M Gmail

Nathan Steiger <nathanjohnsteiger@gmail.com>

LMR proxy database
8 messages

Nathan Steiger <nsteiger@ldeo.columbia.edu>   Mon, May 22, 2017 at 5:07 PM
To: "Horlick, Kaleb" <horlickk@oregonstate.edu>, Robert Tardif <rtardif@u.washington.edu>
Cc: Julien Emile-Geay <julieneg@usc.edu>

Kaleb and Robert (cc Julien),

So I've been in discussion with Nick McKay and Chris Heiser about getting the LMR data that's on GitHub
into a format I can use in Matlab. The trouble seems to be that in the conversion from NOAA files to LiPD,
the proxy data isn't consistently and correctly transported over. Last I spoke with them, Nick and Chris
aren't actively working on improving the NOAA to LiPD conversion and so from the perspective of people
interested in using the LMR data for reconstructions, unfortunately the LMR database is not in a usable
state. Which is a shame because it's the largest such database out there and it seems to give much better
hydroclimate reconstructions than the PAGES database.

So I'm wondering if it would be possible to put the LMR proxies into LiPD on your end of things? Then it
would be much more straightforward for people to use the Python and Matlab utilities built by Nick and
Chris to ingest the data for reconstructions.

Or perhaps I could help to convert the python data that Robert extracted, into Matlab files or a file type that
Matlab could read? Because it's pretty hopeless for Matlab to read the NOAA files directly. Perhaps then
we could even contribute that version of the data on the LMR GitHub page?

Nathan

Robert Tardif <rtardif@u.washington.edu>   Tue, May 23, 2017 at 12:07 PM
To: Nathan Steiger <nsteiger@ldeo.columbia.edu>
Cc: "Horlick, Kaleb" <horlickk@oregonstate.edu>, Julien Emile-Geay <julieneg@usc.edu>, Greg Hakim
<ghakim@uw.edu>

Nathan,

Here is an update on the LMR proxy database.

- Data converted to LiPD: Won't happen soon as I do not have time available for that.
- I would suggest that you do not use the "NCDC" files "as provided". I found many problems with data in
there. Some data are duplicates, some are slightly modified versions of the same record, some are
mislabelled etc., particularly in ice core records, but not exclusively.

Now on a more positive note:

I now have a cleaned up version of the database, which is a merge of the PAGES2k phase 2 data (paper in
press by Julien et al) and extra "NOAA" records. More importantly for you, these "extra" records include the
Breitenmoser tree ring collection. Below is a summary of the counts of proxies per "types" considered in
LMR.

 FINAL SUMMARY:

```
    Total number of merged proxy chronologies   : 3198
    Total number of eliminated chronologies     : 152
    Number of proxy chronologies included in df : 3046
    ------------------------------------------------------
     Bivalve_d18O                    : 1
     Corals and Sclerosponges_Rates        : 12
     Corals and Sclerosponges_SrCa         : 54
     Corals and Sclerosponges_d18O         : 132
     Ice Cores_Accumulation           : 3
     Ice Cores_MeltFeature            : 3
     Ice Cores_d18O                   : 133
     Ice Cores_dD                     : 19
     Lake Cores_Misc                  : 3
     Lake Cores_Varve                 : 8
     Marine Cores_d18O                : 1
     Speleothems_d18O                 : 28
     Tree Rings_Isotopes              : 1
     Tree Rings_WidthPages2             : 347
     Tree Rings_WidthBreit            : 2242
     Tree Rings_WoodDensity             : 59
    ------------------------------------------------------
     Total:                    : 3046
    ----------------------------------------------------------------------
```

I have this in a python-specific format, but I have some ideas of how to get this to you in matlab-friendly format. Stay tuned, hopefully later today.

If there aren't any objections about sharing this data from Julien & Greg (CC'd here), I will work something out.

From a tree ring forward modelling perspective:

The Breitenmoser collection seems clean, meaning that they are all reported as trsgi in a consistent manner (normalized around 1) (thanks to Kaleb). Be aware that tree ring records in the PAGES are consistently reported for the most part, but some (I do not have the list)
are said to be trsgi but use different standardization. I have come across records with values around 0, some around 1000. These represent curve balls w.r.t. VSLite I imagine. Just a heads up ...

If everyone is on board with this, I will move forward with transfer of this data collection.

Will keep you updated.

-R

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Robert Tardif
Department of Atmospheric Sciences
University of Washington
(206) 543-9411
www.atmos.washington.edu/~rtardif
[Quoted text hidden]

Greg Hakim <ghakim@uw.edu>   Tue, May 23, 2017 at 1:13 PM
To: Robert Tardif <rtardif@u.washington.edu>
Cc: Nathan Steiger <nsteiger@ldeo.columbia.edu>, "Horlick, Kaleb" <horlickk@oregonstate.edu>, Julien
Emile-Geay <julieneg@usc.edu>

Robert,

I am fine sharing the data with Nathan, but I suggest that you just
give him the pickle file. He can read that in python and write it to
.mat using a built-in scipy function.

Greg
[Quoted text hidden]

Nathan Steiger <nsteiger@ldeo.columbia.edu>   Tue, May 23, 2017 at 3:31 PM
To: Greg Hakim <ghakim@uw.edu>
Cc: Robert Tardif <rtardif@u.washington.edu>, "Horlick, Kaleb" <horlickk@oregonstate.edu>, Julien
Emile-Geay <julieneg@usc.edu>

Great! Thanks Greg!
[Quoted text hidden]
--
Nathan Steiger
NOAA Climate and Global Change Postdoctoral Fellow
Lamont-Doherty Earth Observatory | Columbia University
61 Route 9W | P.O. Box 1000 | Palisades, NY 10964
http://www.ldeo.columbia.edu/~nsteiger/

Nathan Steiger <nsteiger@ldeo.columbia.edu>   Tue, May 23, 2017 at 3:33 PM
To: Robert Tardif <rtardif@u.washington.edu>

Robert,

How difficult would it be for you to write the pickle file to matlab? I still don't really know python, so I'm not
exactly sure how to proceed if I had just the pickle file.

Nathan
[Quoted text hidden]

Robert Tardif <rtardif@u.washington.edu>   Tue, May 23, 2017 at 4:04 PM
To: Nathan Steiger <nsteiger@ldeo.columbia.edu>

Nathan,

I am not sure about how difficult yet. I am thinking of some a couple of possibilities that I need to test. Will
let you know. I have a couple more things I need to finish but hope to get to this soon.

-R
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Robert Tardif
Department of Atmospheric Sciences
University of Washington
(206) 543-9411
www.atmos.washington.edu/~rtardif

[Quoted text hidden]
  [Quoted text hidden]
  Nathan SteigerNOAA Climate and Global Change Postdoctoral Fellow
  Lamont-Doherty Earth Observatory | Columbia University
  61 Route 9W | P.O. Box 1000 | Palisades, NY 10964

http://www.ldeo.columbia.edu/~nsteiger/

--
Nathan SteigerNOAA Climate and Global Change Postdoctoral Fellow
Lamont-Doherty Earth Observatory | Columbia University
61 Route 9W | P.O. Box 1000 | Palisades, NY 10964
http://www.ldeo.columbia.edu/~nsteiger/

Robert Tardif <rtardif@u.washington.edu>   Fri, May 26, 2017 at 3:55 PM
To: Nathan Steiger <nsteiger@ldeo.columbia.edu>

Nathan,

I apologize but I could not get to the pickle to matlab conversion issue.

I have put our database files available at:

http://www.atmos.washington.edu/~rtardif/xfer/LMRusers/

Files are:

NCDC_v0.2.0_Metadata.df.pckl : Contains the metadata for every proxy record
NCDC_v0.2.0_Proxies.df.pckl : Contains the actual timeseries data for every proxy record

Here is what I do to read those files in python:

python
import pandas as pd

meta = pd.read_pickle('NCDC_v0.2.0_Metadata.df.pckl')
data = pd.read_pickle('NCDC_v0.2.0_Proxies.df.pckl')

meta and data are pandas DataFrames
You can print them to screen to get sense of the structure.

Once again, I apologize for not being more helpful. And I will be out of the office next week, with much to
do on my other projects before I leave tonight. Let me know if you figure out how to convert to a format that
is more friendly for you.

-R

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Robert Tardif
Department of Atmospheric Sciences
University of Washington
(206) 543-9411
www.atmos.washington.edu/~rtardif

[Quoted text hidden]
  [Quoted text hidden]
  Nathan SteigerNOAA Climate and Global Change Postdoctoral Fellow

Lamont-Doherty Earth Observatory | Columbia University
61 Route 9W | P.O. Box 1000 | Palisades, NY 10964
http://www.ldeo.columbia.edu/~nsteiger/

--
Nathan SteigerNOAA Climate and Global Change Postdoctoral Fellow
Lamont-Doherty Earth Observatory | Columbia University
61 Route 9W | P.O. Box 1000 | Palisades, NY 10964
http://www.ldeo.columbia.edu/~nsteiger/

Nathan Steiger <nsteiger@ldeo.columbia.edu>　　　　　　　　Fri, May 26, 2017 at 6:06 PM
To: Robert Tardif <rtardif@u.washington.edu>

No worries! Thanks for looking into it and getting me the data.
[Quoted text hidden]