R tools users meeting

An-Chi Ho

07/04/2022

contributor: Carlos Delgado

Agenda

1. Ice-breaker

2. News
   - startR
3. User presentation: New s2dv functions [Carlos]
4. Q&A
   - CDORemap from coarse to fine grid, mismatch lat/lon grid problem [Jaume] https://earth.bsc.es/gitlab/es/csdownscale/-/issues/2
   - Shapefiles in esarchive [Núria]

Ice-breaker

The shiny apps we have…

[An-Chi]
ECEARTH experiment diagnostics/cmorfiles visualization https://earth.bsc.es/shiny/cp_shiny-figure/

[Carlos   D.]
C3S_34c   https://earth.bsc.es/shiny/C3S_34c/
EUCP   wind   case   study   https://earth.bsc.es/shiny/EUCP-wind-case-study/
FOCUS   Africa   https://earth.bsc.es/shiny/FOCUS-Africa/
[Bala]
Decadal SPEI6 https://earth.bsc.es/shiny/decadal_spei6/

[Lluís]   *modules*   *used*
AI4D   https://earth.bsc.es/shiny/AI4D/
VITIGEOSS https://earth.bsc.es/shiny/ess-oper/

[Pep]
Medprojections   https://earth.bsc.es/shiny/medprojections-shiny_app/
Projection   Constraints   https://earth.bsc.es/shiny/projection-constraints/
[Bala]
Observed   SPEI   https://earth.bsc.es/shiny/observed_spei/
[Nicola]   (?)
Scorecards   https://earth.bsc.es/shiny/scorecards-shiny_app/

startR

Development & Bugfix

- Reshape the metadata

<mark>Status</mark>: in master branch, has memory problem

Start() provides the reshaping parameters "merge_across_dims" and "split_multiselected_dims". Metadata now can be reshaped accordingly.

Memory problem:
(Part of) the metadata reshaping process seems to take up much memory and crashes the console even if `retrieve = FALSE`.

Development & Bugfix

- Dependency between inner dim and multiple file dim (e.g., region varies with both sdate and member)

  Status: in branch develop-multi_dim_dependency

  The metadata do not have correct dimensions now. But data are fine.

Use case:

https://earth.bsc.es/gitlab/es/startR/-/blob/develop-multi_dim_dependency/inst/doc/usecase/ex1_13_implicit_dependency.R#L89

FAQ about defining dependency by array:

https://earth.bsc.es/gitlab/es/startR/-/blob/develop-multi_dim_dependency/inst/doc/faq.md#22-define-the-selector-when-the-indices-in-the-files-are-not-aligned

New s2dv functions

DiffCorr()

Correlation difference to compare two **deterministic** forecasts

Statistical significance: Siegert et al. (2017)
- It does not assume that the forecasts are independent
- Suitable for forecasts that are strongly correlated
- More power (more probability of correctly detecting skill improvements)
- Time series autocorrelation taken into account

Parameters:
- *exp*
- *ref*
- *obs*
- *N.eff* = NA
- *time_dim* = "sdate"

- *member_dim* = NULL
- *method* = "pearson"
- *alpha* = NULL
- *handle.na* = "return.na"
- *ncores* = NULL

https://earth.bsc.es/gitlab/es/s2dv/-/blob/develop-DiffCorr/R/DiffCorr.R

ResidualCorr()

Residual correlation assesses whether a **<u>deterministic</u>** forecast (*exp*) captures any of the observed variability (*obs*) that is not already captured by a deterministic reference forecast (*ref*).
- Suitable for comparison of forecasts that are strongly correlated
- Described in Smith et al. (2019)
- Ranges from -1 to 1

Procedure:
- Compute the residuals of *exp* and *obs* by linearly regressing out *ref*
- Residual correlation computed as the correlation between the residuals of *exp* and *obs*
- Statistical significance computed with a two-sided t-test (autocorrelation taken into account)

Parameters:
- *exp*
- *ref*
- *obs*
- *time_dim* = "sdate"
- *member_dim* = NULL
- *method* = "pearson"
- *alpha* = 0.05
- *ncores* = NULL

https://earth.bsc.es/gitlab/es/s2dv/-/blob/develop-ResidualCorr/R/ResidualCorr.R

RPS()

Ranked Probability Score (RPS)
- Described in Wilks (2011)
- Similar to the RMSE formula but with probabilities
- Useful to estimate the quality of a **multi-categorical probabilistic** forecasts
- If nºcategories = 2 → Brier Score (BS)
- If nºcategories = ∞ → Continuous Ranked Probability Score (CRPS), but not implemented
- Ranges from 0 (perfect forecast) to nºcategories-1 (worst forecast)
- FairRPS → potential RPS with infinite members
  - suitable to estimate the potential skill
  - not suitable for climate services

$$RPS = \sum_{m=1}^{J} \left[\left(\sum_{j=1}^{m} y_j \right) - \left(\sum_{j=1}^{m} o_j \right) \right]^2$$

$$\overline{RPS} = \frac{1}{n} \sum_{k=1}^{n} RPS_k$$

Parameters:
- *exp*
- *obs*
- *prob_thresholds* = c(⅓,⅔)
- *indices_for_clim* = NULL

- *Fair* = FALSE
- *time_dim* = "year"
- *member_dim* = "member"
- *ncores* = NULL

https://earth.bsc.es/gitlab/es/s2dv/-/blob/develop-RPSS/R/RPS.R

RPSS()

Ranked Probability Skill Score (RPSS)

RPSS = 1 - RPS_exp / RPS_ref

- Described in Wilks (2011)
- Useful to compare two **multi-categorical probabilistic** forecasts
- If nºcategories = 2 → Brier Score (BSS)
- If nºcategories = ∞ → Continuous Ranked Probability Score (CRPSS), but not implemented
- Ranges from -∞ to 1
    - positive value → forecast is more skillful
    - negative value → reference forecast is more skillful
- FairRPSS → potential RPSS with infinite members
    - suitable to compare two models with different ensemble size
    - not suitable to compare two actual forecasts
- Statistical significance: Random Walk test (DelSole and Tippett, 2016)

Parameters:

- *exp*
- *obs*
- *ref* = NULL

- *prob_thresholds* = c(⅓,⅔)
- *indices_for_clim* = NULL
- *Fair* = FALSE

- time_dim = "year"
- *member_dim* = "member"
- *ncores* = NULL

https://earth.bsc.es/gitlab/es/s2dv/-/blob/develop-RPSS/R/RPSS.R

Q & A

CDORemap: from coarse to fine grid

Gitlab issue: https://earth.bsc.es/gitlab/es/csdownscale/-/issues/2

Q & A

Next meeting: 5th May 2022 (11 am)

Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación