Substorm Onset Prediction using Machine Learning Classified Auroral Images

Pascal Sado[1], Lasse Boy Novock Clausen[1], Wojciech Jacek Miloch[1], and Hannes Nickisch[2]

[1]University of Oslo
[2]Philips Research

November 22, 2022

Abstract

We classify all sky images from 4 seasons, transform the classified information into time-series data to include information about the evolution of images and combine these with information on the onset of geomagnetic substorms. We train a lightweight classifier on this dataset to predict the onset of substorms within a 15 minute interval after being shown information of 30 minutes of aurora. The best classifier achieves a balanced accuracy of 61% with a recall rate of 47% and false positive rate of 24%. We show that the classifier is limited by the strong imbalance in the dataset of approximately 50:1 between negative and positive events. All software and results are open source and made freely available.

Substorm Onset Prediction using Machine Learning Classified Auroral Images

P. Sado[1], L. B. N. Clausen[1], W. J. Miloch[1], H. Nickisch[2]

[1]Department of Physics, University of Oslo, Oslo, Norway
[2]Philips Research, Hamburg, Germany

Key Points:

- Auroral images are classified, then time series information is introduced and noise removed by means of a Hidden Markov Model
- A linear classifier predicts the onset of substorms within 15 minutes of seeing 30 minutes of images
- The best classifier recalls 47% of substorms with 61% balanced accuracy and 24% false positive rate

Corresponding author: Pascal Sado, `Pascal.Sado@fys.uio.no`

Abstract

We classify all sky images from 4 seasons, transform the classified information into time-series data to include information about the evolution of images and combine these with information on the onset of geomagnetic substorms. We train a lightweight classifier on this dataset to predict the onset of substorms within a 15 minute interval after being shown information of 30 minutes of aurora. The best classifier achieves a balanced accuracy of 61% with a recall rate of 47% and false positive rate of 24%. We show that the classifier is limited by the strong imbalance in the dataset of approximately 50:1 between negative and positive events. All software and results are open source and made freely available.

Plain Language Summary

When charged particle originating from the sun travel into near earth space, they interact with the Earth's natural magnetic field. These interactions are what leads to the aurora, but can also cause problems with electric installations or satellite communications. Knowing when and where these can occur can be used to mitigate effects. Such forecasts are also beneficial for research, as rockets could be launched into regions of interest or paths of satellites can be adjusted to arrive at the same time as the occurrence of such events. Our model takes images from ground based cameras to predict the onset of strong space weather occurrences.

1 Introduction

Solar wind is the driving force of space weather on earth. Energy can be stored in the Earth's magnetosphere and will subsequently be released. These so called substorms are not only cause for the spectacle we know as the aurora, but have also the potential to cause serious harm to modern technology. Especially nowadays society reliance on digital communication delivered by satellites has made this a major concern in the last few decades.

Heating and expansion of the atmosphere by the aurora can lead to an increase in drag on satellites, possibly reducing lifespan, warranting course correction or at the very least cause observations of the changed course to avoid collisions (Marcos et al., 2010). Geomagnetically induced currents can affect man-made electrically conducting structures such as the power-grid, under-sea communication cables or pipelines, causing disruption in various services (Pirjola, 2000). GNSS systems can provide exact timing and location services, based on the distance to the satellite calculated from the known position and travel time of the signal to a ground based receiver. However, ionospheric disturbances can change the travel time by several nanoseconds or few microseconds, giving errors in the position by a few meters (Kintner et al., 2007).

Although there is the potential for global events to occur, these are extremely rare and localised events are much more likely. In order to mitigate the risks, it is important to know when and where they will occur.

Originally based on images (Akasofu, 1964; Akasofu et al., 1965) the study of substorms has moved on to satellite-supported studies (McPherron et al., 1973), giving us the currently used model of substorms. The solar wind has long since been identified as the main driving force behind substorms and substorm triggers (Caan et al., 1975). During the growth phase of substorms, energy is stored in the Earth's magnetosphere. This energy is released during the expansion phase and the magnetosphere subsequently returns to its quiet state in the recovery phase of a substorm.

Different phases during a substorm can trigger different mechanism of energy-release which will in turn have different outcomes on the visible aurora (P. T. Newell et al., 2010; Akasofu, 2013; Partamies et al., 2015).

In its simplest form during quiet times, aurora are visible in the shape of a single east-west arc, become larger and brighter, expand poleward during a substorm and form rapidly westward travelling folds, before breaking up into smaller structures, becoming more chaotic and returning to their quiet state again towards the end of a substorm(Akasofu, 1964).

No matter the origin of substorms, their footprint on earth stays the same and subsequent identification can be performed either visually through all sky or satellite images of aurora or measurements of the earth's magnetic field. Visual identification as performed for example by Frey et al. (2004) and Liou (2010) is still based on the definition by Akasofu (1964) consisting of sudden brightening of the aurora followed by poleward motion and increase in intensity of the aurora. Forsyth et al. (2015); P. T. Newell and Gjerloev (2011); Ohtani and Gjerloev (2020) use instrument based identification of substorms, where they used the change in Earth's magnetic field.

The lists of substorms originating from this work have found widespread use in the community for prediction of all kinds of space weather(cf. `https://supermag.jhuapl.edu/publications/`), including the prediction of substorm onsets by Maimaiti et al. (2019) using deep neural networks. With their model the authors also confirmed the importance of the B_z component of the interplanetary magnetic field (IMF) (P. T. Newell & Liou, 2011) and the solar wind speed (P. Newell et al., 2016) on the occurrence of substorms. Their work shows how well solar wind data can be used to forecast onsets of substorms on a global level. Furthermore, Sado et al. (2022) have shown that all sky images contain sufficient information that can be extracted by a neural network and be used to model the behaviour of the Earth's local magnetic field in vicinity to the imager.

Taking the same approach, in this study we obtained approximately 4 million all sky imager data, classified the images and used a time series of images representing half an hour of data to predict the onset of substorms within the next 15 minutes after the time series.

We manage to achieve a balanced accuracy of 61% at a recall rate of 46% and false positive rate of 24%. We show that the classifier often correctly identifies to occurrence of an event, but fails to pinpoint the exact location in time and therefore either misses or overshoots the target prediction. The classifier itself was as lightweight as possible and made it therefore necessary to reduce the input information for training to it's bare essentials. Given its success we estimate that training a more sophisticate model on the raw input data will lead to drastic improvements of this method.

In section 2 we give an overview of which data we use and in section 3 we detail our preprocessing steps for the images and substorm data. Finally in section 4 we present our results and give a summary and outlook in section 5.

2 Description of Data Sources

In this project, we use data from two different sources. Our Images are taken from the THEMIS All Sky Imager array's camera in Gillam, Manitoba. The All Sky Camera takes images every 3 s at a resolution of 256 px by 256 px. The images are taken in the 2009/2010 and 2010/2011 seasons corresponding to conditions of solar minimum and and in the 2014/2015 and 2015/2016 seasons for solar maximum. This gives us a total of approximately 3.7 million images taken over 4 years. The images were taken with a fisheye lense giving a full view of the sky from horizon to horizon. To remove artefacts like trees just above the horizon, a ring 20 px wide was removed. The images were then preprocessed before being classified according to the method developed by Sado et al. (2022).

The images are complemented with physical data in the form of substorm occurrences based on the SuperMAG list of substorms. These were created by Forsyth et al. (2015) using the SOPHIE technique, where substorm expansion and growth phases are identified by finding extrema in the derivatives of the SML (Auroral Electrojet Index) and by Ohtani and Gjerloev (2020) who based their identification on the local develop-

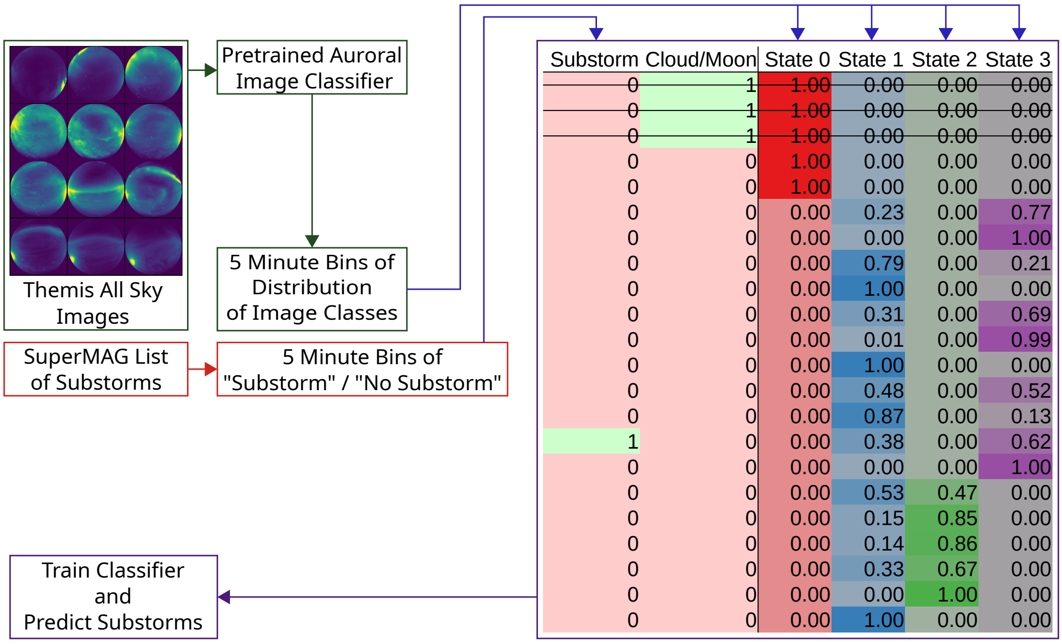


Figure 1: Outline of the workflow. Auroral Images are classified with an established classifier, time information is added by smoothing with a Hidden Markov Model and images are binned into 5 minute intervals with their respective predicted states. Based on the classification's result images with clouds or the moon are removed. To these bins, information about whether a substorm has occurred during the interval is added from the SuperMAG list and finally a classifier is trained to predict whether a substorm will occur after a given interval of images.

ment of the Earth's magnetic field as it is influenced by a substorm. These occurrences are listed by date and time with their respective location. Only substorms occurring within a latitudinal and longitudinal distance of 10° to the camera were used, resulting in approximately 2000 events. Out of these, 245 have been recorded at a time with image coverage.

3 Methods

3.1 Overview of Dataflow

Figure 1 shows an overview of how the data flows through the process. The all sky Images are preprocessed and classified according to the classifier by Sado et al. (2022). This process is detailed in section 3.2. Those images not showing aurora or a clear night sky are removed. Furthermore, we use a Hidden Markov Model for smoothing the time series of images. We show this in detail in section 3.3. The classified images are condensed into bins containing 5 min of distribution of image classes. During periods of full camera coverage, a 5-minute-bin will contain 100 images. However, since coverage is not perfect or images have been removed because they were not relevant, bins might contain less images. Each bin is then assigned a Boolean value based on whether a substorm has occurred during this time or not according to the SuperMAG list.

The processing of substorm data and details on the classifier can be found in sections 3.4 and 3.5 respectively.

3.2 Image Preprocessing

Individual images are classified based on the classifier developed and demonstrated by (Sado et al., 2022; Clausen & Nickisch, 2018). In this process, the images are analysed by a pretrained neural network and the image features as defined by this network are extracted. A classifier that has been trained on a labelled set of images that have undergone the same process of feature-extraction is then used to classify the images. This returns a probability for each image to be in either of the following six classes:

- **arc** The image shows mostly a single auroral arc spanning from east to west (left to right in the frame of the image)
- **diffuse** The image shows diffuse aurora without any clear structure
- **discrete** Discrete aurora show structure but not in the form of well-defined arcs. The structures can be of any other shape.
- **cloud** The image shows clouds
- **moon** The image shows the moon
- **clear** The image shows a clear night sky

Images where the probability to show the moon is above 40% or the probability to show clouds is above 70% are discarded and will not be used from now on. These probabilities are removed from the distribution of classes for the remaining images and the probabilities rescaled to 100%.

3.3 Hidden Markov Model for Time Series Information

Individual images can contain false information and noise within the data. A person walking by the imager holding a flashlight and illuminating the dome of the camera can cause misclassification for individual images. The images however belong to a time series of images where images are taken 3 s apart. The change from one individual image to another should therefore be small as should be the probability assigned to two consecutive images. This can be used to smooth the distribution of probabilities and classes for individual images along a longer time series of images.

To do this, we have employed a Hidden Markov Model (HMM) a widely used smoothing model that can be adapted to data. The observable state of a HMM is based on the state of a hidden variable. The transition of the hidden variable from one state to another is based on a Markov process, giving the model its name Rabiner (1989).

For our purposes we assume that our observed probabilities for each image are based on a set number of hidden states. For each image, the hidden state can transition from one to another giving different probabilities for the output of each image.

We tested various amounts of hidden states between 2 and 100 and judged these by two metrics:

- Transition Matrix Sparsity
- State Distribution Entropy

The transition matrix sparsity measures how the transition of states are distributed. The higher this value, the more non-zero elements are present in the transition matrix. The more possible transitions between different states there are, the higher this value will be. If the model only transitions between selected states or does not populate some states at all, this metric will be smaller. The Sparsity is calculated as $\frac{\text{Count of non-zero Elements in the Matrix}}{\text{Total Elements in the Matrix}}$. State distribution entropy measures how evenly the states are populated. The Entropy is defined as $S(P) = -\frac{1}{N}\sum_{i=0}^{N} p_i log(p_i)$ where N is the total amount of states, P the distribution and p_i the probability for the i-th state. This value is lowest with $S(P) = 0$ if a single state holds all values or highest with $S(P) = log(N)$ if the states are evenly distributed. We demonstrate these values in Figure 2, where we have split the data ran-

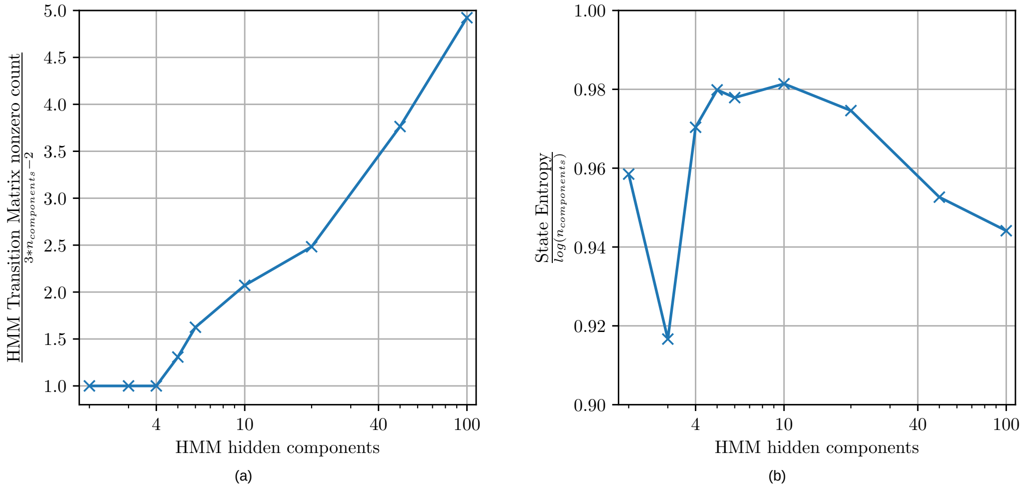


Figure 2: Sparsity of the transition matrix adjusted for expected values (a) and Entropy of state distribution adjusted by maximum entropy for each possible distribution (b)

domly into 60% training and 40% for testing. Because this is only the model-selection process and it takes a considerable amount of time to train models with a high number of hidden states, we did not employ cross-validation but opted for this rather large split of test data. The sparsity (Figure 2a) has been normalised on the expected value if only the main and first off-diagonal were to be populated, because if transitions happen only between neighbouring states, the numerator will grow linearly and the denominator quadratically The entropy (Figure 2b) has been normalised to the maximum value for the respective distribution because it increases naturally with an increasing amount of states. We can see that for up to four hidden components, the matrix is only populated along the main diagonal and its first off-diagonal. This means that transitions only happen between neighbouring states and that the HMM does not skip states when transitioning from one state to another. This behaviour is lost when using more hidden states. In terms of entropy we are looking for a model that does not neglect some states. This would be reflected in a low value for entropy as seen for three hidden components. Based on the entropy, we deem all models between four and twenty hidden components to be valid choices.

On the principle of selecting the simplest possible model that is able to perform the task, we have settled for a HMM with 4 hidden layers. This way the model does not have to infer too much information from the given probabilities or too much information is lost. The importance of its property to not skip states will also become apparent later.

The output of the final model with four hidden layers for a selected time frame is demonstrated in Figure 3. This Figure shows the probabilities for images taken by the camera on 2009-12-14 as determined by our classifier in the top panel. The bottom panel shows the states determined by the HMM. We see that although the states are not mapped back onto the predicted classes in a 1-to-1 fashion, the HMM has learned to interpret the transition of images in a physical sense. An event of auroral images beginning at approximately 04:00 starts with images classified in the zeroth state at first, then transitions to the first, further to the second and later to the third state, before going backwards through the same cycle. It is interesting to note that the HMM always transitions in this order and never skips any of the output states, i.e. transitioning from state zero to state two immediately, without going through state one.

Furthermore, we show the transition matrix of the model in Figure 4. We see that the model prefers to stay in its current state and then follows the logical progression trough

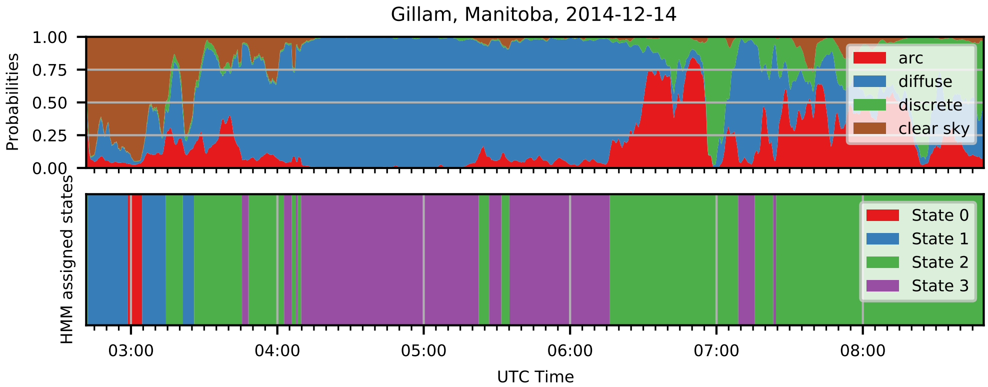
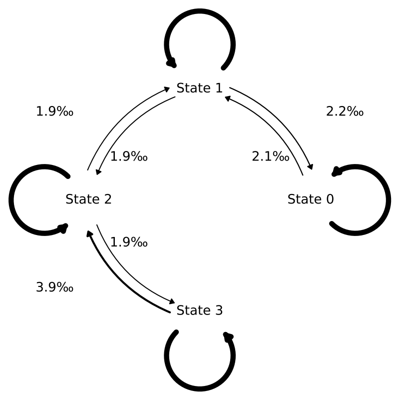


Figure 3: The original images' probabilities are transformed into 4 states using a Hideen Markov Model. This ensures time information between images and their transition from one to another is encoded into the information given to the classifier and removes noise between the images. The original images' class probabilities are shown in the top panel and their respective states assigned by the HMM are shown below.

←	State 0	State 1	State 2	State 3
State 0	997.9	2.2	0	0
State 1	2.1	995.9	1.9	0
State 2	0	1.9	996.2	3.9
State 3	0	0	1.9	996.1

(a)



(b)

Figure 4: Transition matrix of the Hidden Markov Model in per mille (a) and visualisation (b). The arrows corresponding to the off-diagonal values have been scaled by a factor of 100.

Table 1: (a) Mapping between labels assigned by the classifier and by the HMM. (b) The same as (a), but instead of the assigned classes, their probabilities have been used. This means, instead of adding 1 or 0 for "hit" or "miss", for each image, its probabilities are added to the row for the state it has been assigned by the HMM.

(a)

	Arc	Diffuse	Discrete	Clear	Total
State 0	0	0	0	1 057 963	1 057 963
State 1	0	0	0	1 046 727	1 046 727
State 2	190 952	543 830	150 155	172 092	1 057 029
State 3	0	552 354	0	0	552 354
Total	190 952	1 096 184	150 155	2 276 782	3 714 073

(b)

	Arc	Diffuse	Discrete	Clear	Total
State 0	16 150	26 258	3 588	1 011 967	1 057 963
State 1	112 841	127 480	20 968	785 438	1 046 727
State 2	225 749	425 709	198 075	207 496	1 057 029
State 3	9 761	519 496	17 748	5 349	552 354
Total	364 501	1 098 943	240 379	2 010 250	3 714 073

the different states, from a clear sky to diffuse aurora. For all states, the probability to transition into a different state is approximately 0.4%, at one image per 3 seconds, this gives an expected lifetime of a state of $750\,\mathrm{sec} = 12.5\,\mathrm{min}$. In our interpretation, the states can be roughly described such that the zeroth state describes a clear sky, the first state the beginning of aurora, the second state structured aurora and the third state diffuse aurora.

This is further supported by observing the mapping between the classes assigned by the original classifier and the hidden states assigned by the HMM as shown in Table 1a. Clearly, the HMM puts a lot of emphasis on the images labelled as "clear", assigning these exclusively to the zeroth and first state. The auroral classes have to share the second state and the third state is left for use by about half of the images labelled diffuse. One would assume, that based on this, the HMM would not be a good descriptor of the states. However, the HMM also has information about the probability of the classes, not just the actual class value available.

Modifying the same Table with the probabilities assigned by the classifier to each class, we obtain a distribution as shown in Table 1b. Previously we looked up each images state and label and counted the overlap. Now, instead of counting "1" for "hit" and "0" for "miss", we add the probabilities that have been assigned by the classifier for every image assigned to each state. This gives less of a hard count and more of an expected value of the mapping if a finer thresholding was possible. This gives us a broader picture for analysis while the overall result stays unchanged.

With this modification we can see how much of a role the probabilities for the arc and diffuse classes of images have already played in the first state of the HMM. Although none of the images were labelled as such, the expected accuracy of these images is high and increasing towards the next state, whereas the expected accuracy for the clear class is decreasing. The discrete class becomes most important during the second state and the diffuse class most for and during the third state. Interestingly however, the diffuse class plays a dominant role compared to the other auroral classes throughout all the states. This is because it is the overall dominant class, maybe because cloud-removal has not been working optimally and the original classifier does not manage to discern properly between a cloudy sky and diffuse aurora. Overall however, this supports our first interpretation of the models progression from clear skies to diffuse aurora through its four hidden states.

At one image every 3 s, we expect a total of 100 images per 5 minutes. However, since we have removed some images because they are not interesting or the camera may have had problems, not all bins contain the full 100 images. This problem is solved by binning the time series into 5 minute intervals. This also reduces the amount of total data as input for the classifier. Each bin will only contain the distribution of hidden states for each of the bins, which gives us a very condensed view of how the observed aurora evolves over time.

3.4 Substorms

The list of substorms contains substorms measured and registered all over the world. Because we are only interested in substorms that we will be able to recognise visually based on our images, we remove all substorms outside of a 10° region around the location of the classifier. We also remove any substorms registered at a time where there is no image data available. Doing this we obtain 245 individual substorms.

3.5 Classification

Our classifier is a simple Linear Ridge Model. As input we use 6 bins of 5 minutes of image data, giving us an input vector containing $6 \times 4 = 24$ cells of input data. As output to be predicted we use a Boolean value whether there will be a substorm within the next 15 minutes after the end of the input interval.

In Figure 5 we demonstrate how the input is prepared for the model. In the upper row the states for each image up to sixty minutes before and after a substorm has been identified are plotted. In the middle row, the distribution of states for each of the 5 minute bins is calculated and shown. For every 5-minute-Interval the occurrence of each state has been counted and divided by the total amount of images per interval. This removes the problem if less than the maximum possible amount of images have been taken in a given interval. The bottom two panels show a visualisation of the input for the classifier. Each contains a 30-minute-interval of data. The first interval ends less than 15 minutes before the substorm occurs and has therefore been given a positive label. The second interval ends more than 15 minutes before the substorm and has therefore been given a negative label.

4 Results

4.1 Distribution of Hidden States

In Figure 6 we show the distribution of hidden states up to an hour before and after a substorm has been observed. A total of 261 substorms have been observed within an hour before or after images have been taken. In the upper panel each individual event is plotted, the bottom panel shows the average distribution of hidden states. About 20% of substorms are accompanied by images in the zeroth state. We interpreted this state to be equal to images showing a clear sky. The remaining images start with structures identified as the first or second state an hour before the substorm. Towards the onset of the substorm, the first state becomes less prevalent compared to the second and at half an hour after the substorm the third state is the dominant one.

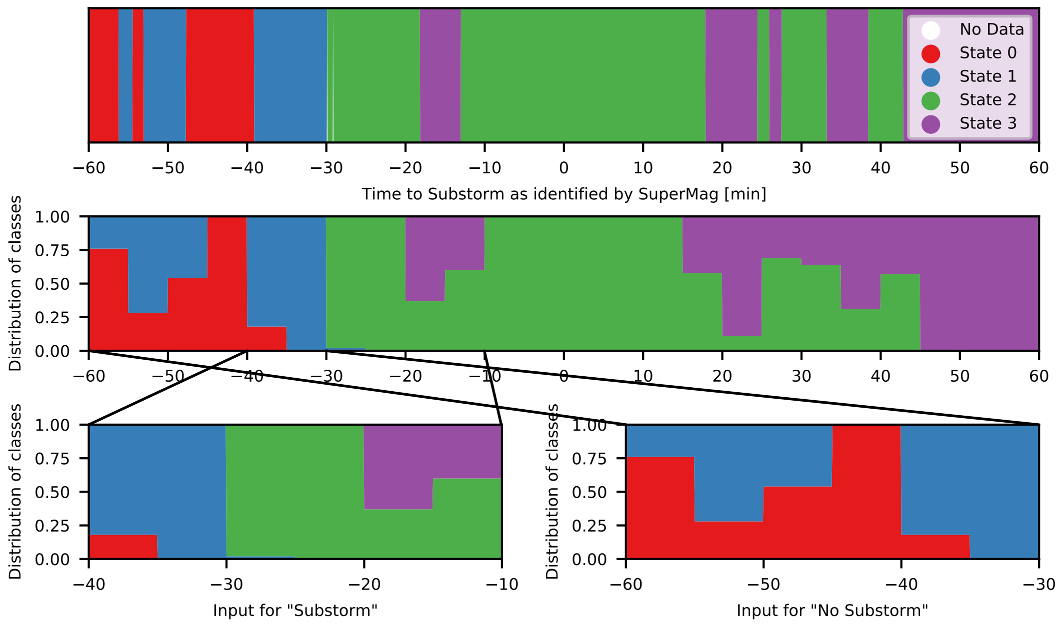


Figure 5: HMM-predicted states per image (top), binned distribution of classes (middle) and input for "no substorm" (bottom left) or "substorm" (bottom right)

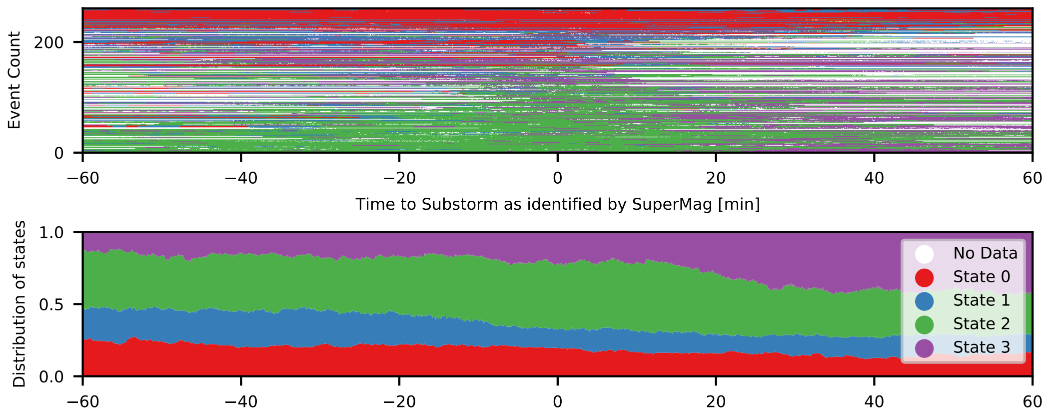


Figure 6: Distribution of Image states around substorms

4.2 Classification

For the classification task we prepare the classifier as described in section 3.5.

In Figure 7a, we see a ROC-curve for the classifier, which we obtain by choosing different thresholds. The area under the curve is 0.67. Overall precision is at only 3.5% because of the imbalance in data between negative and positive events of approximately 50:1. Still we are able to choose a working regime with a balanced accuracy of 61%, a false positive rate of 24% and a true positive rate of 47%.

Figure 7b displays the classification for a specific date. In the top panel we show ground based magnetometer measurements for the evening and the keogram for the time-frame in the panel below. The third panel shows the probabilities of individual images over time, the fourth panel the states assigned by the Hidden Markov Model. These states are prepared as displayed in Figure 5 and the probabilities for a substorm to occur based on our model are shown in the fifth panel. A threshold is chosen accordingly, giving a binary classification for "substorm" and "no substorm" as displayed in the sixth panel. Notably, there are periods where "substorm" and "no substorm" are switching back and forth, these have been filled in manually. The last panel shows the known true test data.

We see that the substorm occurring at 08:56 has been identified correctly, while overshooting slightly with the duration. Between 7:10 and 08:05 another event has been identified. As we can see from the magnetometer measurements plotted alongside, another substorm happened earlier with its onset identified at 06:56 by Ohtani and Gjerloev (2020). This substorm is not in our list of true positive data, because it occurred too early after onset of observations. Even if it was, it would not have been identified at the correct time, but the classifier has correctly identified that there was an ongoing event during the time. The substorm was also a longer lasting event, which was picked up by the classifier.

Both cases lead us to believe that the classifier prefers to identify ongoing substorms instead of the substorm onsets, it was trained on. This is most likely due to the fact that the definition of a substorm onset is rather arbitrary with respect to image data and the effect on the images heavily depends on the duration and strength of the substorm.

Nevertheless, the fact that the classifier managed to roughly identify the time both events occurred, is a huge success given the very limited model and training data. It has been trained on data only giving information about the onset of the substorm, resulting in a large imbalance between true and false cases of about 1:50. This means that just by guessing "false" all the time the classifier would achieve an accuracy of about 98%. This would correspond to the top-right corner of the ROC-curve.

Using the correct threshold it is possible to obtain a working regime that is performing better than this trivial case. Given the fact that the original input for half an hour of data has been condensed down from 600 images at 256 px by 256 px giving approximately 40M data points total to just 24 input values, this is a good achievement for a linear classifier.

To show that the preprocessing step with the HMM is an improvement over the raw input data, we have added two test cases, one without HMM-preprocessing and one with more hidden states. In the first, we use the classes as obtained by the original classifier's probabilities, with a Gaussian filter smoothing the transitions between images and filtering out some of the noise. Figure 7c shows the ROC-curve. The overall curve resembles that obtained by the HMM-preprocessed model, however the best case performs considerably worse and it is more difficult to choose a working regime for this classifier. Two regions between a false positive rate of approximately 0.2-0.4 and approximately 0.5-0.95 show a straight line because thresholding cannot resolve these regimes better. Finetuning a classifier to be within this region is therefore not possible, making this classifier not feasible.

Increasing the number of hidden states to 10 (see Figure 7d) does not yield an improvement either. The thresholding is a bit more stable, but overall worse and comes at the cost of increased training time.

The classifier based on 4 hidden states (cf. Figure 7a) shows these unstable regions as well, but here they are smaller and overall less problematic than for the classifier trained without the usage of a HMM for preprocessing.

5 Conclusion & Outlook

We have shown that a simple linear classifier based on the distribution of image classes of auroral images for up to half an hour can predict the onset of a substorm with respectable accuracy. Given the limitations of a linear model and how much information was discarded in the preprocessing stages to allow for quick training and evaluation, this raises the assumption that a more complex model utilising more information can achieve a much better result. The input data also only contained information directly obtained from images. Replacing the model with a neural network and supplementing the input data with for example solar wind data could lead to an accurate prediction of the local onset and possibly duration of substorms.

Because this method and underlying source code is made freely available, it could be used to forecast substorms live. While we have not undertaken such steps, the time-limiting factor in a project like this would be preprocessing of images. Since our methods operate much faster on commercial hardware than the limit of one image every three seconds, a proper implementation should be possible.

Acknowledgements

This work is funded in part by the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (ERC CoG grant agreement No 866357). The All Sky Image Classifier was made available by Sado et al. (2022) on `http://tid.uio.no/TAME`. We provide the data and code for this project openly and freely on `https://doi.org/10.11582/2022.00040` and `http://tid.uio.no/SOP` respectively. We acknowledge NASA contract NAS5-02099 and V. Angelopoulos for use of data from the THEMIS Mission. Specifically: S. Mende and E. Donovan for use of the ASI data, the CSA for logistical support in fielding and data retrieval from the GBO stations, and NSF for support of GIMNAST through grant AGS-1004736. We acknowledge the substorm timing list identified by the SOPHIE technique (Forsyth et al., 2015), the SMU and SML indices (P. T. Newell & Gjerloev, 2011), the Ohtani and Gjerloev technique (Ohtani & Gjerloev, 2020), the SMU and SML indices (P. T. Newell & Gjerloev, 2011); and the SuperMAG collaboration (Gjerloev, 2012).

References

Akasofu, S.-I. (1964, April). The development of the auroral substorm. *Planetary and Space Science*, *12*(4), 273–282. Retrieved 2022-07-28, from `https://linkinghub.elsevier.com/retrieve/pii/0032063364901515` doi: 10.1016/0032-0633(64)90151-5

Akasofu, S.-I. (2013, March). Auroral Morphology: A Historical Account and Major Auroral Features During Auroral Substorms. In A. Keiling, E. Donovan, F. Bagenal, & T. Karlsson (Eds.), *Geophysical Monograph Series* (pp. 29–38). Washington, D. C.: American Geophysical Union. Retrieved 2022-08-22, from `http://doi.wiley.com/10.1029/2011GM001156` doi: 10.1029/2011GM001156

Akasofu, S.-I., Chapman, S., & Meng, C.-I. (1965, November). The polar electrojet. *Journal of Atmospheric and Terrestrial Physics*, *27*(11-12), 1275–1305. Retrieved 2022-07-28, from `https://linkinghub.elsevier.com/retrieve/pii/0021916965900875` doi: 10.1016/0021-9169(65)90087-5

Caan, M. N., McPherron, R. L., & Russell, C. T. (1975). Substorm and interplanetary magnetic field effects on the geomagnetic tail lobes. *Journal of Geophysi-*

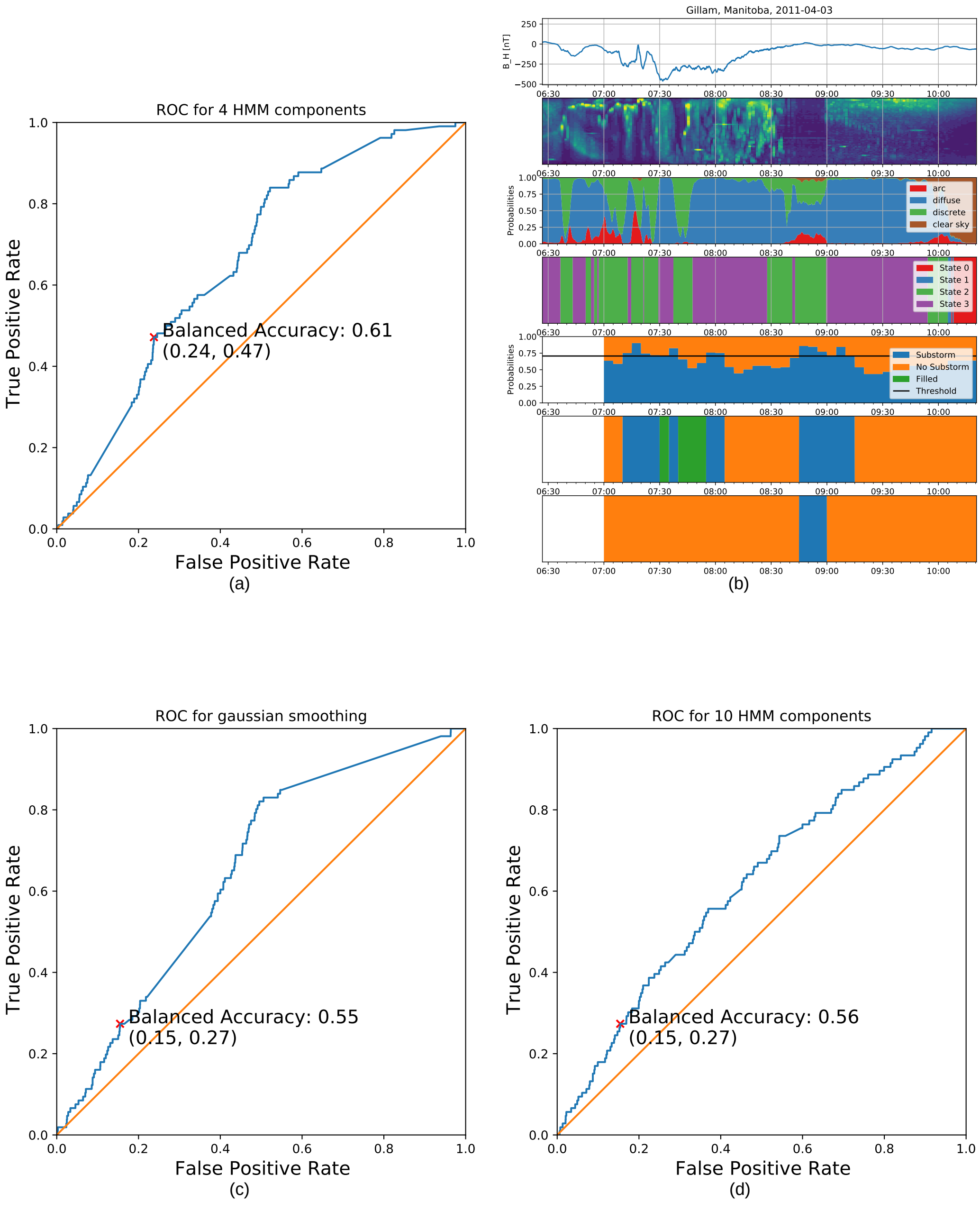


Figure 7: Overall Goodness of classification for the hmm-based classifier with 4 hidden states (a), for 10 hidden states (d) and for a classifier based on the original classes (c). A demonstration of Classification of Time series is shown in (b) for the 4 hidden classes case.

cal Research, *80*(1), 191–194.

Clausen, L. B. N., & Nickisch, H. (2018). Automatic classification of auroral images from the oslo auroral themis (oath) data set using machine learning. *Journal of Geophysical Research: Space Physics*, *123*(7), 5640-5647. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JA025274 doi: https://doi.org/10.1029/2018JA025274

Forsyth, C., Rae, I. J., Coxon, J. C., Freeman, M. P., Jackman, C. M., Gjerloev, J., & Fazakerley, A. N. (2015). A new technique for determining substorm onsets and phases from indices of the electrojet (sophie). *Journal of Geophysical Research: Space Physics*, *120*(12), 10,592-10,606. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015JA021343 doi: https://doi.org/10.1002/2015JA021343

Frey, H., Mende, S., Angelopoulos, V., & Donovan, E. (2004). Substorm onset observations by image-fuv. *Journal of Geophysical Research: Space Physics*, *109*(A10).

Gjerloev, J. W. (2012). The supermag data processing technique. *Journal of Geophysical Research: Space Physics*, *117*(A9). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012JA017683 doi: https://doi.org/10.1029/2012JA017683

Kintner, P. M., Ledvina, B. M., & de Paula, E. R. (2007, September). GPS and ionospheric scintillations: GPS AND IONOSPHERIC SCINTILLATIONS. *Space Weather*, *5*(9), n/a–n/a. Retrieved 2022-08-22, from http://doi.wiley.com/10.1029/2006SW000260 doi: 10.1029/2006SW000260

Liou, K. (2010). Polar ultraviolet imager observation of auroral breakup. *Journal of Geophysical Research: Space Physics*, *115*(A12).

Maimaiti, M., Kunduri, B., Ruohoniemi, J., Baker, J., & House, L. L. (2019). A deep learning-based approach to forecast the onset of magnetic substorms. *Space Weather*, *17*(11), 1534–1552.

Marcos, F., Lai, S., Huang, C., Lin, C., Retterer, J., Delay, S., & Sutton, E. (2010, August). Towards Next Level Satellite Drag Modeling. In *AIAA Atmospheric and Space Environments Conference.* Toronto, Ontario, Canada: American Institute of Aeronautics and Astronautics. Retrieved 2022-08-22, from https://arc.aiaa.org/doi/10.2514/6.2010-7840 doi: 10.2514/6.2010-7840

McPherron, R. L., Russell, C. T., & Aubry, M. P. (1973, June). Satellite studies of magnetospheric substorms on August 15, 1968: 9. Phenomenological model for substorms. *Journal of Geophysical Research*, *78*(16), 3131–3149. Retrieved 2022-07-28, from http://doi.wiley.com/10.1029/JA078i016p03131 doi: 10.1029/JA078i016p03131

Newell, P., Liou, K., Gjerloev, J., Sotirelis, T., Wing, S., & Mitchell, E. (2016). Substorm probabilities are best predicted from solar wind speed. *Journal of Atmospheric and Solar-Terrestrial Physics*, *146*, 28–37.

Newell, P. T., & Gjerloev, J. W. (2011). Evaluation of supermag auroral electrojet indices as indicators of substorms and auroral power. *Journal of Geophysical Research: Space Physics*, *116*(A12). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011JA016779 doi: https://doi.org/10.1029/2011JA016779

Newell, P. T., Lee, A. R., Liou, K., Ohtani, S.-I., Sotirelis, T., & Wing, S. (2010, September). Substorm cycle dependence of various types of aurora: SUBSTORM DEPENDENCE OF AURORAL TYPES. *Journal of Geophysical Research: Space Physics*, *115*(A9), n/a–n/a. Retrieved 2022-08-22, from http://doi.wiley.com/10.1029/2010JA015331 doi: 10.1029/2010JA015331

Newell, P. T., & Liou, K. (2011). Solar wind driving and substorm triggering. *Journal of Geophysical Research: Space Physics*, *116*(A3). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JA016139 doi: https://doi.org/10.1029/2010JA016139

Ohtani, S., & Gjerloev, J. W. (2020). Is the substorm current wedge an ensemble of wedgelets?: Revisit to midlatitude positive bays. *Journal of Geophysical Research: Space Physics*, *125*(9), e2020JA027902. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020JA027902 (e2020JA027902 2020JA027902) doi: https://doi.org/10.1029/2020JA027902

Partamies, N., Juusola, L., Whiter, D., & Kauristie, K. (2015, July). Substorm evolution of auroral structures. *Journal of Geophysical Research: Space Physics*, *120*(7), 5958–5972. Retrieved 2022-08-03, from https://onlinelibrary.wiley.com/doi/10.1002/2015JA021217 doi: 10.1002/2015JA021217

Pirjola, R. (2000, December). Geomagnetically induced currents during magnetic storms. *IEEE Transactions on Plasma Science*, *28*(6), 1867–1873. Retrieved 2022-08-22, from http://ieeexplore.ieee.org/document/902215/ doi: 10.1109/27.902215

Rabiner, L. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257-286. doi: 10.1109/5.18626

Sado, P., Clausen, L. B. N., Miloch, W. J., & Nickisch, H. (2022). Transfer learning aurora image classification and magnetic disturbance evaluation. *Journal of Geophysical Research: Space Physics*, *127*(1), e2021JA029683. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021JA029683 (e2021JA029683 2021JA029683) doi: https://doi.org/10.1029/2021JA029683