

# VENUS CLIMATE DATABASE v2.3

## DETAILED DESIGN DOCUMENT

(ESA Contract N° 4000138542/22/NL/CRS “Venus Climate Database”)

S. Lebonnois, E. Millour, F. Forget, A. Martinez, T. Pierron, A. Boissinot (LMD, Paris)

May 2023

### **Abstract**

This is the Detailed Design Document for version 2.3 of the Venus Climate Database (VCD). It contains a detailed description of the database and addresses technical aspects of how the data is represented, manipulated and post-processed by the VCD access software. Instructions on how to install the VCD and use the provided access software and post-processing tools are given in the VCD v2.3 **User Manual**.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>General Description of Database Contents</b>	<b>3</b>
2.1	VCD Files . . . . .	3
2.2	VCD Software . . . . .	3
2.3	VCD Datasets and Datafiles . . . . .	4
2.4	Database Grid Structure . . . . .	6
2.4.1	Horizontal Structure . . . . .	6
2.4.2	Vertical Structure . . . . .	7
2.4.3	Temporal Structure . . . . .	9
<b>3</b>	<b>EUV and cloud albedo scenarios in the VCD</b>	<b>9</b>
3.1	The EUV Scenarios . . . . .	10
3.2	The Cloud albedo Scenarios . . . . .	10
<b>4</b>	<b>Technical description of Methods Used to Retrieve Data from the VCD</b>	<b>11</b>
4.1	Temporal Interpolations . . . . .	11
4.2	Spatial Interpolations . . . . .	12
4.2.1	Horizontal Interpolation . . . . .	12
4.2.2	Vertical Interpolation . . . . .	12
4.2.3	Specific Treatments of Vertical Interpolation . . . . .	13
<b>5</b>	<b>Variability Models in the VCD</b>	<b>14</b>
5.1	Hourly RMS of Variables . . . . .	15
5.2	Day-to-Day RMS of Variables . . . . .	16
5.3	The Large-Scale Variability Model . . . . .	16
5.3.1	Calculation of EOFs and PCs . . . . .	16
5.3.2	Storage of variability with EOFs . . . . .	17
5.3.3	Reconstructing a realistic variability . . . . .	18
5.4	The Small-Scale Variability Model . . . . .	18
<b>6</b>	<b>High Resolution Outputs</b>	<b>20</b>
6.1	High Resolution Topography . . . . .	20
6.2	Deriving High Resolution Surface Pressure . . . . .	20
6.3	Computing High Resolution Values of Atmospheric Variables . . . . .	21
6.3.1	Redefinition of high resolution levels . . . . .	21
6.3.2	Interpolation of atmospheric temperature . . . . .	23
6.3.3	Interpolation of atmospheric density . . . . .	23
6.3.4	Interpolation of other variables . . . . .	24
6.3.5	Modification of column densities . . . . .	24
<b>A</b>	<b>Computing Solar Longitude and Local Time</b>	<b>26</b>
A.1	Some constants . . . . .	26
A.2	Computing aerocentric solar longitude . . . . .	26
A.2.1	The three anomalies . . . . .	26
A.2.2	Aerocentric solar longitude $L_s$ . . . . .	26
A.2.3	Converting Julian date $JD$ to $L_s$ . . . . .	26
A.3	Computing Local True Solar Time . . . . .	27

# 1 Introduction

The Venus Climate Database (VCD) is a database of atmospheric statistics compiled from state-of-the art Global Climate Model (GCM) simulations of the Venusian atmosphere.

The Venus PCM (Venus Planetary Climate Model) computes in 3D the atmospheric circulation taking into account radiative transfer through the gaseous atmosphere, includes a representation of non-LTE processes, EUV heating, conduction and molecular diffusion in the thermosphere. Sub-grid processes such as convection in the boundary layer, non orographic and orographic gravity waves are also simulated. Photochemistry that controls the atmospheric composition is also implemented in the model which extends into the upper thermosphere.<sup>1</sup>

The model used to compile the statistics has been extensively validated using available observational data and aims at representing the current best knowledge of the state of the Venusian atmosphere given the observations and the physical laws which govern the atmospheric circulation and surface conditions on the planet.

The Venus Climate Database access software adds several capabilities to better represent the Venusian environment variability and accurately compute the atmospheric fields at high spatial resolution.

This document provides the user of the VCD with a detailed description of the database structure and of technical aspects of the access software.

The VCD is freely available online at request. Its content may also be accessed to make plots and figures using the interactive server on our WWW site at:

<http://www-venus.lmd.jussieu.fr>

## 2 General Description of Database Contents

### 2.1 VCD Files

The VCD main directory contains three folders: `docs` which contains documentation, `data` which contains the datafiles and `vcd` which contains the access software (see the User Manual).

### 2.2 VCD Software

Subdirectory `vcd` contains the following Fortran programs:

- `VCD.F90`: This is the main module, containing the main subroutine 'call\_vcd' which should be used to retrieve data from the database. The file also contains a collection of subsidiary routines. It requires the module `VCD_var.F90` to run. Input and output arguments of `call_vcd` are documented in the VCD User Manual.
- `julian.F90`: A subroutine which converts Earth date into the corresponding Julian date (one of the forms of date input used by `call_vcd`).

---

<sup>1</sup>A detailed description of the LMD Global Climate Model is out of the scope of the present document. Nevertheless, details can be found in the scientific papers provided with the database in the `docs/articles` directory:

- `Lebonnois.etal.2010.pdf`: 'Superrotation of Venus atmosphere analyzed with a full general circulation model'
- `Lebonnois.etal.2016.pdf`: 'Wave analysis in the atmosphere of Venus below 100km-altitude, simulated by the LMD Venus GCM'
- `Gilli.etal.2017.pdf`: 'Thermal structure of the upper atmosphere of Venus simulated by a ground-to-thermosphere GCM'
- `GarateLopez-Lebonnois.2018.pdf`: 'Latitudinal variation of clouds structure responsible for Venus cold collar'
- `Gilli.etal.2021.pdf`: 'Venus upper atmosphere revealed by a GCM: II. Model validation with temperature and density measurements'
- `Martinez.etal.2023.pdf`: 'Exploring the variability of the Venusian thermosphere with the IPSL Venus GCM'

- `test_vcd.F90`: A sample program provided to display (and test) the way to call subroutine `call_vcd` and to retrieve data from the VCD. The program can be run using as inputs the `test_vcd.def` file.

Apart from these Fortran programs, the `vcd` directory also contains subdirectories with examples of interfaces for other languages: `idl`, `matlab`, `scilab`, `python` and `c_interfaces` (which contains both C and C++ interface examples). See the accompanying README files and the VCD User Manual for details.

## 2.3 VCD Datasets and Datafiles

VCD data is stored in various folders and files in directory `data` using the Network Common Data Form (NetCDF) developed and distributed by Unidata. The NetCDF libraries are freely available for numerous platforms from the Unidata web site:

<http://www.unidata.ucar.edu/software/netcdf>

- Essential and common datafiles
  - File `surface.nc` which contains maps (at model resolution, i.e. a 96x96 longitude-latitude grid) of topography.
  - File `relief.nc` which contains a high resolution (23 pixels per degree) map of Venusian topography.
  - File `EUV_julian.txt` which contains the value of the E10.7 factor (in sfu) at a given Julian Date, computed from the 'Solar2000 empirical solar irradiance model' of *Tobiske et al.* (2000) for interpolation of EUV solar flux done in `call_vcd`.
  - A subdirectory `VIRA` containing the density, pressure and temperature data from VIRA (Venus International Reference Atmosphere).
- Individual cloud albedo and solar EUV scenarios (see section 3 for a detailed description) datafiles are in corresponding subdirectories:
  - `mean_aveEUV`, `mean_minEUV` and `mean_maxEUV` for the standard cloud albedo scenario. It represents the baseline climatology<sup>2</sup> scenarios,
  - `low_aveEUV`, `low_minEUV` and `low_maxEUV` for the low cloud albedo scenario
  - `high_aveEUV`, `high_minEUV` and `high_maxEUV` for the high cloud albedo scenario

The file naming convention for each file is as follows :

Data file name = `'scen'_'type'.nc` for low atmospheric files (up to around 100km – 50 levels)

Data file name = `'scen'_thermo_'solar'_'type'.nc` for thermospheric files (from around 100km up to around 250km – 40 levels).

Data file name = `'scen'_'solar'_eof.nc` for the Empirical Orthogonal Function (EOF) data files.

'scen' denotes the albedo scenario type (mean, low, high)

'type' indicates the type of data in the file :

'type' = 'me' means "mean data" contains mean fields but also rms data at constant pressure.

'type' = 'ad' means "altitude wise standard deviation data" contains rms data at constant altitude,

'solar' indicates the type of solar EUV scenario for thermosphere and EOF files:

'solar' = 'min' for fixed minimum EUV,

'solar' = 'ave' for fixed average EUV,

'solar' = 'max' for fixed maximum EUV,

---

<sup>2</sup>Note that only the `mean_aveEUV` directory is required; all others are optional, in the sense that they are only required when data for these specific scenarios is sought.

Each of the `me` files contain mean fields, pressure wise Venusian day-to-day RMS and Venusian hourly RMS. All the fields are stored with 24 values corresponding to 24 solar times of day (i.e. every 1 LT hour) for the variables shown in Table 1. The `ad` data files contain altitude wise Venusian day-to-day RMS and Venusian hourly RMS values of variables (see Table 2 for details). The fields are also stored with 24 values (i.e. every 1 LT hour).

Table 1: Variables stored in database mean data files

Variable	symbol	units	2-D or 3-D
Surface temperature	tsol_ave	K	2-D
V-hourly RMS Surface temperature	tsol_rms	K	2-D
V-day to day RMS Surface temperature	tsol_dtd	K	2-D
Surface pressure	psol_ave	Pa	2-D
V-hourly RMS Surface pressure	psol_rms	K	2-D
V-day to day RMS Surface pressure	psol_dtd	K	2-D
CO <sub>2</sub> column	col_co2_ave	kg/m <sup>2</sup>	2-D
CO column	col_co_ave	kg/m <sup>2</sup>	2-D
H <sub>2</sub> column	col_h2_ave	kg/m <sup>2</sup>	2-D
H <sub>2</sub> O column	col_h2o_ave	kg/m <sup>2</sup>	2-D
H <sub>2</sub> SO <sub>4</sub> column	col_h2so4_ave	kg/m <sup>2</sup>	2-D
H <sub>2</sub> SO <sub>4</sub> column	col_h2so4_ave	kg/m <sup>2</sup>	2-D
H column	col_h_ave	kg/m <sup>2</sup>	2-D
HCl column	col_hcl_ave	kg/m <sup>2</sup>	2-D
HCl column	col_hcl_ave	kg/m <sup>2</sup>	2-D
He column	col_he_ave	kg/m <sup>2</sup>	2-D
N <sub>2</sub> column	col_n2_ave	kg/m <sup>2</sup>	2-D
O <sub>2</sub> column	col_o2_ave	kg/m <sup>2</sup>	2-D
O <sub>2</sub> nightglow column	col_o2dg_ave	kg/m <sup>2</sup>	2-D
O <sub>3</sub> column	col_o3_ave	kg/m <sup>2</sup>	2-D
O column	col_o_ave	kg/m <sup>2</sup>	2-D
OCS column	col_ocs_ave	kg/m <sup>2</sup>	2-D
SO <sub>2</sub> column	col_so2_ave	kg/m <sup>2</sup>	2-D
SO column	col_so_ave	kg/m <sup>2</sup>	2-D
Solar Flux (SW) at the top of the atmosphere	tops_ave	W/m <sup>2</sup>	2-D
Atmospheric mean molar mass	mmean_ave	g/mol	3-D
Atmospheric density	rho_ave	kg/m <sup>3</sup>	3-D
V-hourly RMS Atmospheric density	rho_rms	kg/m <sup>3</sup>	3-D
V-day to day RMS Atmospheric density	rho_dtd	kg/m <sup>3</sup>	3-D
Atmospheric temperature	temp_ave	K	3-D
V-hourly RMS Atmospheric temperature	temp_rms	K	3-D
V-day to day RMS Atmospheric temperature	temp_dtd	K	3-D
Zonal wind (West-Est)	vitv_ave	m/s	3-D
V-hourly RMS Zonal wind (West-Est)	vitv_rms	m/s	3-D
V-day to day RMS Zonal wind (West-Est)	vitv_dtd	m/s	3-D
Meridional wind (South-North)	vitv_ave	m/s	3-D
V-hourly RMS Meridional wind (South-North)	vitv_rms	m/s	3-D
V-day to day RMS Meridional wind (South-North)	vitv_dtd	m/s	3-D
Vertical wind (down-up)	vitwz_ave	m/s	3-D
V-hourly RMS Vertical wind (down-up)	vitwz_rms	m/s	3-D
V-day to day RMS Vertical wind (down-up)	vitwz_dtd	m/s	3-D
CO <sub>2</sub> volume mixing ratio	co2_ave	mol/mol	3-D
CO volume mixing ratio	co_ave	mol/mol	3-D

Table 1: Variables stored in database mean data files

H <sub>2</sub> volume mixing ratio	h2_ave	mol/mol	3-D
H <sub>2</sub> O volume mixing ratio	h2o_ave	mol/mol	3-D
H <sub>2</sub> O liquid volume mixing ratio	h2oliq_ave	mol/mol	3-D
H <sub>2</sub> SO <sub>4</sub> volume mixing ratio	h2so4_ave	mol/mol	3-D
H <sub>2</sub> SO <sub>4</sub> liquid volume mixing ratio	h2so4liq_ave	mol/mol	3-D
H volume mixing ratio	h_ave	mol/mol	3-D
HCl volume mixing ratio	hcl_ave	mol/mol	3-D
He volume mixing ratio	he_ave	mol/mol	3-D
N <sub>2</sub> volume mixing ratio	n2_ave	mol/mol	3-D
O <sub>2</sub> volume mixing ratio	o2_ave	mol/mol	3-D
O <sub>3</sub> volume mixing ratio	o3_ave	mol/mol	3-D
O volume mixing ratio	o_ave	mol/mol	3-D
OCS volume mixing ratio	ocs_ave	mol/mol	3-D
SO <sub>2</sub> volume mixing ratio	so2_ave	mol/mol	3-D
SO volume mixing ratio	so_ave	mol/mol	3-D
Electron volume mixing ratio	elec_ave	mol/mol	3-D
SW net flux (positive downward) at given altitude	SWnet_ave	W/m <sup>2</sup>	3-D
LW net flux (positive upward) at given altitude	LWnet_ave	W/m <sup>2</sup>	3-D

Table 2: Variables stored in database altitude wise standard deviation data files.

Altitude-wise variability (3-D RMS)	ARMS symbol	type (Day to day or hourly)	units
Atmospheric Pressure	pressure_arms	Hourly	Pa
Atmospheric Pressure	pressure_adtd	Day to day	Pa
Atmospheric Density	rho_arms	Hourly	kg/m <sup>3</sup>
Atmospheric Density	rho_adtd	Day to day	kg/m <sup>3</sup>
Atmospheric Temperature	temp_arms	Hourly	K
Atmospheric Temperature	temp_adtd	Day to day	K
Zonal wind (West-Est)	vitv_arms	Hourly	m/s
Zonal wind (West-Est)	vitv_adtd	Day to day	m/s
Meridional wind (South-North)	vitw_arms	Hourly	m/s
Meridional wind (South-North)	vitw_adtd	Day to day	m/s
Vertical wind (down-up)	vitwz_arms	Hourly	m/s
Vertical wind (down-up)	vitwz_adtd	Day to day	m/s

The EOF datafiles (eof) contain normalized, multi-dimensional EOFs of zonal wind, meridional wind, atmospheric temperature and surface pressure as well as some normalization factors, eigenvalues and principal component model coefficients. It is recommended that you use the software supplied to access and exploit the data in these files.

## 2.4 Database Grid Structure

### 2.4.1 Horizontal Structure

Variables in the database are stored on the grid on which they are obtained from the general circulation model (GCM) runs: a regular, equispaced horizontal  $96 \times 97$  in East longitude $\times$ latitude. Longitudes thus range from  $-176.25^\circ$  to  $180.00^\circ$  in steps of  $3.75^\circ$  and latitudes from  $90^\circ$  to  $-90^\circ$  in steps of  $1.875^\circ$ .

## 2.4.2 Vertical Structure

Variables in the database are stored on the same vertical grid on which they are computed. This vertical coordinate is a hybrid coordinate in which vertical levels  $l$  are at pressure  $P$ :

$$P(l) = aps(l) + bps(l) \times P_S \quad (1)$$

where  $P_S$  is surface pressure. Coefficients  $aps(l)$  and  $bps(l)$  are respectively hybrid pressure and hybrid sigma levels.

In its present form the database extends over 90 levels and variables in the datafiles are split between lower atmosphere ( $l = 1, \dots, 50$ ) and thermosphere ( $l = 51, \dots, 90$ , for `thermo` datafiles). This is due to the fact that only the thermosphere is affected by solar EUV input, unlike the lower part of the atmosphere. In order to reconstruct a column of data, one must thus extract the first 50 levels from 'lower atmosphere' datafiles and obtain the following 40 from the corresponding 'thermosphere' datafiles.

Note that  $aps$  and  $bps$  are prescribed coefficients such that near the surface (small values of  $l$ ) levels are essentially terrain-following sigma coordinates, whereas at high altitude (large values of  $l$ ) the vertical levels are pressure levels. Values of coefficients  $aps$  and  $bps$  are given in Table 3, along with corresponding pseudo-altitude which are the approximative heights above local surface of the corresponding layer (computed using a surface pressure of  $9 \cdot 10^6$  Pa and a scale height of 10 km; it is thus a rough estimate and particularly inaccurate in the upper atmosphere above 60 km because it does not account for actual changes in temperatures and scale height there).

Layer	hybrid pressure level aps (Pa)	hybrid sigma level bps	Typical pressure (Pa)	pseudo-altitude above surface (km)
1	328,08	0,99934	8994388	0,006237
2	1633,1	0,99674	8972293	0,03083
3	5141,5	0,98975	8912891	0,09725
4	12500	0,97517	8789030	0,2372
5	25142	0,95036	8578382	0,4797
6	44323	0,9132	8263123	0,8542
7	70973	0,8624	7832573	1,389
8	105490	0,79774	7285150	2,113
9	147530	0,72025	6629780	3,056
10	195780	0,6323	5886480	4,245
11	247920	0,53746	5085060	5,709
12	300770	0,44011	4261760	7,475
13	350590	0,34502	3455770	9,571
14	393590	0,25669	2703800	12,02
15	426340	0,17882	2035720	14,86
16	445860	0,11521	1482750	18,03
17	450990	0,068256	1065294	21,33
18	442330	0,035953	765907	24,63
19	417930	0,014785	550995	27,93
20	365060	0,0034636	396232	31,22
21	283180	0,00013352	284381	34,54
22	203790	2,6913E-08	203790	37,87
23	146020	1,3467E-15	146020	41,21
24	104630	8,1189E-30	104630	44,54
25	74969	0	74969	47,87
26	53718	0	53718	51,21
27	38490	0	38490	54,54
28	27580	0	27580	57,87
29	19762	0	19762	61,21
30	14160	0	14160	64,54
31	10146	0	10146	67,87

..... table continued next page .....

Layer	hybrid pressure level aps (Pa)	hybrid sigma level bps	Typical pressure (Pa)	pseudo-altitude above surface (km)
32	7269,9	0	7269,9	71,21
33	5209,1	0	5209,1	74,54
34	3732,5	0	3732,5	77,87
35	2674,4	0	2674,4	81,21
36	1916,3	0	1916,3	84,54
37	1373,1	0	1373,1	87,87
38	983,87	0	983,87	91,21
39	704,98	0	704,98	94,54
40	505,14	0	505,14	97,87
41	361,95	0	361,95	101,2
42	259,35	0	259,35	104,5
43	185,83	0	185,83	107,8
44	133,15	0	133,15	111,2
45	95,408	0	95,408	114,5
46	65,575	0	65,575	118,2
47	39,879	0	39,879	123,2
48	21,278	0	21,278	129,5
49	10,202	0	10,202	136,9
50	4,5292	0	4,5292	145,0
51	2,3732	0	2,3732	151,4
52	1,4441	0	1,4441	156,4
53	0,87589	0	0,87589	161,4
54	0,53125	0	0,53125	166,4
55	0,32222	0	0,32222	171,4
56	0,19544	0	0,19544	176,4
57	0,11854	0	0,11854	181,4
58	0,071897	0	0,071897	186,4
59	0,043608	0	0,043608	191,4
60	0,02645	0	0,02645	196,4
61	0,016042	0	0,016042	201,4
62	0,0097303	0	0,0097303	206,4
63	0,0059017	0	0,0059017	211,4
64	0,0035796	0	0,0035796	216,4
65	0,0021711	0	0,0021711	221,4
66	0,0013168	0	0,0013168	226,4
67	0,00079871	0	0,00079871	231,4
68	0,00048444	0	0,00048444	236,4
69	0,00029383	0	0,00029383	241,4
70	0,00017822	0	0,00017822	246,4
71	0,00010809	0	0,00010809	251,4
72	6,5562E-05	0	6,5562E-05	256,4
73	3,9765E-05	0	3,9765E-05	261,4
74	2,4119E-05	0	2,4119E-05	266,4
75	1,4629E-05	0	1,4629E-05	271,4
76	8,8729E-06	0	8,8729E-06	276,4
77	5,3817E-06	0	5,3817E-06	281,4
78	3,2641E-06	0	3,2641E-06	286,4
79	1,9799E-06	0	1,9799E-06	291,4
80	1,201E-06	0	1,201E-06	296,4
81	7,2834E-07	0	7,2834E-07	301,4
82	4,4164E-07	0	4,4164E-07	306,4
83	2,6785E-07	0	2,6785E-07	311,4
84	1,6248E-07	0	1,6248E-07	316,4
85	9,8571E-08	0	9,8571E-08	321,4
86	5,9802E-08	0	5,9802E-08	326,4
..... table continued next page .....				



Layer	hybrid pressure level aps (Pa)	hybrid sigma level bps	Typical pressure (Pa)	pseudo-altitude above surface (km)
87	3,6279E-08	0	3,6279E-08	331,4
88	2,2006E-08	0	2,2006E-08	336,4
89	1,3345E-08	0	1,3345E-08	341,4
90	8,093E-09	0	8,093E-09	346,4

Table 3: Vertical grid in the GCM and in the VCD datafiles

### 2.4.3 Temporal Structure

In order to store the diurnal behaviour of variables, data from the Global Climate Model was processed to be stored along one full Venus day, which lasts 10087200 s. Time evolution of variables is included in the datafiles where values at 24 times of day are stored. Venusian hours are defined as being 1/24th of a Venus day. To avoid confusion, we do not use (or define) venusian minutes or seconds: any venusian time of day is always given as a fraction of a Venus day or in venusian hours and decimal fractions thereof (e.g. *time* = 18.5 hours means 18 hours and a half).

The database reference time is Venus Universal Time (which is simply “prime meridian time”, i.e. the local time at 0° longitude) and data is stored every venusian hours, i.e. from 0.5 to 23.5 hours in steps of 1 hour. Note that all times are expressed in True Solar Time (the Sun is highest at noon) and not Mean Solar Time (see the description of the Equation Of Time in Appendix A).

The Local True Solar Time  $LTST$  at a given East longitude  $lon$  (expressed in degrees) may easily be computed from the Local True Solar Time at longitude zero  $LTST_0$  (i.e. Venus Universal Time) by the following formula:

$$LTST = LTST_0 - lon/15 \quad (2)$$

## 3 EUV and cloud albedo scenarios in the VCD

Combinations of solar EUV and cloud albedo scenarios are provided.

- On the one hand, the solar conditions describe variations in the Extreme UV input which control the heating of the atmosphere above around 100 km, which typically varies on an 11-year cycle. Depending on the scenarios, either fixed (i.e. constant over time) solar maximum, average and/or minimum conditions are provided, or deduced from Earth date for a realistic solar EUV, i.e. as observed at the desired date. The minimum, average and maximum EUV scenarios correspond, respectively, to a value of E10.7 factor of 70, 140 and 300 sfu (solar flux unit : 1 sfu =  $10^{-22}$  W.m<sup>-2</sup>.Hz<sup>-1</sup>). The VCD user has also the possibility to set the solar EUV at a desired value as input of the `call_vcd` routine. A linear interpolation is used between minimum and average or average and maximum solar EUV scenarios.
- On the other hand, there is still uncertainties about the Venus solar heating rate. As shown in the work of *Lee et al.*, 2019 (Astronomical J. 158, doi:10.3847/1538-3881/ab3120), this Venus solar heating rate is directly related to the cloud albedo through a radiative transfer model. Therefore the heating rate is calculated for each scenario in order to simulate the albedo measured by Hubble and MESSENGER mission during its flight over Venus. Three kinds of cloud albedo scenarios are proposed. The Standard cloud albedo is calculated using our standard heating rate profile, based on *Haus et al.* (2014). The low cloud albedo scenario is calculated using the maximum ratio profile from Fig. 13 (right panel) in *Lee et al.* (2019), with a maximum ratio of 50% (instead of 40% in *Lee et al.* (2019)). The high cloud albedo scenario is calculated by decreasing the standard heating rate profile by 30%. The three profiles are shown on Figure 1.

How these EUV and cloud albedo scenarios are taken into account is detailed in the following subsections.

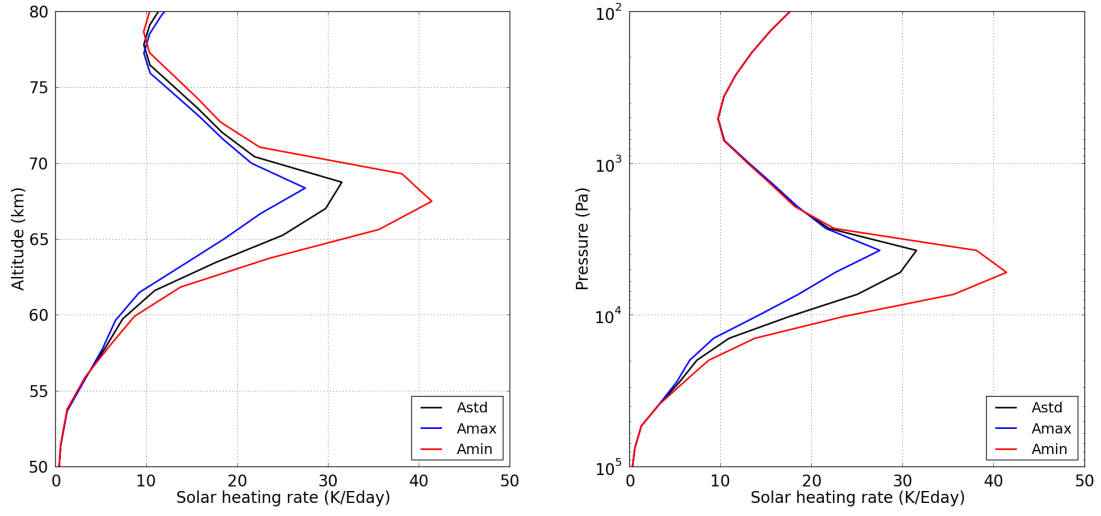


Figure 1: Solar heating rate profiles at local noon time at Equator.

### 3.1 The EUV Scenarios

The radiative output of the Sun is known to vary at different timescales: for example, due to the solar flares (timescale of minutes to hours), to the solar rotation (27 days) or to the magnetic cycle of the Sun (11 years, the so-called solar cycle). This last variability, first detected counting the number of sunspots, is more important in the UV (than in the visible) region of the solar spectrum, as a variability of about a factor 2 in the total EUV irradiance (below 120 nm) with the solar cycle is found. Although the UV spectral region represents a small contribution to the total solar energy, the UV radiation is the major heating source of the Venusian upper atmosphere (above about 100 km). So, its variability has a strong impact over the thermospheric temperatures.

In three VCD simulations, the EUV flux is set to constant values chosen to provide an estimation of the average and extreme (minimum and maximum) fluxes. The “**Solar maximum**” scenario corresponds to conditions when the the solar cycle is at its maximum activity<sup>3</sup> at 1 AU (Astronomical Unit) around 300 sfu (1 sfu =  $10^{-22} \text{ W.m}^{-2}.\text{Hz}^{-1}$ ) and the UV emission is highest. For such conditions, a rise in temperatures in the thermosphere, induced by a more intense UV heating, is expected (and obtained). On the opposite, the “**Solar minimum**” scenario is appropriate for conditions when the Sun is at its minimum activity (approximate value of E10.7 at 1 AU around 70 sfu). In this case, UV emission is low, so a lower temperature due to a lower UV heating is expected. The “**Solar average**” scenario is an intermediate situation between maximum and minimum solar activity (approximate value for E10.7 at 1AU of 140). Figure 2 displays mean temperature profiles corresponding to the three EUV scenarios.

It is also possible to call the VCD with a realistically (daily) evolving EUV input, taken from available E10.7 measurements, from February 1947 to October 2020. The user has also the possibility to give as input the desired value of E10.7. In both these cases a linear interpolation is performed between the EUV scenarios. The validity of this interpolation was tested with additional GCM simulations performed with E10.7 values of 100 and 200 sfu (see Validation document).

### 3.2 The Cloud albedo Scenarios

As said before, the uncertainties about the Venus cloud albedo and therefore on the solar heating rate motivates to provide for the VCD three kinds of cloud albedo scenarios :

<sup>3</sup>E10.7 is a proxy for the integrated extreme ultraviolet flux from 1 to 105 nm expressed in the same unit than the canonic F10.7 index (full-disc solar emission at the 10.7 nm wavelength, expressed in units of  $10^{-22} \text{ W m}^{-2} \text{ Hz}^{-1}$ ) originally measured by Earth-based radio-observatory when no solar spectral EUV fluxes is available.

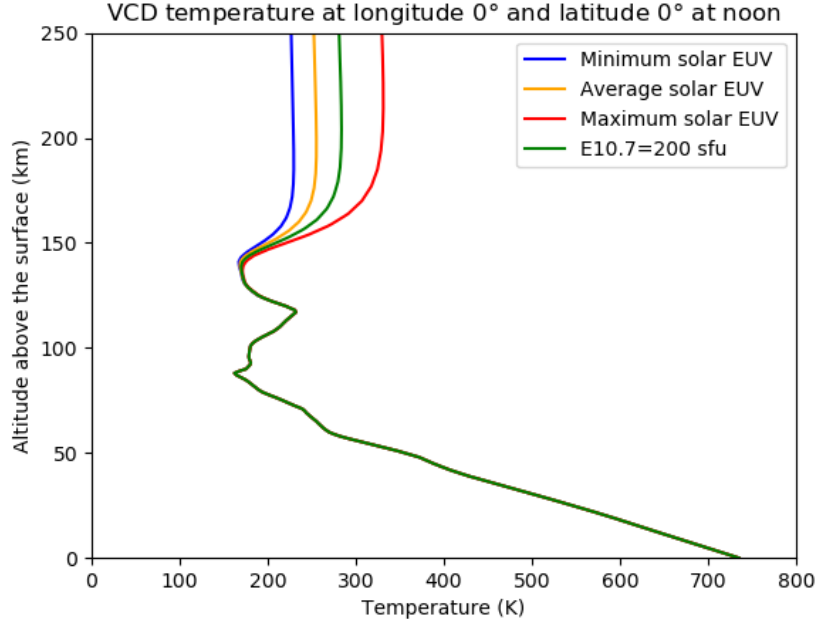


Figure 2: Illustration of temperature profiles at a given location provided by the VCD (with the same “standard” cloud albedo scenario) with the Extreme Ultraviolet (EUV) fluxes corresponding to the solar EUV minimum (min), average (ave) and maximum (max) scenarios, as well as with a chosen E10.7 value (here E10.7=200 sfu).

1. The **“Standard cloud albedo” (standard)** scenario, which is a simulation ran using the latest version of the LMD Global Climate Model (GCM) forced by a solar heating rate calculated from standard cloud albedo observations with 3 solar EUV conditions: solar minimum EUV flux, solar ave, solar max.
2. The **low** scenario corresponds to a low cloud albedo conditions, forced by a solar heating rate calculated from low cloud albedo observations with 3 solar EUV conditions: solar min, solar ave, solar max.
3. The **high** scenario corresponds to high cloud albedo conditions, forced by a solar heating rate calculated from high cloud albedo observations with 3 solar EUV conditions: solar min, solar ave, solar max.

Figure 3 displays mean zonal wind profiles corresponding the three cloud albedo scenarios.

This leads to a total of 15 different scenarios because there are 3 cases for the cloud albedo scenarios and, for each one, 5 possible cases for the solar EUV scenarios. All the combinations are available as long as the corresponding data have been downloaded.

## 4 Technical description of Methods Used to Retrieve Data from the VCD

### 4.1 Temporal Interpolations

Once the time of day at which data is requested is known (see Appendix A for details on how these can be computed for a given Julian date), linear interpolation between datasets of encompassing venusian hours are used to evaluate data on the GCM grid.

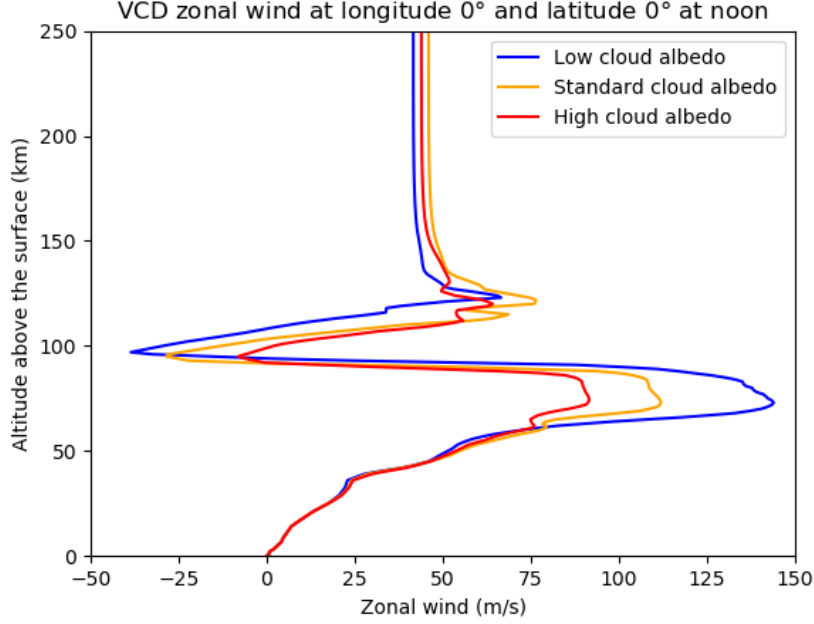


Figure 3: Illustration of zonal wind profiles at a given location provided by the VCD (with the same “Average” solar EUV scenario) with the Cloud albedo corresponding to the Standard, low and high albedo scenarios.

## 4.2 Spatial Interpolations

### 4.2.1 Horizontal Interpolation

To compute the values of variables at a given location (which is not on the GCM grid), one needs to first build a virtual column of data along what would have been GCM levels. This profile is obtained by horizontal interpolation of encompassing grid values. To be specific, values of variables are obtained from encompassing grid values using bilinear interpolation, except for density  $\rho$  (and similar bi-products such as associated RMS or surface pressure), which is obtained using bilinear interpolation of  $\log(\rho)$ .

### 4.2.2 Vertical Interpolation

In order to compute the value of meteorological variables at a given height (or pressure), one must first identify the altitudes (and pressure) of the vertical levels of the profile. The pressure of each level, which correspond to the middle of the layer, is given by equation (1) using the surface pressure which has been interpolated at the desired point with a bi-logarithmic interpolation, as said in the previous section. Then, calculating the altitude of each level is done by integrating the hydrostatic equation over the column:

$$\frac{\partial P}{\partial z} = -g \cdot \rho = -g \cdot \frac{P}{R \cdot T} \quad (3)$$

from which the relation between an increment in altitude  $\delta z$  and average (over  $\delta z$ ), gas constant  $R_m$ , atmospheric temperature  $T_m = \delta T / \delta \log T$ , and log pressure increment  $\delta \log P$  can be derived:

$$\delta z = -\frac{R_m}{g_m} \cdot T_m \cdot \delta \log P \quad (4)$$

This relation can be used to compute the altitudes of all the levels of a profile for which pressure  $P(l)$  (or equivalently sigma levels  $\sigma(l) = P(l)/P_S$ , where  $P_S$  is surface pressure), atmospheric temperature  $T(l)$  as well as gas constant  $R(l)$  at all  $l$  levels are known. Moreover, as the database extends to high altitudes, it must also taken into account that gravity  $g$  varies as the inverse square of the distance to the center of

the planet, i.e. that gravity at altitude  $z$  can be approximated as  $g(z) = g_0.R_0^2/(R_0 + z)^2$ , where  $z$  is the altitude above the reference sphere  $R_0$ , and  $g_0 = 8.87\text{m/s}^2$  is the gravity at  $R_0 = 6051848$  m above the center of Venus.

The altitude of the first atmospheric layer is thus obtained using:

- Gravity  $g(l = 1)$  is approximated to be that at the surface (of known orography, i.e. altitude above reference sphere,  $h$ ):

$$g(l = 1) = g_0.R_0^2/(R_0 + h)^2$$

- The altitude of the first layer is then simply

$$z(l = 1) = -[R(l = 1)/g(l = 1)].T(l = 1). \log(\sigma(l = 1))$$

From there, the altitudes of layers are recursively determined; i.e. for layer  $l + 1$ :

- Approximate  $g(l + 1)$  using the altitude  $z(l)$  of previous layer:

$$g(l + 1) = g_0.R_0^2/(R_0 + h + z(l))^2$$

- Compute altitude of level  $l + 1$ :

$$z(l + 1) = z(l) - [R(l + 1)/g(l + 1)].T_m(l, l + 1). \log[\sigma(l + 1)/\sigma(l)]$$

Once the altitudes  $z(l)$  of layers are known, then the value of a variable  $X$  at a given altitude above the surface  $z_s$  may be obtained from linear interpolation of values at encompassing grid points. For variables which are, to first order, exponentially distributed in altitude (e.g. pressure, density and associated variables) then the interpolation must naturally be performed on the logarithm of these values.

#### 4.2.3 Specific Treatments of Vertical Interpolation

Interpolation, as explained above, is feasible when the altitude (or pressure) at which the value of a variable is sought falls in the range of the VCD grid. In the more extreme cases, either above the topmost atmospheric layer or below the lowest one, different approaches must be used. These are detailed here.

##### Above the VCD Topmost Layer

If above the VCD topmost layer  $l = L$ , then the concentration of component  $j$  of the neutral atmosphere at the altitude  $z_s$  is:

$$n_j(z_s) = n_j(L) \frac{T(L)}{T(z_s)} \exp\left\{-\int_{z(L)}^{z_s} \frac{g(z)}{R_j T(z)} dz\right\} \quad (5)$$

where  $n_j$  and  $R_j$  are the concentration and gas constant of species  $j$  (in  $\text{m}^2 \text{s}^{-2} \text{K}^{-1}$ ). This results from the fact that, in the upper thermosphere and exosphere the collision rate of molecules and atoms is so small that each species tend to have its own scale height.

Assuming that the gravity decreases inversely to the square of planetocentric radial coordinate as:

$$g(z) = g(L) \left( \frac{R_0 + h + z(L)}{R_0 + h + z} \right)^2 \quad (6)$$

Neglecting the variation of neutral temperature with altitude  $T(z > z(L)) \cong T(L)$  the pressure  $P(z_s)$  at altitude  $z_s$  is:

$$P(z_s) = \sum_j P_j(L) \exp\left\{-\frac{g(L)}{R_j T(L)} \frac{(R_0 + h + z(L))(z_s - z(L))}{R_0 + h + z_s}\right\} \quad (7)$$

With  $P_j(L)$  the partial pressure of species  $j$  in the top layer of the model. When the vertical coordinate is requested in pressure units the conversion into meters is implemented numerically by the bisection method with 5% precision. Therefore, small discrepancies between cases with vertical coordinate defined in [Pa] and [m] may happen.

The pressure and volume mixing ratios of neutral species are computed according to these relations. The density and its variabilities (Venus day-to-day or Venus hourly) are extrapolated above surface at altitude  $z_s$  as follows :

$$a(z_s) = a(L) \sum_j \frac{P(L)}{R_j T(L)} \frac{vmr_j(L)}{\rho(L)} \exp\left\{-\frac{g(L)}{R_j T(L)} \frac{(R_0 + h + z(L))(z_s - z(L))}{R_0 + h + z_s}\right\} \quad (8)$$

where  $a$  is either the density  $\rho$ , the day to day density variability  $\rho_{dtd}$  or the hourly density variability  $\rho_{RMS}$ .

However, all other variables (e.g. winds) are not extrapolated and are assumed to remain constant above the topmost layer.

### Below the Lowest VCD Layer

When below the lowest VCD layer  $l = 1$  (which typically lies at 6 m above the surface, see table in Section 2.4.2), then pressure  $P(z_s)$  at altitude  $z_s$  above the surface is computed using a hydrostatic vertical profile, i.e.:

$$P(z_s) = P(1) \exp[(z(1) - z_s).g(z_s)/(R(1).T(1))] \quad (9)$$

where  $P(1)$ ,  $R(1)$  and  $T(1)$  are the pressure, gas constant and temperature at the first layer,  $z(1)$  its altitude and  $g(z_s)$  is gravity at altitude  $z_s$ . This formulation enables recovery of the value of surface pressure  $P_s$  when  $z_s = 0$ .

Density  $\rho$ , as well as the RMS of these two variables, are treated similarly.

Temperature is linearly interpolated with surface temperature :

$$T(z_s) = T_{surf} + (T(1) - T_{surf}) * \frac{z_s}{z(1)} \quad (10)$$

where  $T_{surf}$  is the surface temperature. This formulation enables recovery of the value of surface temperature when  $z_s = 0$ .

Winds are extrapolated using a logarithmic profile such that :

$$u(z_s) = u(1) * \frac{\ln \frac{z_s}{z_0}}{\ln \frac{z(1)}{z_0}} \quad (11)$$

where  $u$  is the zonal wind,  $z_0$  is the roughness length assumed constant and equal to  $z_0 = 0.01$  m. The same extrapolation is done for meridional wind  $v$  and vertical wind  $w$ . The winds under  $z_0$  are set to zero.

Other variables (except horizontal winds and atmospheric temperature) are not extrapolated and are taken to be constant from the middle of the first atmospheric level down to the surface.

## 5 Variability Models in the VCD

Firstly, the variability in the VCD is modeled with RMS calculated from the GCM outputs. There are two type of RMS: the day to day RMS which represents the variability of a variable from one day to an other and the hourly RMS which represents the variability within one Venusian hour. The hourly RMS includes the day to day RMS. Both are computed on the GCM grid. Since the vertical coordinate is essentially (see Section 2.4.2) a pressure coordinate, the obtained RMS values are evaluated at constant pressure.

Apart from this pressure-wise RMS, the VCD also provides altitude-wise RMS, which is computed in the same way, but on time series of GCM outputs which have been interpolated on a fixed altitude vertical

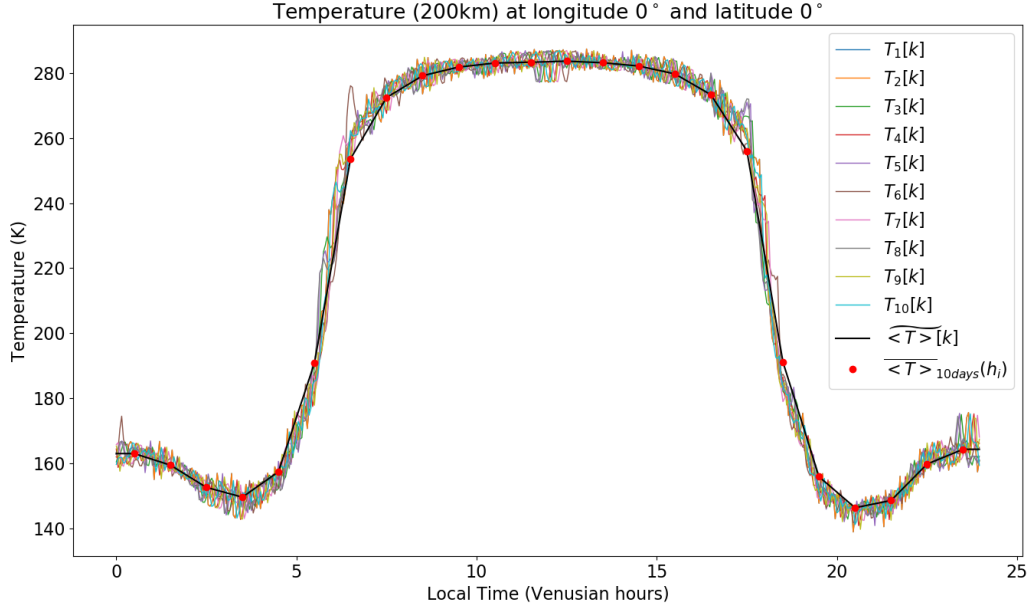


Figure 4: Atmospheric temperature diurnal cycles that intervenes in the calculation of the VCD hourly RMS of temperature.

coordinate grid. For some variables, which are dependent on pressure, e.g. density, the difference between pressure-wise and altitude-wise RMS can be quite significant.

Secondly, the variability can be modeled adding large scale or small scale perturbations to the signal, as we will see in the next sections.

### 5.1 Hourly RMS of Variables

The VCD provides the Venus hourly variability of variables computed from the outputs of the Global Circulation Model (GCM). This hourly RMS of variable  $X$  is computed over each of the 24 venusian hours  $h_i$  from 0.5 to 23.5 with 1 hour steps as:

$$RMS\_HOURLY\_X(h_i) = \sqrt{\langle (X - \langle X \rangle)^2 \rangle^{h_i}} \quad (12)$$

where  $\langle X \rangle$  is the climatological average of the variable  $X$  over the 10 GCM days and  $\bar{X}$  is the moving average over a centered interval of 1 hour such that:

$$\langle X \rangle = \frac{1}{N_{days}} \sum_{j=1}^{N_{days}} X_j \quad (13)$$

and:

$$\bar{X}^{h_i} = \frac{1}{N_p(h_i)} \sum_{k=1}^{N_p(h_i)} X[k] \quad (14)$$

with  $N_p(h_i)$  the number of GCM samples of  $X(t)$  for  $t \in [h_i - 0.5, h_i + 0.5]$  and  $N_{days} = 10$  the number of days of GCM simulation. So we compute the instantaneous standard deviation of the set of 10 GCM days then, instead of sampling directly one value per hour which could be not representative for the whole hour, we compute the average over one hour centered on the time we are interested in. Figure 4 shows what the different terms of the equations above represent in the case of exospheric temperature.

Note that this RMS takes into account any large time scale variations from one hour to the next one, so the RMS obtained  $RMS\_HOURL\_X(h_i)$  represents the variability of a variable within one hour around a given time of day  $h_i$ . Note also that this RMS already includes the day-to-day variability so there is no need to add the last to the first to get an estimation of the total variability.

## 5.2 Day-to-Day RMS of Variables

The VCD provides the day-to-day variability of variables computed from the outputs of the Global Circulation Model (GCM). This day-to-day RMS of variable  $X$  is computed over each of the 24 venusian hours  $h_i$  from 0.5 to 23.5 with 1 hour steps as:

$$RMS\_DTD\_X(h_i) = \sqrt{\langle (\overline{X}^{h_i} - \langle \overline{X}^{h_i} \rangle)^2 \rangle}^{h_i} \quad (15)$$

with the same conventions as in the previous subsection (namely the section 5.1). The moving averages which intervene inside the square root are used as filters on the variable  $X$  to remove the waves. The RMS thus obtained  $RMS\_DTD\_X(h_i)$  represents the variability of a variable from one day to the next at a given time of day  $h_i$ . Note that the day-to-day variability is taken into account in the hourly RMS.

## 5.3 The Large-Scale Variability Model

In the VCD, data are stored at 24 local times of day. Although this captures the main diurnal components of variability, any intra-hour or day-to-day variations are averaged out. Thus there is a need to simulate this variability, especially if the user wishes to produce an ensemble of realizations of a variable at a particular local time of day which covers a realistic range of variability.

For this purpose, we store the 3D variability of pressure, temperature and winds using Empirical Orthogonal Functions (EOFs) over the different simulated days.

### 5.3.1 Calculation of EOFs and PCs

Consider a time series of longitude-pressure vectors of zonal wind, meridional wind, temperature and surface pressure deviations from the climatological means for one GCM day at  $M$  discrete time points and on  $L$  spatial points respectively noted  $\delta u(\phi, p, t)$ ,  $\delta v(\phi, p, t)$ ,  $\delta T(\phi, p, t)$  and  $\delta p^*(\phi, t)$ . We form a time series of vectors  $\mathbf{D}(t)$  such as:

$$\mathbf{D}(t) = (\hat{u}(\phi, p, t), \hat{v}(\phi, p, t), \hat{T}(\phi, p, t), \hat{p}^*(\phi, t)) \quad (16)$$

The  $\hat{\cdot}$  denotes a removal of the series mean and normalization of variance operator:

$$\overline{\delta u} = \frac{1}{ML} \sum_{m=1}^M \sum_{l=1}^L \delta u_{ml} \quad (17)$$

$$\hat{u}_{ml} = \frac{\delta u_{ml} - \overline{\delta u}}{\sqrt{\frac{1}{ML} \sum_{m=1}^M \sum_{l=1}^L (\delta u_{ml} - \overline{\delta u})^2}} \quad (18)$$

where the  $m$  denotes time and the  $l$  denotes spatial point. Hence, with this normalization, the entire time series of  $\mathbf{D}(t)$  has a variance equal to 1 and a mean equal to 0.

We then form the  $(N \times N)$  covariance matrix,  $\mathbf{C}$ , such that:

$$\mathbf{C} = \frac{1}{N} \mathbf{D} \mathbf{D}^T \quad (19)$$

where  $N = (\text{number of horizontal points}) \times (3 \times \text{number of vertical points} + 1)$ ,  $\mathbf{D}$  is the  $(N \times M)$  matrix whose rows are the vectors  $\mathbf{D}(t)$  and the superscript  $T$  indicates the transpose.



The matrix  $\mathbf{C}$  is real symmetric and we can find the eigenvectors and eigenvalues and order them in decreasing eigenvalue magnitude. We note that if  $\mathbf{E}_i$  is the  $i$ th eigenvector then:

$$|\mathbf{E}_i| = 1 \quad (20)$$

and

$$\sum_{i=1}^N \lambda_i = 1 \quad (21)$$

where  $|\cdot|$  is the Euclidean Norm and  $\lambda_i$  is the  $i^{\text{th}}$  eigenvalue.

The  $i$ th principal component (PC) at time  $m$ ,  $p_{mi}$  is defined as

$$p_{mi} = \sum_{n=1}^N D_{mn} E_{ni} = \mathbf{D} \cdot \mathbf{E}_i \quad (22)$$

and we note the result

$$\frac{1}{MN} \sum_{m=1}^M (p_{mi})^2 = \lambda_i \quad (23)$$

Each principal component has 480 values during one day.

### 5.3.2 Storage of variability with EOFs

As decribed above, using PCs and EOFs provides a mean to store the variability. In practice, any of the fields  $\delta X$  (cf. section 5.3.1) can be decomposed as:

$$\delta X_j(\theta, \phi, p, t) = AVE_j^X(\theta, \phi, p) + NORM_j^X(\phi) \sum_i PC_{j,i}(t) \times EOF_{j,i}^X(\theta, \phi, p)$$

where  $AVE^X$  is simply an average over the whole time series (*i.e.* one Venusian day),  $NORM^X$  is a normalization factor,  $PC_{j,i}$  are the principal components and the  $EOF_{j,i}^X$  are the empirical orthogonal functions (note that these are function of spatial coordinates only). The superscript  $X$  simply denotes a dependency on the variable  $X$  – which is the case of  $AVE$ ,  $NORM$  and the  $EOF_i$  but not the  $PC_i$  – while the subscript  $j$  denotes a dependency on the GCM day.

By design the individual  $PC_{j,i}$  and  $EOF_{j,i}$  are ordered along  $i$  from the “most significant” (contains the most information) to the “least significant” (contains the less information), so that truncating the series and keeping the  $N$  first EOFs will suffice to reconstruct the original signal. Therefore **this method allows to store several days of meteorological variability without having to store the huge amount of data initially produced by the model**. Nevertheless, to further limit the size of the final dataset (and also limit the size of the eigensystem to solve) the retained dataset consists of:

- Atmospheric temperature, zonal wind, meridional wind and surface pressure variables
- Fields sub-sampled on a reduced 24x24 longitude x latitude grid (a sixteenth of the computational GCM 96x96 grid), and for three-dimensional quantities, on all 90 vertical levels.
- A time sampling of 480 data points per day, which is sufficient to reconstruct an intra hourly variability (note that the daily evolution of variables is already include in the VCD datasets).
- Overall, this means that a single EOF contains  $24 \times 24 \times (3 \times 90 + 1) = 156096$  elements, hence the need to retain as few as possible.

Concerning the number  $N$  of EOFs suitable to retain to be able to realistically represent the original signal and its inherent variability, it is quite straightforward to test, by direct comparison between the original signal and truncated reconstructions, which value of  $N$  is sufficient. Our tests have shown that  $N=200$  is a reasonable choice.

### 5.3.3 Reconstructing a realistic variability

Rather than reconstructing the original signal, our goal is to represent the original variability to recreate a realistic one (in space and time). In order to do that, we have computed the EOFs for the difference between a chosen GCM day and the climatological day. Thus the average  $AVE_j^X(\theta, \phi, p)$  is the difference between the daily average of the chosen day and the daily average of the climatological day.

If we compute the EOFs for each GCM day  $j$  then randomly choose both the day number  $j$  within  $\llbracket 1; 10 \rrbracket$  and a time shift  $dt$  within  $\pm 0.5$  hour, the mean and the standard deviation of a sample of  $\tilde{X}$  defined as:

$$\tilde{X}(\theta, \phi, p, t) = X_{\text{VCD}}(\theta, \phi, p) + \delta X_j(\theta, \phi, p, t + dt)$$

will converge respectively toward the mean and the hourly RMS (cf. section 5.1) stored in the VCD.

However, store the EOFs for 10 GCM days for 9 scenarios is quite expensive (about 11 Go). *Because the variability is quite similar between the different days, we decide to store the EOFs for only one chosen day.* That implies to modify the formula in order to conserve the right mean value. Indeed, although the average of  $\delta X_j$  over  $dt$  and  $j$  is 0, nothing allows us to think the average over  $dt$  at a given  $j$  is 0 too. So we compute the perturbed fields  $\tilde{X}$  as follows:

$$\tilde{X}(\theta, \phi, p, t) = X_{\text{VCD}}(\theta, \phi, p) + \text{Pert}^X(\theta, \phi, p, t) \quad (24)$$

with

$$\begin{aligned} \text{Pert}^X(\theta, \phi, p, t) &= \delta X_j(\theta, \phi, p, t + dt) - \overline{\delta X_j}(\theta, \phi, p, t) \\ &= \text{NORM}^X(\phi) \sum_i (PC_i(t + dt) - \overline{PC_i}(t)) \times \text{EOF}_i^X(\theta, \phi, p) \end{aligned} \quad (25)$$

If we assume the variability at a given time is the same for every days, then the perturbed fields have a variability reduced by the variability contained in the 10 daily offsets  $AVE^X(\theta, \phi, p)$ . Nevertheless, the mean of an increasingly growing sample of perturbed events will still converge toward the VCD climatological mean.

The key point in this procedure is that, by construction, the spatio-temporal coherence of transient phenomena is preserved and thus the true internal variability of the original signal can be reproduced.

An illustration is shown in Figure 5 where we can see that the reconstructed perturbed temperature signal with the EOF method is bracketed by the unperturbed signal including the intra-hourly variability.

## 5.4 The Small-Scale Variability Model

The small-scale variability model simulates perturbations of density, temperature and wind due to the upward propagation of small-scale gravity waves. The model is based on the parameterization scheme used in the Mars Climate Database (MCD). This parametrization might be improved in future versions of the VCD. The details of this parametrization is given in (Forget et al., *J. Geophys. Res.* 104, 24,155-24,176, 1999). In the VCD case, these waves are generated at the top of the cloud convective layer, around 60 km above the surface.

The surface stress exerted by a vertically-propagating, stationary gravity wave can be written

$$\tau_0 = \kappa \rho_0 N_0 |\mathbf{v}_0| \sigma_0^2 \quad (26)$$

where  $\kappa$  is a characteristic gravity wave horizontal wave number,  $\rho_0$  is the surface density,  $N_0$  is the surface Brunt Väisälä frequency,  $\mathbf{v}_0$  is the surface vector wind and  $\sigma_0$  is a measure of the orographic variance. It represents the model sub-grid scale topographic variance. However in the VCD case, we set it to a constant value of 5 m. This empirical value is chosen such that the amplitude of the density wave lies in around 20%, as shown by the measurement from Venus Express with Torque experiment. Moreover, in the VCD case, instead of using the variable at the surface like in the MCD, we use the variables at the top

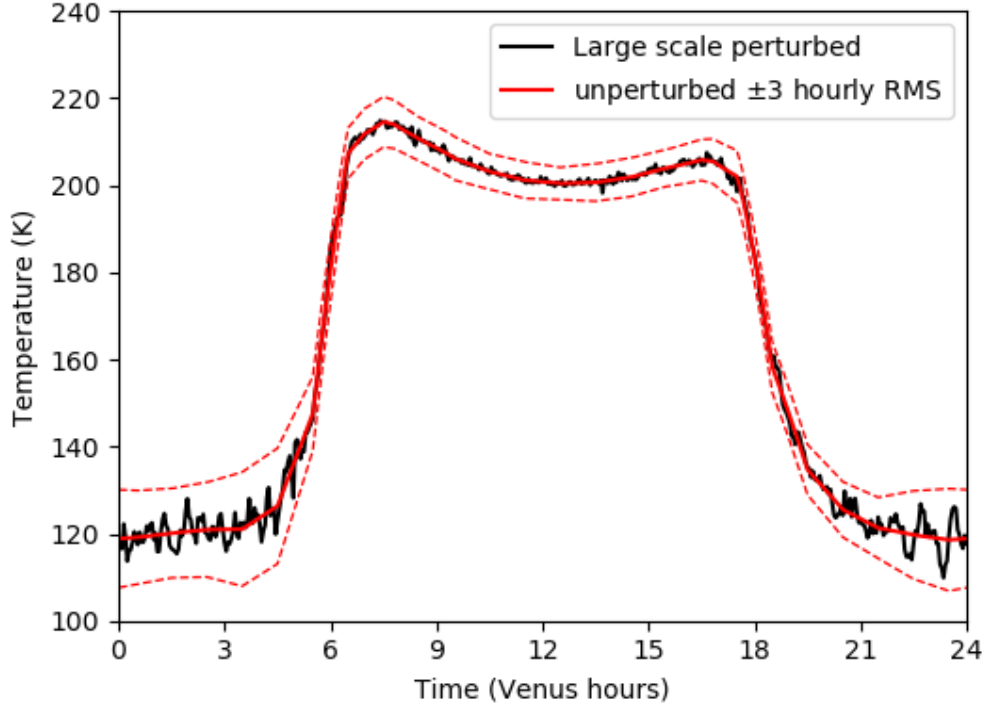


Figure 5: Example of atmospheric temperature diurnal cycle, computed with the VCD using the large scale perturbation with EOF method compared with the hourly RMS stored in the VCD at  $0^\circ\text{E}$ ,  $0^\circ\text{N}$  and 150 km of altitude, for a standard albedo and an average solar EUV flux.

of the cloud convective layer, which is assumed to be the  $27^{th}$  layer of the model. The surface stress (or in our case, the cloud stress) can be related to the gravity wave vertical isentropic displacement,  $\delta h$ , by

$$\tau = \kappa \rho N |\mathbf{v}| \delta h^2. \quad (27)$$

We then assume that the stress,  $\tau$ , above the surface is equal to that at the surface (or in the cloud for the VCD)  $\tau_0$ . This leads to an expression for the wave displacement  $\delta h$ , at height  $z$ ,

$$\delta h = \sqrt{\frac{\rho_0 N_0 |\mathbf{v}_0| \sigma_0^2}{\rho N |\mathbf{v}|}} \quad (28)$$

where  $\rho$ ,  $N$  and  $\mathbf{v}$  are the density, Brunt Väisälä frequency and wind vector at height  $z$ .

The gravity wave perturbation to a meteorological variable is calculated by considering vertical displacements of the form

$$\delta z = \delta h \sin \left( \frac{2\pi z}{\lambda} + \frac{2\pi Dist}{\lambda_{Dist}} + \phi_0 \right) \quad (29)$$

where  $\lambda$  is a characteristic vertical wavelength for the gravity wave,  $\lambda_{Dist}$  is the horizontal wavelength, such as  $\lambda_{Dist} = 100 \times \lambda$ , and  $\phi_0$  is a randomly generated surface phase angle. A set of points constituting a horizontal trajectory (varying latitude and longitude) is characterized by the dynamic distance  $Dist$ , representative for the wave propagation direction, and calculated between the trajectory's initial point ( $lat_0, lon_0$ ) and its following points ( $lat, lon$ ) (where  $lat$  and  $lon$  are respectively the latitude and longitude).  $Dist$  is expressed as follows:

$$Dist = 2R_0 \arcsin \left( \sqrt{\sin^2 \left( \frac{lat - lat_0}{2} \right) + \cos(lat_0) \cos(lat) \sin^2 \left( \frac{lon - lon_0}{2} \right)} \right) \quad (30)$$

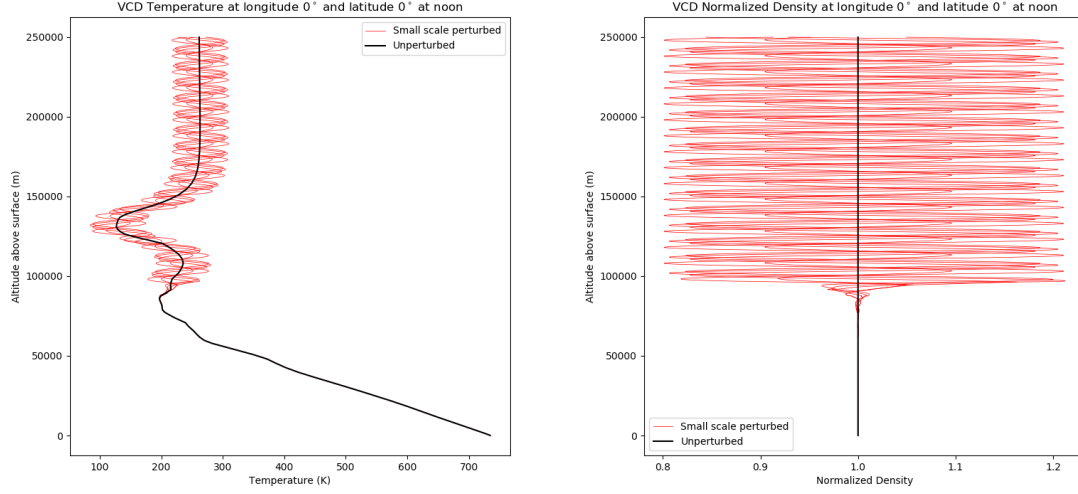


Figure 6: A series of several perturbations, all with the same wavelength, generated by the small scale variability model added to mean profiles of temperature and density from the VCD on a vertical trajectory. Left plot: Atmospheric temperatures (including unperturbed profile). Right plot: density deviations (i.e. perturbed over unperturbed density ratios). Profiles obtained at longitude  $0^\circ$  East, latitude  $0^\circ$  North and a local time of 12 hours.

where  $R_0 = 6051848$  m is the reference sphere. The trajectory's initial point (lat0,lon0) remains unchanged for a complete set of points launched by the same seed, and is resaved at every new seed launch.

Perturbations to temperature, density and wind at height  $z$  are then found by using the value at  $z + \delta z$  on the background profile, with the perturbations to temperature and density calculated on the assumption of adiabatic motion to the valid height.

A value can be chosen for  $\lambda$  in the range of 2-30 km; longer vertical wavelength should be well resolved by the model and shorter wavelengths result in negligible perturbations. By fixing the value of  $\lambda$ , the user will also fix the one of  $\lambda_{Dist}$ , such as  $\lambda_{Dist} = 100 \times \lambda$ , which is a realistic order of magnitude.

The amplitude of the waves, growing with the altitude, is limited by a maximum value of  $\delta z$ ,  $\delta z_{max}$ , corresponding to their saturation.

An example of several small scale perturbations is shown in Figure 6.

## 6 High Resolution Outputs

### 6.1 High Resolution Topography

The Venus Climate Database has been compiled from the output of a general circulation model in which the topography is very smoothed because of its low resolution. The VCD includes (and uses) topography at the database's resolution ( $96 \times 96$  in longitude $\times$ latitude).

The access software includes a "high resolution" mode with a high resolution (23 pixels/degree) topography measured with Magellan radar altimeter (combined with a few topography measurements from Pioneer Venus to fulfill the holes) in order to compute surface pressure as accurately as possible.

This 23 pixel per degree topography file is available in NetCDF format. This file, named 'relief.nc' can be found in the directory `data` of the VCD and is used by the VCD software in high resolution mode.

### 6.2 Deriving High Resolution Surface Pressure

High resolution surface pressure may be obtained from GCM surface pressure by taking into account the change in altitude from the (coarse and smoothed) GCM topography to more realistic high resolution

topography. In order to correct the atmospheric mass, we also add a correction with a surface pressure measurement. Five available observations of surface pressure were available : the four Pioneer Venus probes (Sunder, Day, Night, North) and the one from the Vega 2 mission. Vega 2 was chosen as the reference since its position is the one known with the less incertitudes, knowing that the terrain around the landing site is quite flat, which minimize the incertitudes of the altitude of the landing site. This is not the case for the four Pioneer Venus probes for which the incertitude of position and the variability of the terrain height around their theoretical landing positions are high.

In practice, an estimation of the high resolution surface pressure  $P_s$  at a given location and time is given by:

$$P_s = P_{s_{\text{GCM}}} \frac{P_{\text{Vega2}_{\text{OBS}}}}{P_{\text{Vega2}_{\text{GCM}}}} e^{-(z-z_{\text{GCM}})/H} \quad (31)$$

where  $P_{s_{\text{GCM}}}$  is the pressure predicted by the GCM at the same location and time (bilinear interpolations from the VCD grid),  $P_{\text{Vega2}_{\text{OBS}}} = 8902500$  Pa the Vega 2 surface pressure measurement,  $P_{\text{Vega2}_{\text{GCM}}}$  the Vega2 surface pressure predicted by the GCM (interpolated vertically and horizontally),  $z$  is the altitude of the local surface retrieved from the Magellan dataset, and  $z_{\text{GCM}}$  is the altitude at the location interpolated from the coarse GCM topography grid.  $H$  is the scale height used in the hydrostatic equation to vertically interpolate the pressure:

$$H = \frac{R}{g} \left( T + \frac{\Gamma}{2} (z - z_{\text{GCM}}) \right) \quad (32)$$

with  $R$  ( $\text{m}^2 \text{s}^{-2} \text{K}^{-1}$ ) the gas constant,  $g = 8.87 \text{ m s}^{-2}$  the acceleration of gravity,  $T$  the atmospheric temperature extracted from the GCM near the surface (3rd atmospheric layer) and  $\Gamma = -g/C_p$  the adiabatic lapse rate,  $C_p$  being the heat capacity.

An example of the impact of the high resolution scheme is given in Figure 7 where is plotted the surface pressure at GCM resolution and at high resolution using the scheme described above. A comparison with the values measured at the four Pioneer Venus probes landing sites and at the Vega 2 site is given.

## 6.3 Computing High Resolution Values of Atmospheric Variables

### 6.3.1 Redefinition of high resolution levels

The VCD contains the atmospheric variables obtained from the corresponding GCM runs, on the same grid (as described in Section 2.4), and thus at given pressure  $P(l)$  for a given location and layer  $l$  (computed from surface pressure  $P_{s_{\text{GCM}}}$  interpolated from the  $1.875^\circ$  latitude by  $3.75^\circ$  longitude  $96 \times 97$  VCD grid, using Equation 1 as shown in Section 2.4.2). If, however, the surface pressure should in fact be  $P_{s_{\text{HR}}}$ , as given by Equation 31, then a legitimate question is : How should the atmospheric variable profile above  $P_{s_{\text{HR}}}$  be interpolated?

First of all, we redefined the “high resolution” pressure levels  $P_{\text{HR}}(l)$  such that  $\forall z(l) < z_{\text{lim}}$ :

$$P_{\text{HR}}(l) = P_{\text{GCM}}(l) \left[ \frac{P_{s_{\text{HR}}}}{P_{s_{\text{GCM}}}} + \left( 1 - \frac{P_{s_{\text{HR}}}}{P_{s_{\text{GCM}}}} \right) \frac{z(l)}{z_{\text{lim}}} \right] \quad (33)$$

where  $P_{\text{GCM}}(l)$  is the original pressure level in the database.  $z(l)$  roughly corresponds to the altitude above the surface (in the GCM smoothed topography) of  $l^{\text{th}}$  level and  $z_{\text{lim}} = 10$  km is the limit altitude at which GCM levels corresponds to High resolution levels in term of pressure (where  $P_{\text{HR}}$  will become equal to  $P_{\text{GCM}}$ ). Above this limit altitude we set :  $P_{\text{HR}} = P_{\text{GCM}}$ .

At low altitude, it is designed to ensure that the first levels will be at the same distance from the surface in both low and high resolution (then  $P_{\text{HR}}/P_{s_{\text{HR}}} = P_{\text{GCM}}/P_{s_{\text{GCM}}}$ ). In such conditions the model levels are compressed closer to the surface, to mimic the observed behaviour in high resolution GCM simulations. Conversely the model levels are expanded when there is a topographic low.

Figure 8 displays an example of the pressure field produced by the “high resolution” compared to the “low resolution” version.



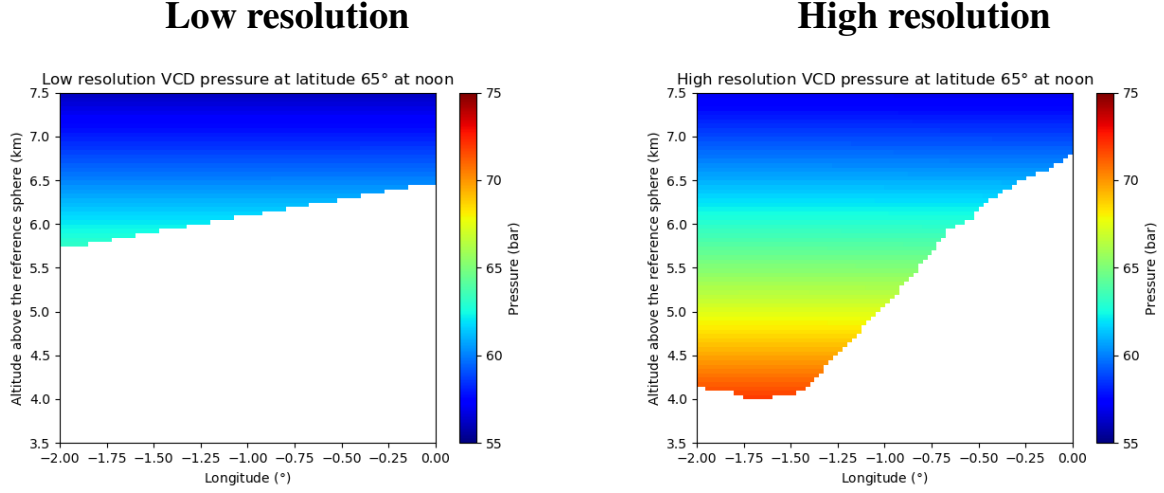


Figure 8: Cross-section of pressure fields near Maxwell Montes from the VCD at low resolution and high resolution.

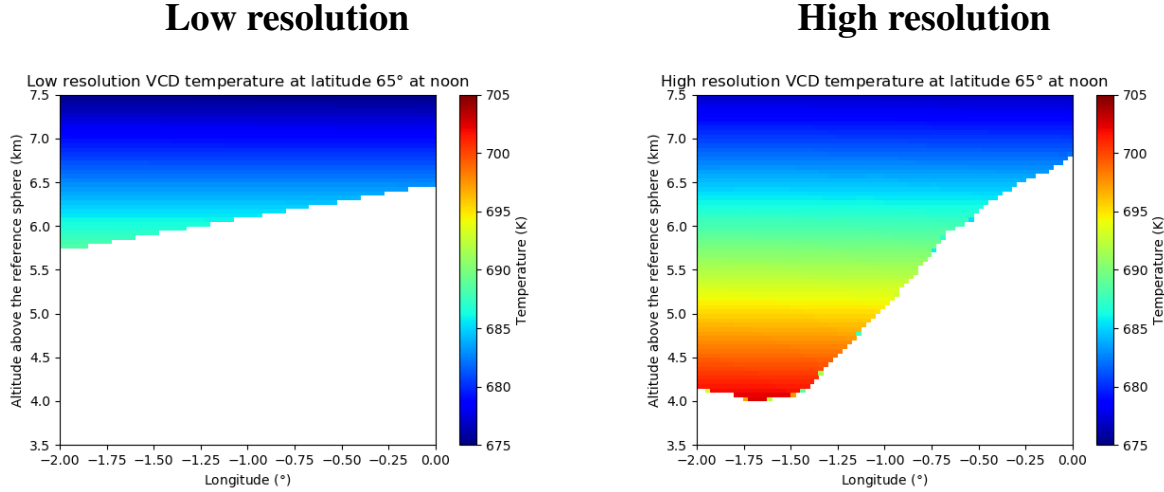


Figure 9: Cross-section of temperature fields near Maxwell Montes from the VCD at low resolution and high resolution.

### 6.3.2 Interpolation of atmospheric temperature

We consider that at a given altitude above the reference sphere, the temperature should remain constant, the topography having no impact on the temperature field on Venus. Therefore we reconstruct a high temporal resolution temperature profile  $T_{HR}(l)$  from the low resolution one  $T_{GCM}(l)$  and from the pressure levels at low ( $P_{GCM}(l)$ ) and high ( $P_{HR}(l)$ ) resolution. A simple linear interpolation in  $\log(P)$  is performed. If the high resolution surface is under the low resolution one, such that  $P_{HR}(l) > P_{GCM}(l)$ , we extrapolate the temperature profile using the vertical temperature gradient calculated from the first two levels.

Figure 9 displays an example of the temperature field produced by the “high resolution” compared to the “low resolution” version.

### 6.3.3 Interpolation of atmospheric density

In the low resolution version of the database, the density  $\rho$  ( $\text{kg m}^{-3}$ ) at altitude  $z$  (m) is interpolated in the vertical by using a weighted interpolation between the density at the level above ( $\rho_{l+1}$ ) and the density at

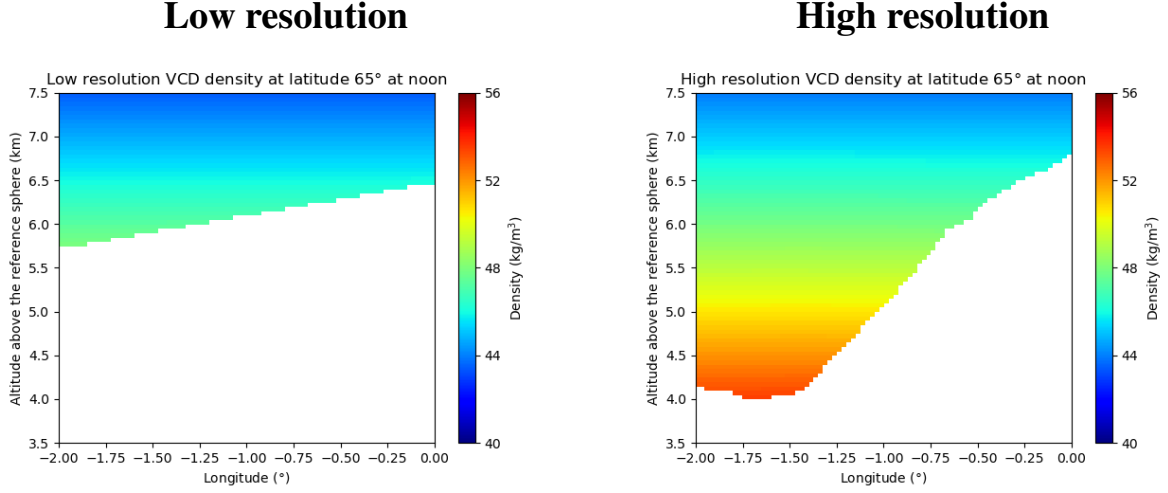


Figure 10: Cross-section of density fields near Maxwell Montes from the VCD at low resolution and high resolution.

the level below ( $\rho_l$ ). In practice the linear interpolation is performed on the logarithm of the density (as mentioned in Section 4.2.2).

Density is proportional to pressure and inversely proportional to temperature and thus, in the “high resolution” mode, it must be recomputed by taking into account the change in pressure (computed using Equation 33) and temperature. The algorithm for vertical interpolation that is used for “low resolution” density is used to compute “high resolution” density but using  $\rho_{HR}(l)$  instead of  $\rho_{GCM}(l)$  such that :

$$\rho_{HR}(l) = \rho_{GCM}(l) \frac{P_{HR}(l)}{P_{GCM}(l)} \frac{T_{GCM}(l)}{T_{HR}(l)} \quad (34)$$

Figure 10 displays an example of the density field produced by the “high resolution” compared to the “low resolution” version.

A similar treatment is applied to the RMS of density, corresponding perturbations, etc...

### 6.3.4 Interpolation of other variables

Most other atmospheric variables  $X$  (winds, gases, mixing ratio...) are treated the following way. The values of the profile  $X(l)$  are kept, and only the pressure of the levels  $P(l)$  are recomputed to account for the “high resolution” post-processing. This allows to take into account the effect of the topography near the surface, ie. under 10km and keep the exact same profile as the low resolution above this limit altitude.

Figure 11 displays an example of the zonal wind field produced by the “high resolution” compared to the “low resolution” version.

### 6.3.5 Modification of column densities

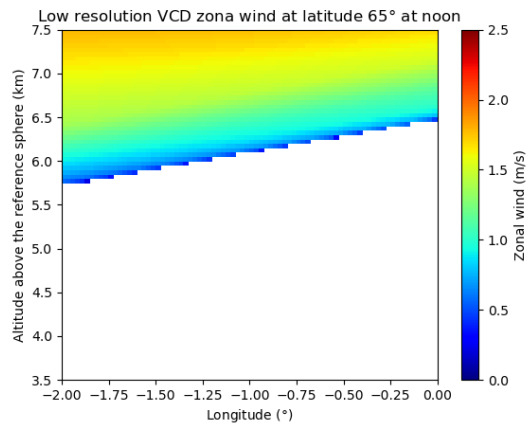
Column densities properly integrated in low resolution are already stored in the VCD data files. In order to get columns for the high resolution surface pressure, the two columns are roughly integrated over high and low resolution sigma levels based on the hydrostatic equation:

$$\begin{aligned} c_{GCM}^j &= \sum_{l=2}^{l=L} \{ (vmr^j(l) R_m(l) [P_{GCM}(l-1) - P_{GCM}(l)] / (g_m(l) k_b) \} \\ c_{HR}^j &= \sum_{l=2}^{l=L} \{ (vmr^j(l) R_m(l) [P_{HR}(l-1) - P_{HR}(l)] / (g_m(l) k_b) \} \end{aligned} \quad (35)$$

where  $vmr^j$  is the volume mixing ratio of specie  $j$ ,  $k_b$  is the Boltzmann constant. Then to scale to the high resolution surface pressure the low resolution column density from the VCD data files is multiplied by the ratio  $c_{HR}/c_{GCM}$ .



### Low resolution



### High resolution

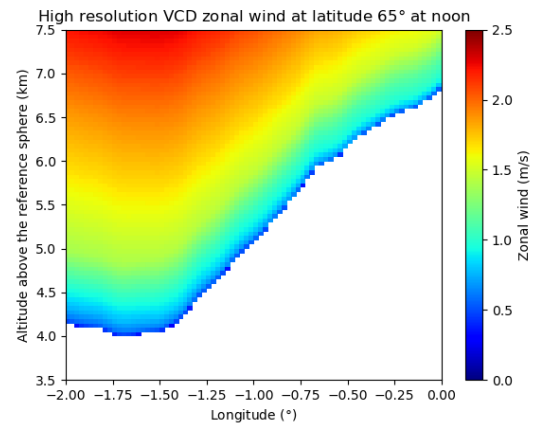


Figure 11: Cross-section of zonal wind fields near Maxwell Montes from the VCD at low resolution and high resolution.

## A Computing Solar Longitude and Local Time

In this appendix we briefly describe how one may compute aerocentric solar longitude of Venus for a given Julian date, as well as how to compute corresponding Local True Solar Time.

### A.1 Some constants

$N_{s1}$	10087200	Number of second in a venusian solar day
$N_{s2}$	19411229	Number of second in a venusian year
$Ls_{peri}$	253.8	Perihelion date (in deg.)
$t_{peri}$	1.36	Perihelion date (in Venus days)
$a$	0.723336	Semi-major axis of orbit (in AU)
$e$	0.00678	Orbital eccentricity
$\epsilon$	2.64	Obliquity of equator to orbit (in deg.)

### A.2 Computing aerocentric solar longitude

#### A.2.1 The three anomalies

For an elliptical trajectory, the **true anomaly**  $\nu$  is given by the polar angle of the position, with the convention that  $\nu = 0$  at the periastron (minimum distance to center, i.e. the perihelion, in Venus' case).

If  $n = 2\pi/T$  is the mean angular frequency of motion ( $T$  is the period of revolution, i.e. a year), then at time  $t$  (taking the time origin as  $t = t_{peri}$ , perihelion date at which  $\nu = 0$ ):

$$n(t - t_{peri}) = 2 \arctan \left[ \sqrt{\frac{1-e}{1+e}} \tan \frac{\nu(t)}{2} \right] - \frac{e\sqrt{1-e^2} \sin \nu(t)}{1 + e \cos \nu(t)} \quad (36)$$

The **eccentric anomaly**  $E$  is related to mean motion  $n$  by:

$$n(t - t_{peri}) = E(t) - e \sin E(t) \quad (37)$$

The **mean anomaly**  $M$  is the angle determining the position of a fictitious point in uniform circular motion of angular frequency  $n$ . Thus,

$$n(t - t_{peri}) = M(t) \quad (38)$$

#### A.2.2 Aerocentric solar longitude $Ls$

The position of the planet on its heliocentric orbit is given by the aerocentric solar longitude  $Ls$ . The origin of  $Ls$  is defined as the vernal equinox (northern hemisphere spring equinox).

$Ls$  is directly related to the true anomaly  $\nu$ :

$$\nu = Ls - Ls_{peri} \quad (39)$$

where  $Ls_{peri}$  is the perihelion date.

#### A.2.3 Converting Julian date $JD$ to $Ls$

First step is to compute a Venus day  $D_s \in [0; N_{s2}/N_{s1}]$  (i.e. number of solar days elapsed since beginning of venusian year defined by  $Ls = 0$ )

$$D_s = (JD - JD_{ref}) \frac{86400}{10087200} \pmod{N_{s2}/N_{s1}} \quad (40)$$

where 86400 and 10087200 are respectively the number of seconds in an earth day and a venusian solar day.  $JD_{ref}$  is a reference Julian date corresponding to an  $Ls = 0$  event.  
Example of reference date:  $Ls = 0$  occurred on 24-05-2020 at 23:00:00, which corresponds to Julian date 2458994.458.

Then we convert this venusian day in  $Ls$  with the following steps :

Step1: Compute mean anomaly  $M(D_s)$  (in radians):

$$M = 2\pi \frac{D_s - t_{peri}}{N_{s2}/N_{s1}} \quad (41)$$

Step 2: Compute eccentric anomaly  $E(D_s)$  (in radians):

This is the tricky bit (also known as Kepler's problem). Eccentric and mean anomalies are related by the nonlinear equation:

$$M = E - e \sin E \quad (42)$$

which must be solved by, for instance, a Newton iterative procedure.

Step 3: Compute true anomaly  $\nu(D_s)$  (in radians) using:

$$\nu = 2 \arctan \left[ \sqrt{\frac{1+e}{1-e}} \tan \left( \frac{E}{2} \right) \right] \quad (43)$$

Step 4: Compute  $Ls(D_s)$  (in degrees):

$$Ls = \left( \nu \frac{180}{\pi} + Ls_{peri} \right) \pmod{360} \quad (44)$$

### A.3 Computing Local True Solar Time

The steps required to compute the Local True Solar Time at a given longitude  $lon$  (in degrees East) for a given Julian Date are the following:

Step 1: Compute Local Mean Time  $LMT_0$  at longitude 0, in venusian hours, for a given Julian date  $JD$  :

$$LMT_0 = \left( LMT_{ref} + 24(JD - JD_{ref}) \frac{86400}{10087200} \right) \pmod{24} \quad (45)$$

Where  $JD_{ref}$  is a reference Julian date, and  $LMT_{ref}$  the Local Mean Time at longitude 0, in venusian hours, at that date. 86400 and 10087200 are respectively the number of seconds in an earth day and a Venus solar day.

Example of reference date and time: 24-05-2020 at 23:00:00 was such that  $JD_{ref} = 2458994.458$  and  $LMT_{ref} = 20.743$ .

The equation above assumes simple Keplerian motion and is accurate to a few minutes with respect to more advanced computations which account for the interactions between main objects of the Solar System.

Step 2: For the given aerocentric solar longitude  $Ls$  (which is obtained from Julian Date  $JD$  as explained previously), compute the Equation Of Time  $EOT$  (which yields, in venusian hours, the difference between Mean Time and True Solar Time) :

$$EOT = \left[ 2e \sin(Ls - Ls_{peri}) - \tan^2 \left( \frac{\epsilon}{2} \right) \sin(2Ls) \right] \frac{24}{2\pi} \quad (46)$$

Where  $Ls_{peri}$  is the perihelion date,  $e$  the orbital eccentricity and  $\epsilon$  the obliquity of equator to orbit.

Step 3: Compute Local True Solar Time at longitude 0,  $LTST_0$ , in venusian hours:

$$LTST_0 = LMT_0 - EOT \quad (47)$$

step 4: At longitude  $lon$  (in degrees east), Local True Solar Time  $LTST$ , in venusian hours, is then:

$$LTST = \left( LTST_0 - \frac{lon}{15} \right) (mod\ 24) \quad (48)$$