

s2dverification

Update meeting – 21/09/2015

Sprint review

Mailing list:

Nico: Unknown yet. We will create new BSC mailing lists.

Wiki:

<https://earth.bsc.es/wiki/doku.php?id=tools:s2dverification>

Outline

- Current state
- Ready to test features
- Issues and comments
- Introducing 'rdoxygen' and 'testthat'
- Useful links

Current state

- Latest release: s2dverification 2.4.3
 - Released internally (2.4.0 on CRAN).
 - Fixed bug that made LOCALPLOT crash.
 - Fixed bug when loading observational global means.
 - NMME models output now loadable.

Ready to test features

develop-trend	François, Chloé
develop-seaicemodesetc	Already tested, last fixes ongoing, depends on develop-trend
develop-plotwind	Verónica, Eleftheria
develop-composite	Chloé, Verónica
develop-crpss	Omar
develop-toymodel	François, Eleftheria
develop-animVSLtime	Chloé, Eleftheria, Omar
develop-int-fusion-Coupled-Modeling-System	Verónica, Omar

Issues and comments

- Change all paths from /cfunas to /esnas
 - To be addressed ASAP. Release tomorrow, see #98.
 - Nico: If you need to use it before, you can edit the configuration file and update the paths. Then specify its file name as 'configfile' parameter in Load().
- Release on CRAN at the end of September?
 - Neven needs to publish some functions. What's the progress in testing these?
 - Neven: I will ask the testers (Verónica, Chloé and François) to test the pending features before 2 weeks.

Issues and comments

- News in develop-THREDDS-subsetting
 - Subsetting enabled. See test results. See MR#62.
 - Other improvements: remap in local. Allow any order of dimensions in NetCDF files, but warn if not in standard order. See #97, #95, #89.
 - Should we go again through a testing period?
 - Nico: As conclusion, subsetting along longitudes doesn't lead to time gain, whereas subsetting along latitudes and leadtimes leads to a gain of 1:1 (e.g: half latitudes, half time).
 - Paco: Is Load() using the entire file when you request a subset of the data?

Issues and comments

- News in develop-THREDDS-subsetting
 - Nico: Load() as is installed in the machines now takes the files entirely. The version in develop-THREDDS-subsetting takes only the required parts and that's the reason of the time gains we can see in the charts. If you need to use it you can checkout the branch, install it locally and use it.
 - Paco: Are we going to use then a THREDDS server? If the user decided to work with the data in the THREDDS server, would them see it as separate files (start dates, chunks, ...) or as a single dataset?

Issues and comments

- News in develop-THREDDS-subsetting
 - PA: We are going to use a THREDDS server because it makes possible to share the data with other institutions. Also, there's a little time gain using THREDDS. The data in the THREDDS server is organized as a collection of separate files now. The user must access them as separate files and Load() accesses them as separated files. There's the option to organize them in a collection and the user would see the dataset as a whole, but the server should be reinstalled. We need to try that yet, there are advantages and disadvantages.

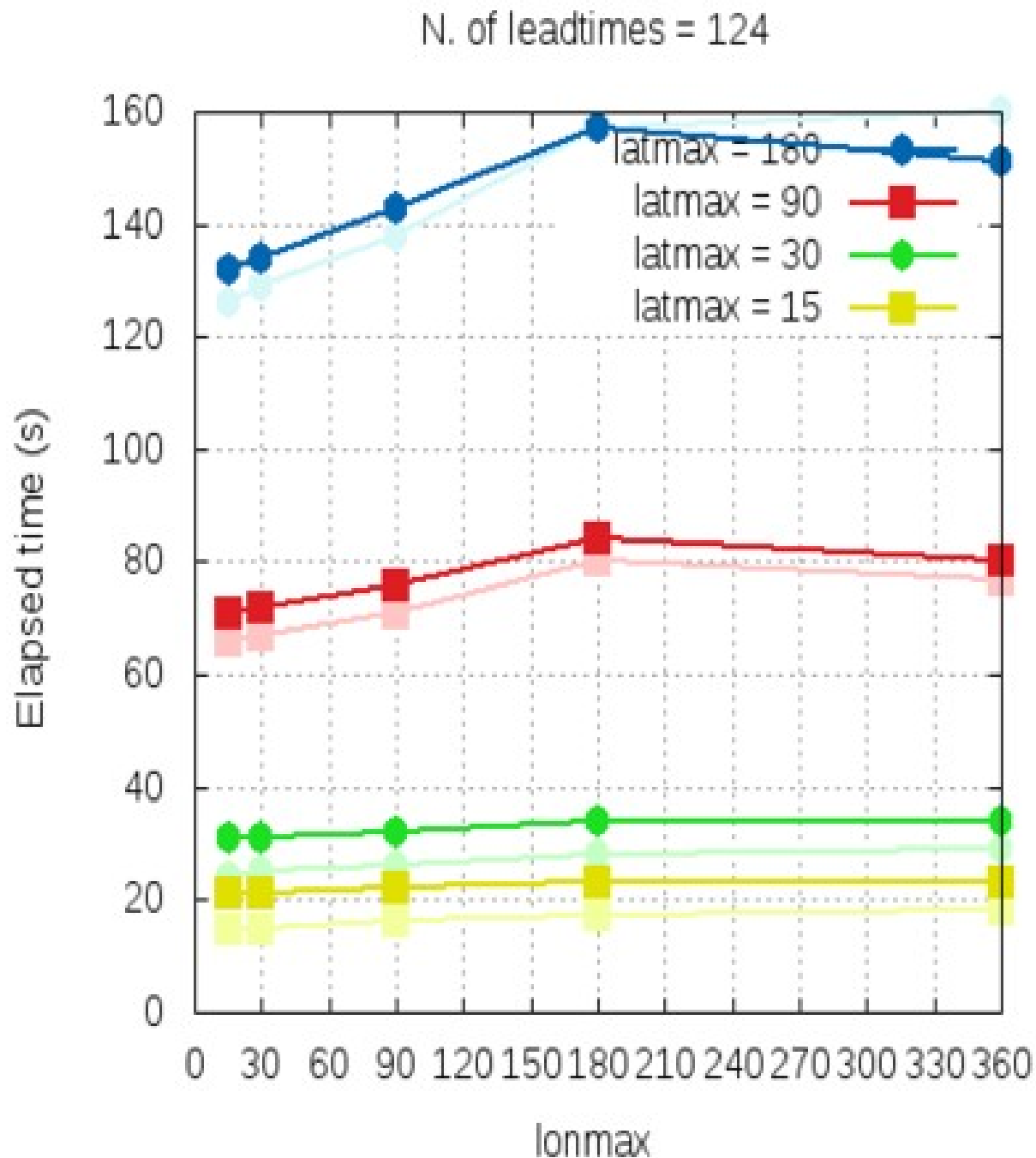
Issues and comments

- News in develop-THREDDS-subsetting
 - Paco: It would be interesting, at least, that the THREDDS could offer different chunks of a run as a single file. There's software that doesn't handle chunks in different files. Additionally, we should try to apply the Load()'s subsetting to try to work with huge datasets to see if makes possible to work with them in R in a reasonable time.

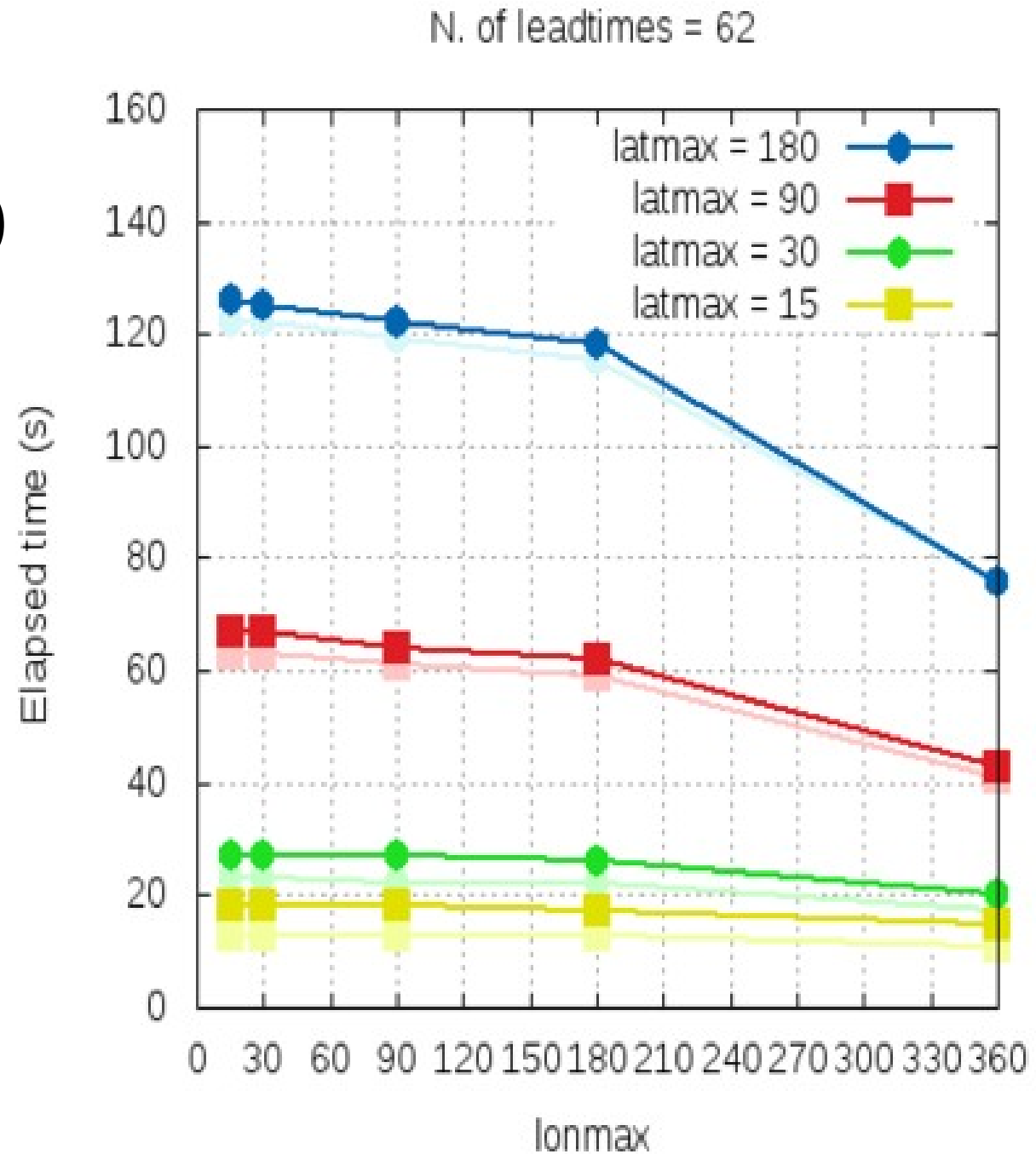
Using Load() + subsetting

Bold lines show tests performed against file server.

Light lines show tests performed against a THREDDS server.

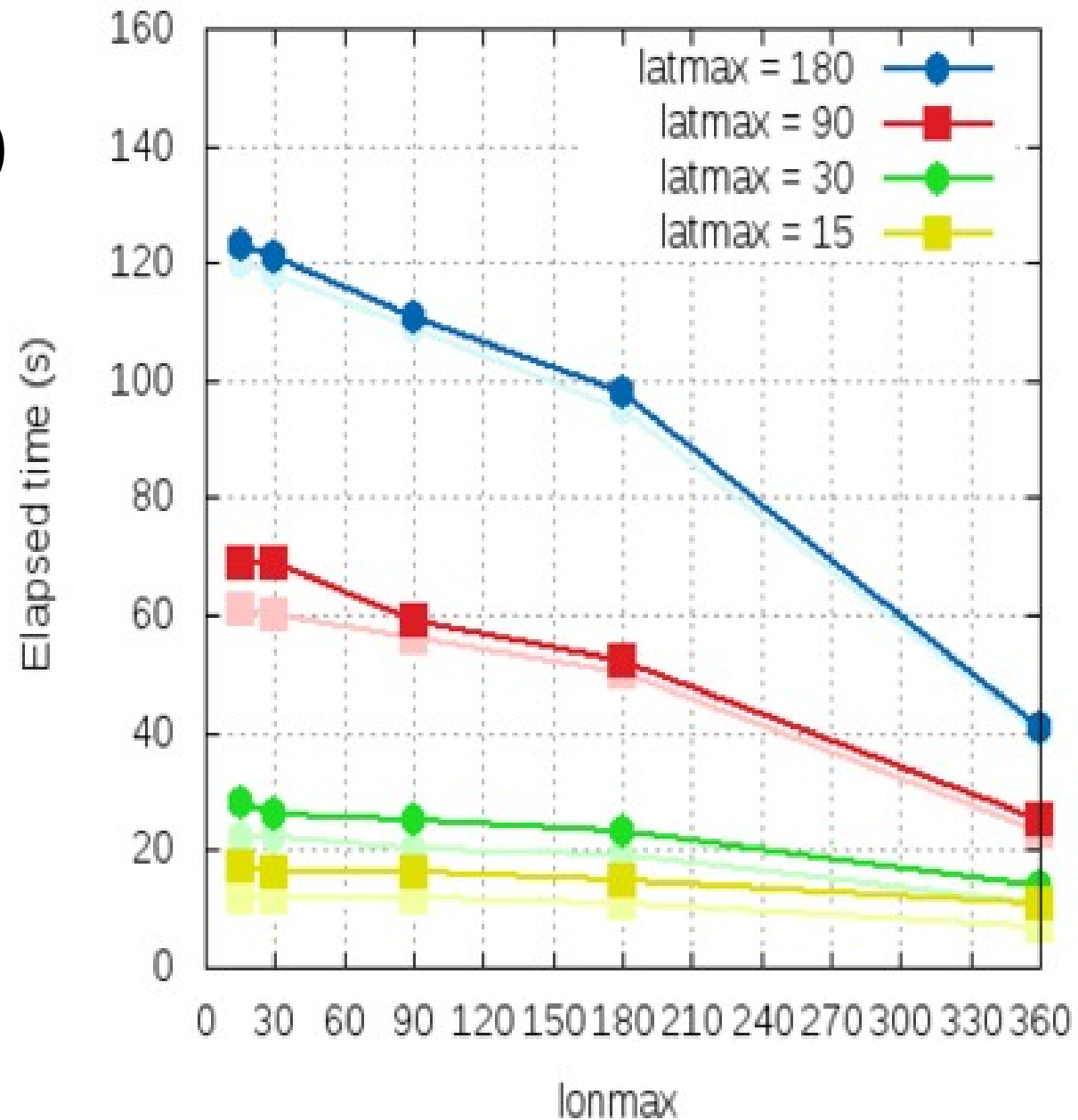


Using Load()
+ subsetting



Using Load()
+ subsetting

N. of leadtimes = 30



Issues and comments

- Should we find a way to interpolate in R?
 - To avoid dependencies with cdo. Paper reviewers!
 - To save time writing data to disk and reading it back.
 - Would give extra value to s2dverification.
 - What kinds of interpolation supported by cdo remap* are the used ones? Can we do without part of them?
 - Nico: There are tools to interpolate in R, but not to interpolate between gaussian grids and regular grids. Anyone know any tool for doing so? We will comment this issue in next meetings.
 - Eleftheria, François, Neven: we should stick to cdo because is a very good tool and specific to do what we do.

Issues and comments

- Moving to 'rdoxygen' and 'testthat'
 - 'rdoxygen': New, automated way of documenting functions.
 - 'testthat': Tools to add unit testing to the functions (not examples!).
 - Extra effort, but recommended before publishing to RJSS.
 - **Danila: Why should we move to 'rdoxygen'? The current documentation system works fine.**
 - **Nico: It works fine, but more and more powerful packages are using it as documenting tool. It keeps together code and documentation avoiding outdated documentation files. It also makes easy some maintenance tasks which otherwise are tedious.**

Documenting with 'rdoxygen2'

- Documentation in the code, in comments ('#')

```
#' Add together two numbers.  
#'  
#' @param x A number.  
#' @param y A number.  
#' @return The sum of \code{x} and \code{y}.  
#' @examples  
#' add(1, 1)  
#' add(10, 1)  
add <- function(x, y) {  
  x + y  
}
```

Testing with 'testthat'

- Test code for each function, in 'tests/testthat'

```
library(stringr)
context("String length")

test_that("str_length is number of characters", {
  expect_equal(str_length("a"), 1)
  expect_equal(str_length("ab"), 2)
  expect_equal(str_length("abc"), 3)
})

test_that("str_length of factor is length of level", {
  expect_equal(str_length(factor("a")), 1)
  expect_equal(str_length(factor("ab")), 2)
  expect_equal(str_length(factor("abc")), 3)
})
```


Issues and comments

- Other issues

- Omar: What about loading observations only for a specified range of time?
- Nico: Verónica and I are working on it. Currently, Load() returns data from the first specified starting date to the current date and is a bit buggy. If anyone wants to contribute is welcome.

Useful links

- How to develop in s2dverification: updated

<https://earth.bsc.es/wiki/doku.php?id=tools:s2dverification#development>

- GitLab

<https://earth.bsc.es/gitlab/s2dverification>

- SCRUM planning

<https://docs.google.com/spreadsheets/d/16-7CvRxNEcpX4k25sYjcV3G4XoqYFwvGXn1W0fv3CAQ/edit?usp=sharing>

Thanks for your attention

nicolau.manubens@bsc.es