A CLIMATOLOGICAL DESCRIPTION OF MCS AND MCD THERMAL PROFILES USING A CLUSTER ANALYSIS.


Introduction

The thermal structure of the Martian atmosphere has been explored up to about 80 km thanks to the Mars Climate Sounder on board the Mars Reconnaissance Orbiter (MRO) for almost a decade to date [Mc Cleese et al. (2007), Kleinbühl et al. (2013)]. This MCS continuous sounding has created a unique and valuable dataset, in particular at mesospheric altitudes, where no previous regular observations existed, thanks to the limb capabilities of this instrument (Kleinbühl et al., 2009). The MCS data have revealed and got new insight of interesting phenomena, like strong inversions in polar regions (McCleese et al., 2008) or the propagation of semi-diurnal tides (McCleese et al., 2010). The present work aims at further exploiting this dataset. This is part of our on-going research intended to apply machine learning tools to analyze datasets and models of the Martian atmosphere. One of these goals is to apply cluster analysis to the very extensive MCS dataset in order to extract regularities and to build a simplified climatology of the Martian atmosphere from the troposphere up to the mesopause, based on such data.

On the other hand, on the modeling side, global circulation models of the Martian atmosphere like the LMD-GCM (Forget et al., 1999) are able to simulate the thermal structure, composition and dynamics up to thermospheric altitudes [González-Galindo et al. (2009), González-Galindo et al. (2015)]. The validation of these models at high altitudes is challenging due to the limited data available there. The MCS temperature profiles offer an excellent benchmark to test the GCM results at mesospheric altitudes. However, a well-known limitation when comparing observations against GCM results is the difficulty of performing such an exercise on a profile-by-profile basis, given the global and statistical nature of these 3D models’ predictions. For these reasons, a comparison based on climatological classes should be more meaningful, and to our knowledge, a fully novel approach to model-data comparisons on Mars atmospheric science (Cala-Hurtado, 2016).

In this on-going project we applied an unsupervised learning algorithm (cluster analysis) to vertical profiles of atmospheric temperatures in order to perform a climatological description of the MCS dataset, and independently, of the LMD-GCM model results, and the first results are presented and discussed below. In addition, we hope this study may eventually supply useful information for the detection of potential shortcomings in the MCS data processing and retrieval techniques.

The MCS and MCD datasets

We have selected four full Martian years of temperature retrievals of MCS, extending from MY29 to MY32, from the PDS Level 2 pipeline v4.3. We defined the vertical profiles as comprising all the nominal retrieval points between pressure levels [200,0.06] Pa, and use the pseudo-altitude \( z^* = -ln(P/P_{max}) \) as the vertical scale, where \( P_{max}=200 \) Pa. The lower boundary is to avoid the high opacity of the Martian atmosphere, which normally saturates the thermal channels in a limb geometry (Kleinbühl et al., 2009). The data coverage of the vertical profiles defined in this way is shown in Figure 1. Few profiles are available during the summer season in the Southern Hemisphere at mid and high latitudes in that hemisphere, due to the high dust loading, and also during Northern Spring and Summer at low latitudes when retrievals also fail below about 10 Pa. An approach for retrieving profiles in conditions with high aerosol opacity is under investigation by the MCS team (A. Kleinbühl, personal communication) and might permit the extension of this study to lower altitudes.

Regarding the model results, instead of direct GCM outputs we used the Mars Climate Database (MCD) [Lewis et al. (1999), Millour et al. (2015)], because it is a handy reference dataset including a set of diverse scenarios which could be analyzed separately. So far we have only focused on the so called “climatological scenario”, intended to represent a dust loading averaged over MY24 to MY30. Instead of the direct web access (http://wwwmars.lmd.jussieu.fr) we used the DVD version 5.2, and Python and R subroutines for intensive access to the netCDF formatted data. In the selection process of the MCD data we used \( P_{max}=300 \) Pa, and the same definition of pseudo-altitude than for MCS. Consequently, for the MCS-MCD comparison we need to recall that \( z_{MCS} = z_{MCD} - ln(300/200) = z_{MCD} - 0.4 \).
The clustering analysis

The application of unsupervised statistical analysis to surface and mineralogical studies is not new on Mars [Marzo et al (2006), Roush et al (2007), Fonti, S. et al (2015)], but has barely been exploited for atmospheric studies. Marzo et al (2008) made a first application of a cluster analysis to the temperature/pressure profiles of the MCD (version 4.3) in order to obtain a simplified climatology and perform a quality assessment of the MCD. They obtained that a “natural number” of clusters of 12 were sufficient to explain the major latitudinal, seasons and local time variations in almost all the scenarios of the MCD. The exception was the dust-storm scenario, which required four additional classes.

The present work is based and inspired on that work, and as Marzo et al (2008) did, we also used a k-means clustering and applied it to our selection of vertical profiles in a fix altitude range. As it is customary in this method, the centroids are found using an euclidean distance in the desired space, which in our case is the altitude-temperature space. The std.dev of the clusters shows the data dispersion at each altitude, and is interpreted here as atmospheric variability. The degree of homogeneity within each cluster and the heterogeneity between clusters is quantified with a quality index. In our case we used the Calinski-Harabasz index. The number K of clusters, which needs to be declared before the clustering is performed, was initially left as a free parameter, i.e. it was varied during this investigation. We tried diverse clustering, between k=6 and 12 for individual seasons and Martian years. We show next some results for K=6.

Preliminary Results

Figure 2 presents results from 16 clustering exercises of the MCS data after split in four seasons and for each of the four Martian years, and using K=6 as the number of clusters. The MCS clusters show a significant degree of inter-annual repeatability when separate seasons are considered. This Figure also shows 4 clustering of the MCD, one for each season of the “climatological” scenario. What are shown is the spatial location of the six clusters, and both for the MCS and MCD datasets. Notice the absence of data in the MCS at high latitudes in the Southern Hemisphere during summer there, due...
REFERENCES

to the elevated dust during that season. Also in low latitudes during Spring and Summer there is a lack of MCS vertical profiles covering the whole altitude range selected in this work. In spite of this incomplete coverage, some similarities and differences stand clear in Figure 2.

A common feature between MCS and MCD, clearly seen in Figure 1, is that most of the clusters are located at high latitudes, where the variations in temperature are largest, as expected. Also, both datasets show strong thermal inversions during in the winter hemisphere, with tropospheric temperatures decreasing with latitude. Also the inversion occurs at higher altitudes the closer to the winter pole. These seem to be robust features of the Martian thermal structure and are in agreement with previous descriptions of the latitudinal variations using MCS (McClleese et al, 2010).

Next we describe two interesting differences between the model and the data. As we can see in Figure 2, in all the seasons except in Northern winter, the dispersion in temperature observed in MCS in the troposphere is larger than in the mesosphere, with a minimum value around $z_{MCS}^*=4$. The MCD only shows such a minimum in winter and a little bit higher, around $z_{MCD}^*=5$. The MCD tropospheric and mesospheric dispersions in Spring and Summer are similar, in contrast to the MCS dataset. The main reason for these differences seems to be the coldest cluster of the MCD dataset (in blue), which does not have a clear correspondence in the MCS. The mesosphere seems to be too cold (compared to the MCS data) during the Spring and Summer periods. The high latitudes in the MCS do have a cold troposphere but present a strong inversion which is not reproduced in the MCD during those seasons.

The opposite also occurs, an observed thermal state missing in the MCD. For example the green cluster during the Spring season at low latitudes in MY29 and MY30 (which is yellow in MY31 and MY32), is not seen in the MCD clustering. It represents a very warm troposphere and a cold mesosphere, with temperatures varying from about 190 to about 135 K, respectively. In contrast, the yellow cluster in the MCS results only varies between about 195 and 150 K. Also the gradient in the top of the figure is negative, while the MCS shows a positive gradient, possibly towards a warmer mesopause than in the MCD.

This work is in progress; we continue exploring model-data differences and the possible reasons behind them.

Acknowledgments

This work has been partially funded by the European Union Horizon 2020 Programme (H2020 - Compet - 08 - 2014) under grant agreement UPWARDS-633127. The Spanish team has also been supported by ESP2015-65064-C2-1-P (MINECO/FEDER).

References


Figure 2: Clusters of the daytime (2 pm) thermal profiles of the MCS and MCD datasets, separated by years (in the MCS only) and by seasons. Colors are arbitrary in each case but correspond to the color scale in Figure 1. See text for details.


