The Time Step Data Format (TSD)

with implementation in R

Arne Johannes Holmin

November 1, 2018

Abstract

This is the documentation for the Time Step Data format (TSD) developed by Arne Johannes Holmin for the purpose of dynamic reading and writing of data of varying type and dimension over multiple time steps. The TSD format includes binary and text files, and reading and writing functions exist in R.

Contents

Chapter 1

Definition

The Time Step Data file format (TSD) is a file format developed for saving data of varying length and dimension over multiple time steps. Accepted data types are character strings, byte integer, short integer, long integer, float, double, complex, and R-functions (saved as strings and converted back to function). The specific file structure is listed in table 1.1 below.

Table 1.1: The TSD file format

#	**Name**	**Length**	**Description {Variable type}**
1	*nvar*	1	Number of variables {float}
2	*numt*	1	Number of time steps {long integer}
3	*lvar*	*nvar*	Lengths of the variables (see Section 1.2) {float}
4	*labl*	*nvar*	Labels/names of the variables (see Section 1.3) {4 characters UTF-8, 1 byte pr character}
5	*dtyp*	*nvar*	Data types (see Section 1.1) {4 characters UTF-8, 1 byte pr character}
6	$X_{1,1}$ ⋮ $X_{nvar,1}$	$lvar[1]$ ⋮ $lvar[nvar]$	Variable number 1 at time step number 1 {dtyp[1]} ⋮ Variable number *nvar* at time step number 1 {dtyp[1]}
6	$X_{1,2}$ ⋮ $X_{nvar,2}$	$lvar[1]$ ⋮ $lvar[nvar]$	Variable number 1 at time step number 2 {dtyp[2]} ⋮ Variable number *nvar* at time step number 2 {dtyp[2]}
⋮			

1.1 Data types

The data types used in the binary TSD file format, are given by four character strings as shown in Table 1.2.

Table 1.2: Legal variable types used in echoIBM.

Data type	Name	Size [bytes]
UTF-8	char	1
Byte	byte	1
Short int	shrt	2
Float	floa	4
Double	doub	8

1.2 Variable lengths

The TSD format was developed to support saving data of variable length or dimension for each time step. A variable can be 3-D array at one time step, a one dimensional vector at the next time step, or missing in another time step. The variable lengths given by the *lvar* variable may be given as a vector of *nvar* elements, in the case that the variable lengths does not change between time steps, or as a vector of length *nvar* times the number of time steps to be saved in the file. Space in *lvar* can be reserved for additional time steps with the R-functions. Otherwise a file saved with variable lengths of the variables cannot be appended to, since the information about the length of the variables is locked in the file. It is always possible to read the data, append time steps, and save the data again.

The dimension information is stored in the file by a special variable named *d000*. This is a numeric vector $(D,\ N_1, \ldots, N_D,\ I_1, \ldots, I_D,\ d_{1,i}, \ldots, d_{1,N_i},\ \ldots,\ d_{D,i}, \ldots, d_{D,N_i})$, where D is the number of variables with dimension, $N_1, \ldots, N_D$ are the number of dimensions for each of these variables, $I_1, \ldots, I_D$ are the index numbers of the positions of these variables in the list of variables, and $d_{1,i}, \ldots, d_{1,N_i}, \ldots, d_{D,i}, \ldots, d_{D,N_i}$ are the dimensions of the variables.

1.3 Variable names

Variable names are restricted to four character names in the current version of the TSD format. This has advantages for reading the files, and is also has the side effect of forcing the user to consider carefully the names of the variables. However, future versions may support variable names of arbitrary length. See info.TSD() for list of four character variables used in the packages TSD, sonR, echoIBM and cpplot3d.