Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2019-481-RC2, 2020

Atmospheric Measurement Techniques Discussions

Interactive comment on "Exploration of machine learning methods for the classification of infrared limb spectra of polar stratospheric clouds" *by* Rocco Sedona et al.

Anonymous Referee #1

Received and published: 26 February 2020

An innovative study is presented, applying machine learning techniques for PSC classification in limb emission infrared spectra. A SVM-based classifier is applied, using input from PCA and KPCA feature extraction from a large set of BTDs, and a RF-based classifier using BTD features without prior feature selection. The methods are compared with an established PSC classification method reported in the literature (Bayesian classifier). Performance of the new classifiers is assessed using MIPAS data from the Northern hemisphere winter 2006/2007 and the Southern hemisphere winter 2009. Potential advantages in comparison to conventional methods are discussed.

The presented study using ML approaches is timely and clearly of interest for PSC

Printer-friendly version

Discussion paper

classification. The manuscript is well organized and mostly written clearly. However, the assessment and comparison of the new ML approaches with the conventional BC method is sometimes difficult to follow. Overlap regions and potential ambiguities of the different classification methods used are not clearly defined. Benefits of the new methods should be elaborated more clearly: why should the user decide to choose one of the new methods instead of established methods that are based on physical understanding and expert knowledge? Do the new methods provide scientifically more robust results? I recommend publication after addressing the following points:

Major points

1) To assess the performance of the new ML methods, different classification schemes are used for the ‘conventional’ reference method (BC) and the new ML methods. A clear comparison of the different classifications schemes, their overlap regions, and potential ambiguities is missing and should be provided prior to chapters 4.3.1 and 4.3.2. It should be clearly defined what is counted as ‘ice’, ‘sts’ and ‘nat’ in the discussion in these chapters (e.g., is ‘nat_sts’ counted to both, ‘nat’ and ‘sts’? What about ‘ice_sts’? Can ‘stsmix’ include ice or nat, too?). Furthermore, are the categories ‘ice’, ‘sts’ and ‘nat’ used in the sense of composition classes such as used by Pitts et al. (2018) in the discussion and Figs. 9-11? Or does it mean that the optical properties of these constituents can be identified/are dominating? Clear definitions of used categories should be provided, and it should be differentiated between PSC types, composition classes and constituents.

2) The benefits of the new methods should be elaborated more clearly. Are the new methods really more ‘objective’? For interpretation, still comparisons with conventional data are needed, and in the end an expert needs to decide which method to trust. Are the results scientifically more robust than conventional methods that are based on physical understanding? From what has been learnt here, would it be possible to set up a robust ML PSC classification without support by a conventional method and expert knowledge?

Specific comments

1.8 is feature extraction really done from both ('these') datasets? If I understood correctly, feature extraction is done using only the CSDB but not the observations.

1.12 This sentence suggests that PCA and KPCA in combination with both, RF and SVM. If I understood correctly, PCA and KPCA are done only in combination with SVM (i.e. PCA+SVM and KPCA+SVM). Cp 8.14: 'RF. . .without prior feature selection'

2.2 PSCs play another important role in ozone depletion by denitrification of the stratosphere; this should be mentioned

2.3 'main types' – the choice of the terms types, constituents and composition classes should be taken with care. In reality, PSCs are often mixtures. Is 'type' used here in the sense of constituents or composition classes such as used by Pitts et al. (2018)? 'main constituents' seems more appropriate here.

2.4 What defines a 'main method'? What about airborne/balloon-borne non-optical in situ observations (mass spectrometry, chemiluminescence) and remote sensing (lidar, limb)? There are many references on other methods in the literature (e.g. Voigt et al., 2000, Molleker et al., 2014, Woiwode et al., 2016, Voigt et al., 2018). Also, microwave observations where shown to be valuable to study PSCs (e.g. Lambert et al. 2012).

2.15 'from the simulated MIPAS spectra' Is 'simulated' missing? If I understood correctly, feature selection is done only from the simulated CSDB data but not from the real MIPAS spectra

2.15 What is meant by 'type'? Composition class? Or does it mean that the optical properties of these constituents can be identified/are dominating?

2.16 'first time that ML methods . . . MIPAS PSC observations' This statement should be revisited. In the literature, Bayesian classifiers such as used Spang et al., 2016 are frequently termed as ML methods (e.g. https://en.wikipedia.org/wiki/Naive_Bayes_classifier, 26.2.2020). Could the work by Spang et al. be considered as first ML application to MIPAS?

2.24f '(PCA) and (KPCA) for feature selection, followed by... (RF) and (SVM)' Is this consistent with 8.14 'RF...without prior feature selection'? If I understood correctly, PCA and KPCA is not done in RF.

3.17ff The use of 'windows' should be revisited. Does 'spectral windows of 1 cmˆ-1' mean that the data is down-sampled or smoothed to a resolution of 1 cmˆ-1 (cp Tab. 1, the 'windows' are broader than 1 cmˆ-1)? Does 'five larger windows' mean larger than 1 cmˆ-1 or larger than R1-R8? The latter does not seem to be the case. It should be differentiated between 'spectral window' (such as R1-8 and W1-5) and spectral resolution.

3.24 What kind of 'background signals' are removed?

4.8 Höpfner et el. 2006 give an upper limit of r=3 μm instead of 2 μm for small NAT particles

4.11ff This section should be revisited. Is there a difference between 'very thin', 'thin' and 'thinnest', or is it the same? PSCs are often (very) thin clouds when compared to other clouds. What is meant by 'atmospheric variability'?

4.23 Here it would be really helpful for the reader to introduce the types or composition classes identified by the BC and discuss here or later the overlap with the ML classification and potential ambiguities. Which classes are summarized as 'nat', 'sts' and 'ice' in the later comparison with the ML results? Just a suggestion: a tabular comparison might be helpful to compare the different classifications and indicate what is counted as 'nat', 'sts' and 'ice'.

6.1ff Section 3.3 gives interesting general information about the ML methods, but the link to the presented work is somehow missing to me. How are the methods used here? What kind of input data is used and what kind of output is generated? What are the critical parameters here? The used classification scheme should be mentioned or at least a reference to the compositions in section 2.2 should be added.

7.14ff 'a closer inspection shows. . .' In Fig. 2a, most of the area look yellowish to me; it might be helpful to adjust the color-coding. Furthermore, it might be helpful to highlight R1, R2. . . directly in Fig. 2a and 2b, since it is difficult to follow the discussion, connect regions and wave numbers with indices by using Tab. 1, and then try to identify index ranges on the panel axes. R1, R2. . . might be indicated also in Fig. 4 for easier reading.

7.17 What do the 'pronounced features' in the real an imaginary refractive index mean physically? How are they related to the spectra? Why does it make sense to feed the complex refractive indices to the ML methods, while the goal is to classify measured limb spectra and not refractive indices? At least a short explanation should be provided.

7.31 'similar clusters as Fig 2a' I have difficulties in finding the similarities, since the discussion uses wave numbers and regions while the figures use 'index'. See above: it might be helpful to indicate R1. . . somehow in the panels.

8.4 'peak in imaginary part' Which peak is meant here? 'minimum in the real part' Which minimum is meant here? See comment to 7.17: how are these refractive index features related to the spectra?

8.25 Possibly I missed it: how is the prediction accuracy determined? See Figs 12ff: How can the prediction accuracy be 99% for all methods while the classification results are relatively heterogeneous?

9.13ff Here and in in the following I got somewhat confused: For BC, the PSC classes 'unspec', 'ice', 'nat', 'stsmix', 'ice_nat', 'nat_sts', 'ice_sts' are used. For the other methods, the classes 'ice', 'nat_large', 'sts_1', 'sts_2', 'sts_3' and 'nat_small' are used. In Fig. 12-14, suddenly 'sts_mix1', 'sts_mix2' and 'sts_mix3' are used (I guess 'sts_1', 'sts_2' and 'sts_3' is meant here). In the text, the types or categories 'nat', 'sts' and 'ice' are used. The used categories and classifications should be clearly defined. Overlap regions of and potential ambiguities between the different classification schemes should be discussed (see comments to 4.23 and 2.3).

11.33 are the new approaches really ‘more objective’, reminding that they need to be assessed using a ‘conventional’ method based on a-priori knowledge and expert knowledge, and finally one needs make a choice?

12.28 Just out of curiosity: would it be possible to make a meaningful search for further PSC constituents not covered by conventional classifications, such as nitric acid dihydrate?

Technical

2.19 approaches

3.21 have been extracted

References

Voigt et al., Science, 290, 1756–1758, 2000

Molleker et al., Atmos. Chem. Phys., 14, 10785–10801, 2014

Woiwode et al., Atmos. Chem. Phys., 16, 9505–9532, 2016

Voigt et al., Atmos. Chem. Phys., 18, 15623–15641, 2018

Lambert et al., Atmos. Chem. Phys., 12, 2899–2931, 2012

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2019-481, 2020.