Column description:
----

1) objectname=CoRoT ID
2) normalized Mahalanobis distance to the centre of the most probable class (class 1)
3) normalized Mahalanobis distance to the centre of the second most probable class (class 2)
4) normalized Mahalanobis distance to the centre of the third most probable class (class 3)
5) classprob1=relative probability for class 1
6) classprob2=relative probability for class 2
7) classprob3=relative probability for class 3
8) classcode1=variability class 1
9) classcode2=variability class 2
10) classcode3=variability class 3
11) Pf1=significance parameter frequency 1 (probability)
12)Pf2=significance parameter frequency 2 (probability)
13)Pf3=significance parameter frequency 3 (probability)
14)f1=frequency 1 (units: cycles per day)
15)f2=frequency 2 (units: cycles per day)
16)f3=frequency 3 (units: cycles per day)
17)amp11=amplitude of 1th harmonic of f1 (units: magnitude)
18)amp12=amplitude of 2th harmonic of f1 (units: magnitude)
19)amp13=amplitude of 3th harmonic of f1 (units: magnitude)
20)amp14=amplitude of 4th harmonic of f1 (units: magnitude)
21)amp21=amplitude of 1th harmonic of f2 (units: magnitude)
22)amp22=amplitude of 2th harmonic of f2 (units: magnitude)
23)amp23=amplitude of 3th harmonic of f2 (units: magnitude)
24)amp24=amplitude of 4th harmonic of f2 (units: magnitude)
25)amp31=amplitude of 1th harmonic of f3 (units: magnitude)
26)amp32=amplitude of 2th harmonic of f3 (units: magnitude)
27)amp33=amplitude of 3th harmonic of f3 (units: magnitude)
28)amp34=amplitude of 4th harmonic of f3 (units: magnitude)
29)phdiff12=phase of amp12, if phase of amp11=0 (units: radians)
30)phdiff13=phase of amp13, if phase of amp11=0 (units: radians)
31)phdiff14=phase of amp14, if phase of amp11=0 (units: radians)
32)phdiff21=phase of amp21, if phase of amp11=0 (units: radians)
33)phdiff22=phase of amp22, if phase of amp11=0 (units: radians)
34)phdiff23=phase of amp23, if phase of amp11=0 (units: radians)
35)phdiff24=phase of amp24, if phase of amp11=0 (units: radians)
36)phdiff31=phase of amp31, if phase of amp11=0 (units: radians)
37)phdiff32=phase of amp32, if phase of amp11=0 (units: radians)
38)phdiff33=phase of amp33, if phase of amp11=0 (units: radians)
39)phdiff34=phase of amp34, if phase of amp11=0 (units: radians)
40)varred=total variance reduction of the light curve, after subtraction of a least-squares fit with the 3 frequencies and their harmonics (values between 0 and 1)

Stellar variability classes (and their codes) considered by the CVC:

MISC Classification uncertain (see below)
MIRA Mira variables
SR Semi-regular variables
RVTAU RV-Tauri stars
RRAB RR-Lyrae stars, subtype ab
RRC RR-Lyrae stars, subtype c
RRD Double mode RR-Lyrae stars
CLCEP Classical Cepheids
DMCEP Double-mode Cepheids
SPB Slowly Pulsating B-stars
BCEP Beta-Cephei stars
DSCUT Delta-Scuti stars
GDOR Gamma-Doradus stars
ELL Ellipsoidal variables
ECL Eclipsing binaries (all types)
ROT Rotational modulation
ACT Activity

Using the CVC results to create candidate lists:

-Users interested in variability types considered by our classifier, should use the class codes ('classcode1-3') and corresponding probabilities ('classprob1-3'), to create candidate lists (see guidelines below).

-Note that also non-variable light curves get class probabilities, the classification results usually make no sense in this case. Non-variables cause a lot of false-positives (incorrect classifications), contaminating the candidate lists. The non-variables can be rejected, using the significance parameter 'Pf1' (related to the dominant detected frequency 'f1' and its harmonics). This parameter has values in the range 0-1. Non-variables will have 'Pf1' close to 1, clear variables will have 'Pf1' close to 0 (or equal to zero within the numerical precision). A reasonable threshold to reject nonvariables or very noisy ones is selecting only those light curves having 'Pf1' smaller than 0.7. If one is only interested in the clearest variables (having a high signal-to-noise ratio for 'f1'), smaller cutoff-values should be taken (0.1 or even smaller). An increasingly large number of objects will be rejected this way. The parameters 'Pf2' and 'Pf3' have the same meaning, but they relate to the significance of the second and third frequencies 'f2' and 'f3'.

-The class codes and corresponding probabilities can be used to make candidate lists with different levels of contamination. For example, users interested in RR-Lyrae stars of subtype ab, and wishing to minimize the risk of missing any, will look for those light curves having one of the 3 classcodes equal to 'RRAB'. Minimizing the risk of missing any objects of that type unfortunately also implies that the candidate list will be more contaminated by false-positives. Stronger selections can be made by selecting only those objects having 'classcode1' or 'classcode2' of the desired type, or only those having 'classcode1' of the desired type. For most classes considered by the classifier, the resulting candidate list will still contain a large fraction of false-positives. Imposing limits on the class

probabilities and the Mahalanobis distance will further clean the selection. In general, the higher the respective class probability and the lower the Mahalanobis distance (*), the better the candidate and the more similar it is to the objects used to define the class. However, low class probabilities (even lower than 0.5) do not necessarily imply that we are dealing with a bad candidate. BCEP and DSCUT stars for example can show very similar pulsation characteristics. Remember that the current classifiers do not take colour information into account, which would allow for a better separation of these 2 classes. Hence, the light curve of a BCEP star might get similar probabilities for the BCEP and DSCUT classes. The same is true for the SPB and GDOR classes, they are also difficult to distinguish without any spectral information.
Note that the listed probabilities are relative ones: all the class probabilities for every light curve sum up to 1. In the classification file, we only list the 3 most probable classes, in order of decreasing probability.

-Many objects get assigned the 'MISC' label as the most probable possibility. These are most often objects without variability, with mixed types of variability, or variability not typical for any of the classes we currently consider. The variables amongst these can be selected using the significance parameters for the frequencies (see description above).

(*) In short, the Mahalanobis distance is a multi-dimensional generalisation of the one-dimensional statistical or standard distance (e.g. distance to the mean value of a Gaussian in terms of sigma). This distance can effectively be used to retain only the objects that are not too far from the class centre in a statistical sense. It can be used together with the probabilities, in order to select the best candidates. Using only the probability values is usually insufficient to select the best candidates. Consider the case e.g., where the probability for one class is 99% (0.99). This high probability value seems to indicate a very certain class assignment. However, these are only relative probabilities, and, even though the relative probability for the class is very high, the object might still be very far away from the class centre. If this is the case, the Mahalanobis distance will have a large value, and one has to conclude that the object is not a good candidate to belong to the class after all. A typical cutoff value for this distance is 2 or 3 (think of outlier removal using 2 or 3-sigma cutoff values). The smaller the cutoff value used, the more similar the selected objects will be to the objects used to define the class.

-We implemented a separate classification of the low frequency part of the light curves (below 0.15 cycles/day) and the higher frequency part (above 0.15 cycles/day): for every object, there are now two lines in the file, those having a CoRoT ID preceded by 'lf-' indicate the classification of the low-frequency part. This split-up is done in order to avoid misclassifications due to jumps in the light curves (jumps mainly affect the low frequency part of the light curves). For most light curves, the classification of the low-frequency signal is not useful, since it is often strongly influenced by instrumental effects. It should allow the identification of clear variables with dominant frequencies below 0.15 cycles/day (e.g. Cepheids). The classification of the low-frequency part is not available for run SRc03, given the too limited time-span of the light curves.

-The definition of the classes ROT and ACT is based on the results obtained in the application of clustering analysis (unsupervised classification) to the CoRoT database (see Sarro et al., 2009; Debosscher et al., 2011, and Sarro et al., 2013).

More detailed information on the classification methods and the definition of the variability classes can be found in the following publications:

-Debosscher et al., 2007 (A&A): Automated supervised classification of variable stars. I. Methodology.
-Sarro et al., 2009 (A&A): Automated supervised classification of variable stars. II. Application to the OGLE database.
-Debosscher et al., 2009 (A&A): Automated supervised classification of variable stars in the CoRoT programme.
-Sarro et al., 2009 (A&A): Comparative clustering analysis of variable stars in the Hipparcos, OGLE Large Magellanic Cloud, and CoRoT exoplanet databases.
-Debosscher et al., 2011 (A&A): Global stellar variability study in the field-of-view of the Kepler satellite.
-Sarro et al., 2013 (A&A): Improved variability classification of CoRoT targets with Giraffe spectra.

Feedback/questions are always welcome:
contact: jonas@ster.kuleuven.be
Last change: 19/02/2013