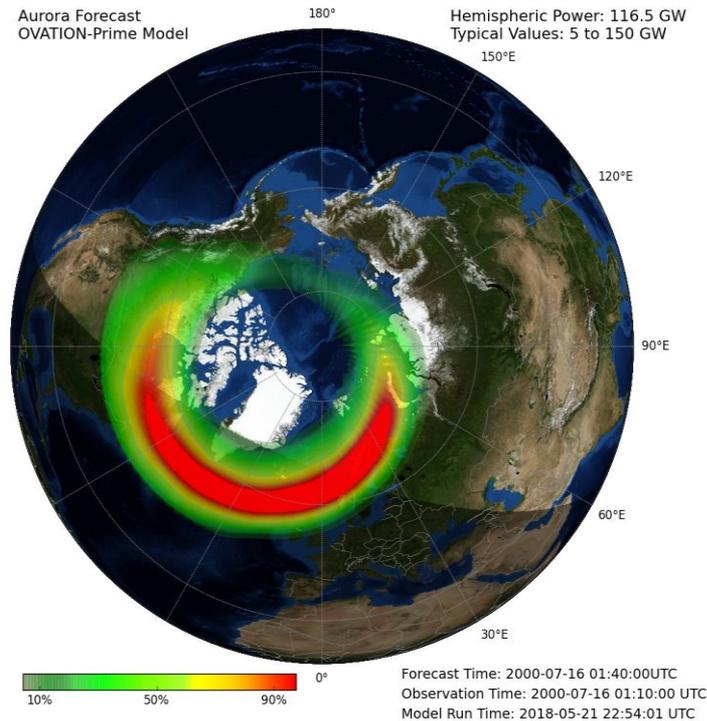
The OP-2013 Model
The operational version of the OP-2013 model uses the conditions of the incoming solar wind measured by satellites situated between the Sun and the Earth. The model also contains averaged maps of the predicted auroral emission based on observations of the incoming plasma particles by the DMSP spacecraft. These average auroral emission maps are scaled linearly depending on the incoming solar wind conditions to forecast the auroral emission 30 minutes in advance. The full details of the model can be found in Newell et al. (2002, 2014). In the operational version of the model, the forecast auroral emission is converted into a probability of the aurora occurring.

Figure 1 shows an example of the auroral forecast from the OP-2013 model during the Bastille Day Storm in July 2000 for the northern hemisphere. The red/green colour scale indicates the probability of aurora occurring with green being a lower probability and red being higher. The OP-2013 forecast output is displayed in geographical coordinates; however, we perform our validation in magnetic coordinates. For the validation study, we ran the OP-2013 model using historic solar wind data to produce 30-minute auroral forecasts.

Figure 1:
An example output from the OP-2013 model during the Bastille Day Strom in July 2000. The model output is shown in geographical coordinates, looking down on the northern polar region. The red/green auroral oval shows the predicted location of the aurora, with the colours corresponding to a probability of the aurora occurring in each location. The probabilities increase from green to red as shown in the colour bar.
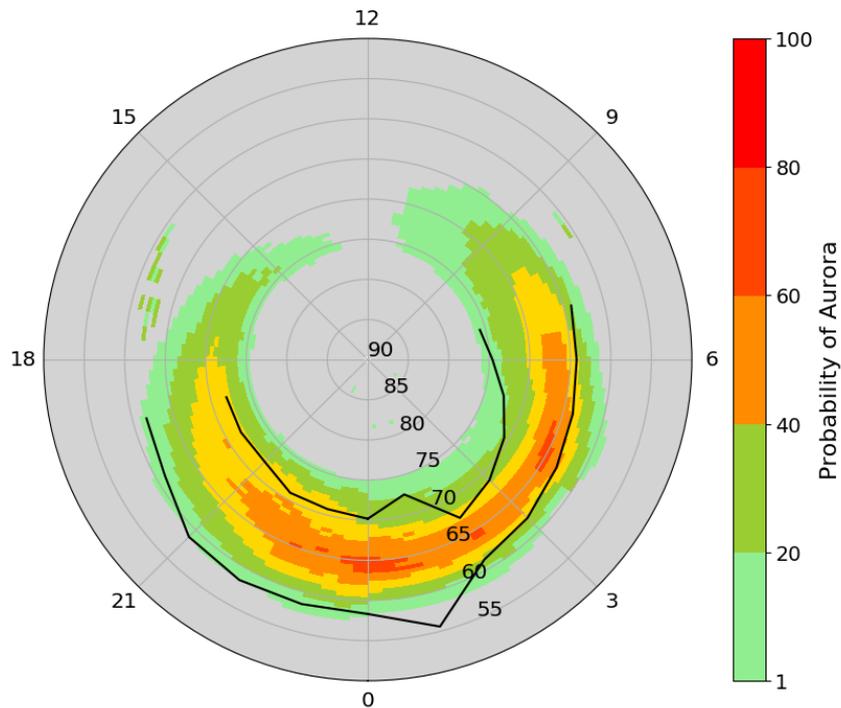
Aurora Forecast
OVATION-Prime Model

Hemispheric Power: 116.5 GW
Typical Values: 5 to 150 GW

10% 50% 90%

Forecast Time: 2000-07-16 01:40:00UTC
Observation Time: 2000-07-16 01:10:00 UTC
Model Run Time: 2018-05-21 22:54:01 UTC

Comparing Forecasts and Observations

Figure 2 shows an example comparison between the auroral forecast from OP-2013 and the auroral boundaries derived by Longden et al. (2010) displayed in the magnetic coordinate system used in our validation. The probabilities of aurora occurring forecast from the OP-2013 model are again shown by the green to red colour scale, showing lower to higher probabilities, respectively. The black lines show the auroral boundaries determined by Longden et al. (2010) from the IMAGE FUV data. Although the automated auroral boundary identification method of Longden et al. (2010) attempts to define the auroral boundaries in all parts of the auroral oval, the auroral emission on the dayside tends to be dimmer and has more interference from sunlight which can make it more difficult to correctly identify the dayside auroral boundaries in some cases. The OP-2013 model is only validated in the local times where there are successfully identified observed auroral boundaries.

Figure 2:
An example of a model forecast-observation pair from October 2002, comparing the OP-2013 model forecast with the probabilities shown in colour from green to red. The black lines indicate the poleward and equatorward boundaries of the observed auroral oval for the same time.

OVATION-Prime 2013 vs Auroral Boundaries Derived from WIC Data
2002-10-29 10:01:13

For each corresponding forecast and observation pair we compare whether the aurora was forecast and subsequently observed at each point on the grid. We repeat this process for all available observation data over the 2.5 year period between May 2000 and October 2002 to build up our contingency table.
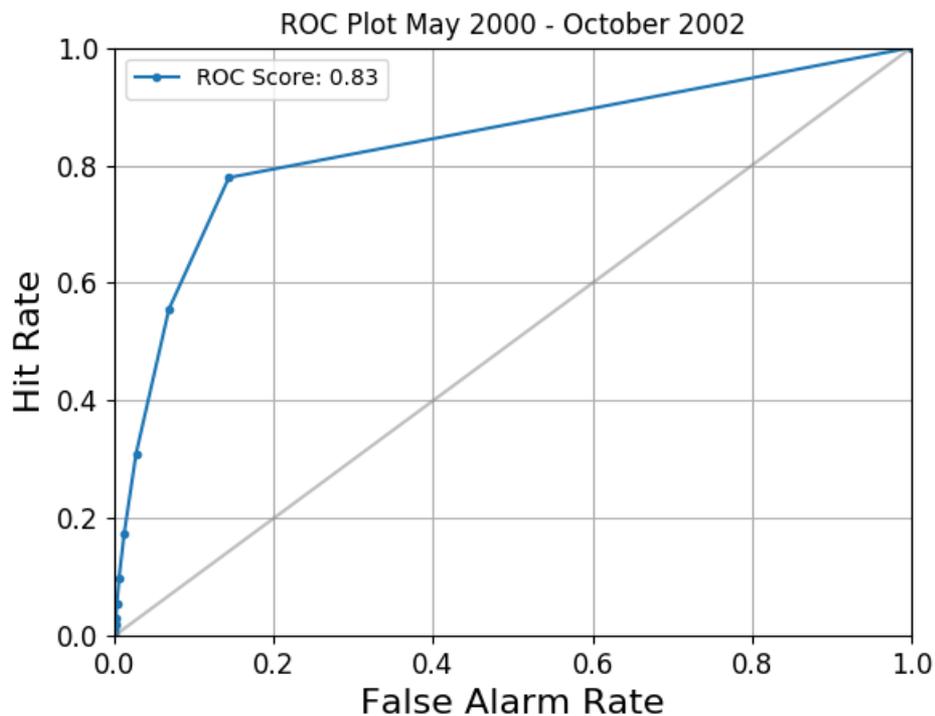
From the contingency table, we can derive the hit rate (the proportion of true positive forecasts out of the total number of times that the aurora was forecast to occur) and false alarm rate (the proportion of incorrect forecasts that aurora would occur which were not subsequently observed) of the model at each probability threshold in 10% increments. The false alarm rate and hit rate are then used to produce a relative operating characteristic (ROC) curve of the model performance. ROC curves provide information on the model discrimination, that is, how well the model distinguishes between regions of aurora and no aurora. The second analysis tool we use is a reliability curve which provides information about the reliability of the probabilities of aurora occurring forecast by the model.

Model Discrimination

To test the model's overall discrimination, we performed a ROC analysis on the 2.5 years of forecast and observation pairs. Figure 3 shows the results of this analysis. The ROC curve indicates that at each 10% probability threshold the model hit rate is larger than the false alarm rate showing that the OP-2013 model is skilful at distinguishing between regions of aurora and no aurora. The ROC score provides a summary value of the model ROC performance and has a value between 0-1 with 1 being the best possible score. The OP-2013 model has a ROC score of 0.83 confirming that the model has high discrimination.
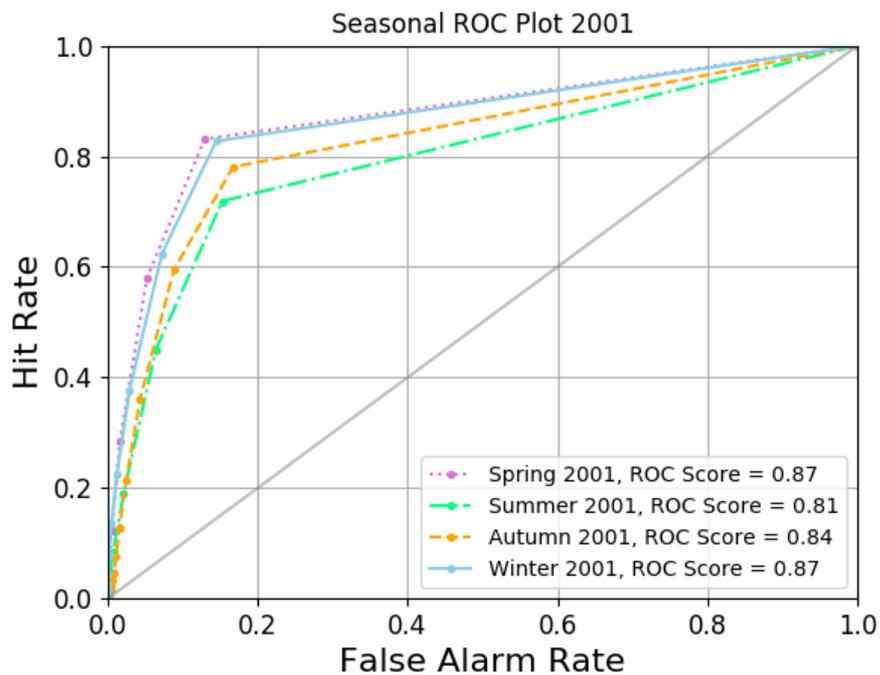
Figure 3:

The results of the ROC analysis for the full 2.5 year observational data set between May 2000 and October 2002. At each probability threshold, the OP-2013 model has a higher hit rate than false alarm rate indicating high discrimination in the model. The high ROC score of 0.83 confirms that the model is good at distinguishing between regions of aurora and no aurora.

We also tested the model for seasonal variability in performance during the four seasons of 2001. Figure 4 shows the ROC analysis of the OP-2013 model for each season in 2001 with the results for spring shown by the dotted pink line, summer shown by a dot-dashed green line, autumn shown by a dashed orange line and winter shown by a solid blue line. The ROC scores for each season vary between 0.81 for summer and 0.87 for spring and winter. Although there is a small variation in the ROC scores between the four seasons, all the ROC scores are above 0.80 indicating that the OP-2013 model performs consistently well all year round.

Figure 4:
The results of the ROC analysis for each season in 2001 with the results for spring shown by the dotted pink line, summer shown by a dot-dashed green line, autumn shown by a dashed orange line and winter shown by a solid blue line. The high ROC scores for each season demonstrate that the OP-2013 model performs well all year round.

Seasonal ROC Plot 2001

Legend:
- Spring 2001, ROC Score = 0.87
- Summer 2001, ROC Score = 0.81
- Autumn 2001, ROC Score = 0.84
- Winter 2001, ROC Score = 0.87

The Reliability of Forecast Probabilities of Aurora Occurring

To evaluate overall reliability of the forecast probabilities from OP-2013 we produced a reliability diagram for the observation and forecast pairs over the full 2.5 year period between May 2000 and October 2002, shown in Figure 5. The grey curve in Figure 5 indicates the line of perfect reliability where the model forecast probabilities have a one-to-one correspondence with the observed frequency of aurora. The reliability curve for the OP-2013 model lies above this perfect reliability line. This indicates that the OP-2013 model consistently under-estimates the probability of aurora occurring at each probability interval i.e. the aurora occurs more frequently than the model predicts. The difference between the forecast probabilities and the observed occurrence of aurora is larger at lower probability values of 0.6 and less. At the highest probability values of 0.7-0.9, the reliability curve tends towards the perfect reliability line. The highest probability values of 0.9-1.0 show a slight tendency towards over-prediction, lying below the perfect reliability line.
Similarly, in Figure 6 showing the reliability curves for each season in 2001, the probabilities forecast by the OP-2013 model are consistently under-predicted, particularly between probability values between 0.1-0.6.

Figure 5
The result of the reliability analysis for the full 2.5 year observational data set between May 2000 and October 2002.
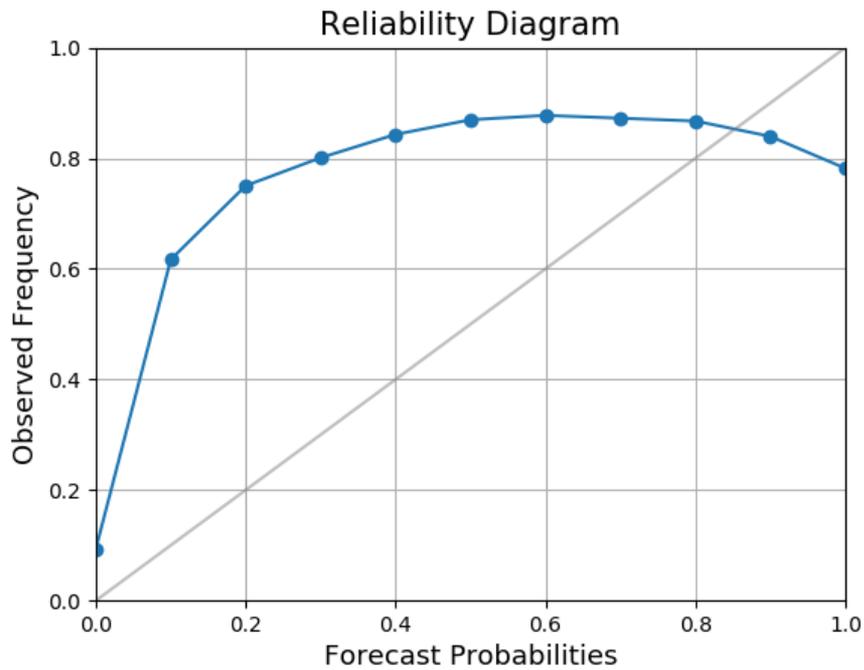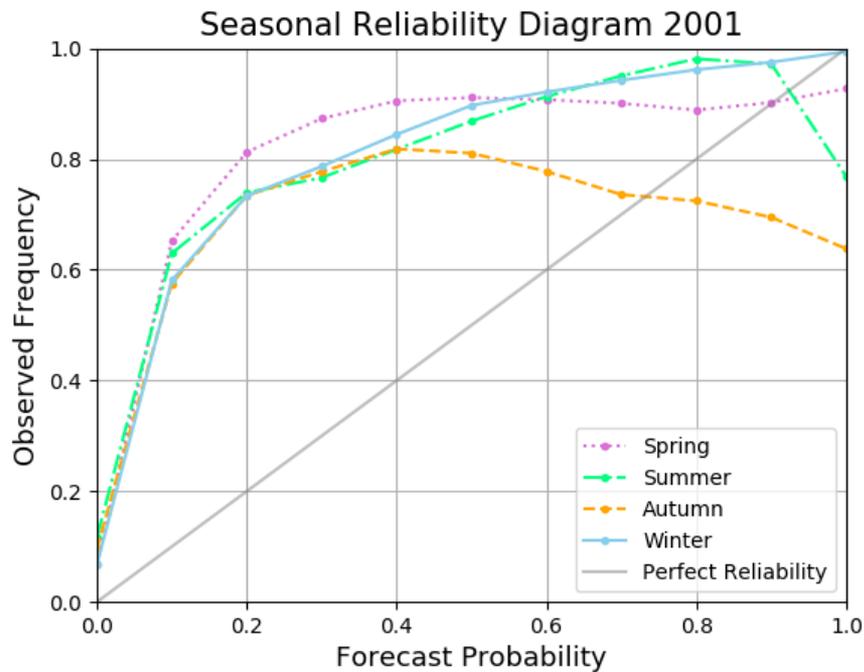
Figure 6
The result of the reliability analysis for each season in 2001. Again, the result of the spring analysis is shown by the dotted pink line, summer shown by a dot-dashed green line, autumn shown by a dashed orange line and winter shown by a solid blue line.

Conclusions

By applying techniques which are routinely used in terrestrial weather model validation, we can evaluate the performance of the operational version of the OP-2013 auroral forecast model which is currently used in daily auroral forecasts at the Met Office. Our results indicate that the OP-2013 model performs well at distinguishing between regions of aurora and no aurora, with consistently high ROC scores above 0.80 for the validation period. However, OP-2013 largely under-predicts the probabilities of aurora occurring. Based on our analysis, it should be possible to recalibrate these probabilities.

This study shows that applying this type of analysis to a space weather model is an effective way to measure the operational performance of a model and could be applied to a wide range of space weather models. Future work will aim to evaluate the performance of the OP-2013 model during geomagnetically active periods such as during geomagnetic storms and substorms.

References

Longden, N., Chisham, G., Freeman, M. P., Abel, G. A., and Sotirelis, T. (2010), Estimating the location of the open-closed magnetic field line boundary from auroral images, Ann. Geophys., 29, 1659-1678, doi: 10.5194/angeo-28-1659-2010.

Machol, J. L., J. C. Green, R. J. Redmon, R. A. Viereck, and P. T. Newell (2012), Evaluation of OVATION Prime as a forecast model for visible aurorae, Space Weather, 10, S03005, doi:10.1029/2011SW000746.

Murray, S. A., Bingham, S., Sharpe, M., and Jackson, D. R., (2017), Flare forecasting at the Met Office Space Weather Operations Centre, Space Weather, 15, 577–588, doi:10.1002/2016SW001579.

Newell, P. T., T. Sotirelis, J. F. Carbary, K. Liou, J. P. Skura, C.-I. Meng, C. Deehr, D. Wilkinson, and F. J. Rich (2002), OVATION: Oval Variation Assessment, Tracking, Intensity, and Online Nowcasting, Ann. Geophys., 20, 1039, doi:10.5194/angeo-20-1039-2002

Newell, P. T., T. Sotirelis, K. Liou, A. R. Lee, S. Wing, J. Green, and R. Redmon (2010), Predictive ability of four auroral precipitation models as evaluated using Polar UVI global images, Space Weather, 8, S12004, doi:10.1029/2010SW000604.

Newell, P. T., K. Liou, Y. Zhang, T. Sotirelis, L. J. Paxton, and E. J. Mitchell (2014), OVATION Prime-2013: Extension of auroral precipitation model to higher disturbance levels, Space Weather, 12, 368–379, doi:10.1002/2014SW001056.

Sharpe, M. A., & Murray, S. A. (2017), Verification of space weather forecasts issued by the Met Office Space Weather Operations Centre. Space Weather, 15, 1383–1395. doi: 10.1002/2017SW001683