

Charles has given me permission to examine some of the programming code and data that has been used by Climate Research Unit at East Anglia University for their climate modeling. My information comes from material that was “hacked” from the CRU servers in Britain and released to the public. This material included program code, databases, private emails and documents. Charles indicated that I could use anything I deemed necessary to make my case.

<blockquote>

re: #723 Charles

OK, then go ahead and post whatever you think makes your case. I seriously doubt that you're going to be able to show an "inaccuracy," which is what you claimed. And quibbling over quick and dirty methods used to parse a flat file is not the same thing as demonstrating that the code produced inaccurate results.

</blockquote>

The legacy program suite and databases that are discussed below have nothing to do with “quick and dirty methods used to parse a flat file.” It has to do with the legacy data modeling software that CRU had used since at least the 1990’s and the datadases that were produced by these legacy programs. As part of Charles’ permission, he stated...

<blockquote>

re: #731 Charles

And for your part, you have to prove that a section of code actually produced inaccurate or mistaken results. Not simply that it could have done so. Because that's what you claimed.

</blockquote>

As Charles said “I seriously doubt that you're going to be able to show an "inaccuracy,"” Charles may be right. I wasn’t at CRU, I didn’t see the actual data from the time it was generated till the time it appeared in a report or paper (or disappeared), no more than Charles actually saw William Shakespeare pen any of his plays. In the same way Charles could use the historical records to prove Shakespeare did create all of those wonderful plays, I am going to use the historical record of the programmer who was responsible for maintaining, modifying and correcting the legacy programs at CRU who can show us the inaccuracies he found. Meet Mr. Ian “Harry” Harris.

(end comment 1)

(comment 2)

Who is Ian “Harry” Harris? He is a staff member at the Climatic Research Unit at East Anglia University. His short bio on the CRU staff page says this... “Dendroclimatology, climate scenario development, data manipulation and visualisation, programming.” (<http://www.cru.uea.ac.uk/cru/people/>). He was tasked with maintaining, modifying and rewriting programs from the existing climate modeling software suite that existed at CRU since at least the 1990’s. He kept copious notes of his progress from 2006 through 2009, including his notes and comments internally in the programs themselves and in a 314 page document named “harry_read_me.txt.” If you revel in the minutia of programmer’s notes you can easily find this document on the internet.

I will document 4 different aspects of Ian “Harry” Harris’ notes

- 1) General comments, inaccurate data bases
- 2) CRU Time Series 3.0 dataset
- 3) a RUN dialog
- 4) Faulty code

Quotes are verbatim, including typos, misspellings and language differences)... any other mistakes are mine.

(end comment 2)
(comment 3)

- 1) General comments from the “harry_read_me.txt.” about the CRU programs and data.

Here is Ian “Harry” Harris talking about both the legacy programs and legacy climate databases and the new data he is trying to create.

<blockquote>

“Oh GOD if I could start this project again and actually argue the case for junking the inherited program suite!!”

WLN note: This is the program suite that has been generating data for years for CRU and staff.

...knowing how long it takes to debug this suite - the experiment endeth here. The option (like all the anomdtb options) is totally undocumented so we’ll never know what we lost.”

WLN note: Remember, Dr. Phil Jones, head of CRU initially said they never lost any data.

“Sounds familiar, if worrying. am I the first person to attempt to get the CRU databases in working order?!! The program pulls no punches. I had already found that tmx.0702091313.dtb had seven more stations than tmn.0702091313.dtb, but that hadn’t prepared me for the grisly truth:”

“Getting seriously fed up with the state of the Australian data. so many new stations have been introduced, so many false references.. so many changes that aren't documented. Every time a cloud forms I'm presented with a bewildering selection of similar-sounding sites, some with references, some with WMO codes, and some with both. And if I look up the station metadata with one of the local references, chances are the WMO code will be wrong (another station will have it) and the lat/lon will be wrong too.”

WLN note: How were they generating temperature data on their world grid in the past if they couldn't even match up stations?

“I am very sorry to report that the rest of the databases seem to be in nearly as poor a state as Australia was. There are hundreds if not thousands of pairs of dummy stations, one with no WMO and one with, usually overlapping and with the same station name and very similar coordinates. I know it could be old and new stations, but why such large overlaps if that's the case? Aarrggghhh!”

“So.. should I really go to town (again) and allow the Master database to be 'fixed' by this program? Quite honestly I don't have time - but it just shows the state our data holdings have drifted into. Who added those two series together? When? Why? Untraceable, except anecdotally. It's the same story for many other Russian stations, unfortunately - meaning that (probably) there was a full Russian update that did no data integrity checking at all. I just hope it's restricted to Russia!!”

WLN note: Fixed? What does that mean? And why the quotes? This is live data Ian is talking about.

“This still meant an awful lot of encounters with naughty Master stations, when really I suspect nobody else gives a hoot about. So with a somewhat cynical shrug, I added the nuclear option - to match every WMO possible, and turn the rest into new stations (er, CLIMAT excepted). In other words, what CRU usually do. It will allow bad databases to pass unnoticed, and good databases to become bad, but I really don't think people care enough to fix 'em, and it's the main reason the project is nearly a year late.”

WLN note: This is about the strongest statement Ian makes about the state of the data at CRU

“The big question must be, why does it have so little representation in the low numbers? Especially given that I'm rounding erroneous negatives up to 1!! Oh, sod it. It'll do. I don't think I can justify spending any longer on a dataset, the previous version of which was completely wrong (misnamed) and nobody noticed for five years.”

“This was used to inform the Fortran conversion programs by indicating the latitude-potential_sun and sun-to-cloud relationships. It also assisted greatly in understanding what was wrong – Tim was in fact calculating Cloud Percent, despite calling it Sun Percent!! Just awful.”

WLN note: Dr. Tim Mitchell or Dr. Tim Osborn? CRU - <http://www.cru.uea.ac.uk/~timm/index.html>

“They aren't percentage anomalies! They are percentage anomalies /10. This could explain why the real data areas had variability 10x too low. BUT it shouldn't be - they should be regular percentage anomalies! This whole process is too convoluted and created myriad problems of this kind. I really think we should change it.”

“Am I the first person to attempt to get the CRU databases in working order?!?”

“Right, time to stop pussyfooting around the niceties of Tim's labyrinthine software suites - let's have a go at producing CRU TS 3.0! since failing to do that will be the definitive failure of the entire project..”

 “OH FUCK THIS. It's Sunday evening, I've worked all weekend, and just when I thought it was done I'm hitting yet another problem that's based on the hopeless state of our databases. There is no uniform data integrity, it's just a catalogue of issues that continues to grow as they're found.”

</blockquote>

Remember, he is talking about legacy programs and legacy data.

(end comment 3)

(comment 4)

2) About the CRU Time Series 3.0 dataset.

Remember all the comments I posted here about HARCRT3 dataset, which contains global temperature readings from 1850 onward and the possible problems with the data in that database. Well, HADCRUT3 is built from CRUTEM3 and the Hadley SST data. CRUTEM3 is built partially from CRU TS 3.0 which is mentioned above. And much of the data used for climate modeling in the past was contained in earlier versions of this data table CRU TS 2.1, CRU TS 2.0, CRU TS 1.1 and CRU TS 1.0... table used for earlier climate models. (see history of CRU TS at <http://csi.cgiar.org/cru/>).

Evidently Ian “Harry” Harris managed to finally produce the dataset CRU TS 3.0 and here is a question from Dr Daniel Kingston, addressed to “Tim.”

<blockquote>

So, you release a dataset that people have been clamouring for, and the buggers only start using it! And finding problems. For instance:

</blockquote>

Hi Tim (good start! -ed)

I realise you are likely to be very busy at the moment, but we have come across something in the CRU TS 3.0 data set which I hope you can help out with.

We have been looking at the monthly precipitation totals over southern Africa (Angola, to be precise), and have found some rather large differences between precipitation as specified in the TS 2.1 data set, and the new TS 3.0 version. Specifically, April 1967 for the cell 12.75 south, 16.25 east, the monthly total in the TS 2.1 data set is 251mm, whereas in TS 3.0 it is 476mm.

The anomaly does not only appear in this cell, but also in a number of neighbouring cells. This is quite a large difference, and the new TS 3.0 value doesn't entirely tie in with what we might have expected from the station-based precip data we have for this area.

Would it be possible for you could have a quick look into this issue?

Many thanks,

Dr Daniel Kingston
 Post Doctoral Research Associate
 Department of Geography
 University College London

</blockquote>

</blockquote>

And here is Ian "Harry" Harris' answer.

<blockquote>

Well, it's a good question! And it took over two weeks to answer. I wrote angola.m, which pretty much established that three local stations had been augmented for 3.0, and that April 1967 was anomalously wet. Lots of non-reporting stations (ie too few years to form normals) also had high values. As part of this, I also wrote angola3.m, which added two rather interesting plots: the climatology, and the output from the Fortran gridder I'd just completed. This raised a couple of points of interest:

1. The 2.10 output doesn't look like the climatology, despite there being no stations in the area. It ought to have simply relaxed to the clim, instead it's wetter.
2. The gridder output is lower than 3.0, and much lower than the stations!

I asked Tim and Phil about 1., they couldn't give a definitive opinion. As for 2., their guesses were correct, I needed to mod the distance weighting. As usual, see gridder.sandpit for the full info.

So to CLOUD. For over a year, rumours have been circulating that money had been found to pay somebody for a month to recreate Mark New's coefficients. But it never quite gelled. Now, at last, someone's producing them! Unfortunately.. it's me.

The idea is to derive the coefficients (for the regressing of cloud against DTR) using the published 2.10 data. We'll use 5-degree blocks and years 1951-2002, then produce coefficients for each 5-degree latitude band and month. Finally, we'll interpolate to get half-degree coefficients. Apparently.

Lots of 'issues'. We need to exclude 'background' stations - those that were relaxed to the climatology. This is hard to detect because the climatology consists of valid values, so testing for equivalence isn't enough. It might have to be the station files *shudder*.

Using station files was OK, actually. A bigger problem was the inclusion of strings of consecutive, identical values (for cloud and/or dtr). Not sure what the source is, as they are not == to the climatology (ie the anoms are not 0). Discussed with Phil - decided to try excluding any cell with a string like that of >10 values. Cloud only for now. The result of that was, unfortunately, the loss of several output values,

</blockquote>

(end comment 4)

(comment 5)

3) Run dialogs

Ian "Harry" Harris did a very good job of documenting his different "runs" of the programs, clipping and pasting the "run time dialog" into his "harry_read_me.txt." document. Run time dialog is the text, messages and input prompts that appear on the screen when you run the program. You can see below that the original programmers of the CRU program suite had a "lively" style of informative messages to the end user. Here is a message you get when running an "update" program to merge temperature reporting stations.

<blockquote>

Before we get started, an important question: If you are merging an update - CLIMAT, MCDW, Australian - do you want the quick and dirty approach? This will blindly match on WMO codes alone, ignoring data/metadata checks, and making any unmatched updates into new stations (metadata permitting)?

Enter 'B' for blind merging, or <ret>: B

</blockquote>

Do you know what this program produced? Bad records, an incomplete dataset. Records with station identifiers missing, stations duplicated, no checks for missing data. And if the program had data it didn't know what to do with, it turned the data into a new station, even if it didn't really know what that data was in reference to.

Remember, these are the legacy programs that CRU used to generate data. These were live programs, live data. Ian "Harry" Harris was trying to fix and modify these programs, because many of them produced invalid data.

(end comment 5)
(comment 6)

4) Example of faulty code.

Here is one example, from Ian "Harry" Harris, about an already existing function, one that had been used to generated data in the past.

<blockquote>

Back to precip, it seems the variability is too low. This points to a problem with the percentage anomaly routines. See earlier escapades - will the Curse of Tim never be lifted?

A reminder. I started off using a 'conventional' calculation

$\text{absgrid}(\text{ilon}(i), \text{ilat}(i)) = \text{nint}(\text{normals}(i, \text{imo}) + * \text{anoms}(\text{ilon}(i), \text{ilat}(i)) * \text{normals}(i, \text{imo}) / 100)$ which is: $V = N + AN/100$

This was shown to be delivering unrealistic values, so I went back to anomdtb to see how the anomalies were contracted in the first place, and found this:

$\text{DataA}(\text{XAYear}, \text{XMonth}, \text{XAStn}) = \text{nint}(1000.0 * ((\text{real}(\text{DataA}(\text{XAYear}, \text{XMonth}, \text{XAStn})) / \& \text{real}(\text{NormMean}(\text{XMonth}, \text{XAStn}))) - 1.0))$ which is: $A = 1000((V/N) - 1)$

So, I reverse engineered that to get this: $V = N(A + 1000) / 1000$

And that is apparently also delivering incorrect values. Bwaaaahh!!

</blockquote>

Harry eventually fixed this, so in the future it would produce accurate data, but one wonders how many times data was pushed through this formula in the past and how much invalid data was generated from this faulty function.

(end comment 6)

(comment 7)

Epilog:

Remember Ian “Harry” Harris was working on a legacy program suite, not some “quick and dirty methods.” A suite of programs and datasets used by CRU for climate modeling and in use for many years. If you want to, read his 314 pages of notes that detail better than I could all of the problems he ran into trying to work with those existing legacy programs.

Does this information presented here disprove AGW? Of course not. There are many other scientific organizations besides the CRU. But it does highlight, with provable facts that the CRU in themselves have been responsible for bad data, bad programs and as we have seen by the dust up about the ignored Freedom of Information Act requests that was issued to CRU, responsible for trying to cover up their mistakes. This is bad science and unfair to all the honest scientist the world over who are diligently working on honest climate science.

Addendum:

You have to give Ian “Harry” Harris a lot of credit. Evidently he has been responsible for cleaning up a lot of the mistakes that have existed in climate based datasets in the past. This little narrative represents some of his work with NCEP/NCAR Reanalysis. (National Centers for Environmental Prediction – NOAA - <http://www.ncep.noaa.gov/>)

<http://www.cru.uea.ac.uk/cru/data/ncep/>

<blockquote>

1948-1957 Data Added (Ian Harris, 22 Jul 2008)

2007 Data Added (Ian Harris, 17 Apr 2008)

2006 Data Added (Ian Harris, 11 Mar 2007)

2005 Data Added (Ian Harris, 13 Jan 2006)

2004 Data Added (Ian Harris, 28 Nov 2005)

2003 Data Added (Ian Harris, 11 May 2004)

SURFACE TEMPERATURE ADDED (Ian Harris, 10 December 2003)

WARNING NOTE ADDED FOR SURFACE FLUX TEMPERATURES (Ian Harris, 10 December 2003)

ALL DATASETS UPDATED TO 2002 (Ian Harris, 23 June 2003)

LAND/SEA MASKS ADDED (Ian Harris, 16 December 2002)

Land/Sea Masks for regular and Gaussian grids have been added.

NEW WINDOW ONLINE (Ian Harris, 9 July 2002)

The new Quarter-Spherical Window (0N-90N; 90W-90E) is now in use (see table below).

The old window data (here) has now been entirely replaced.

Please address any requests for new variables to me.

BAD DATA REPLACED (Ian Harris, 23 May 2002)

The TOVS Problem has been resolved and only corrected data appears on this site.

Anyone wishing to access the old (potentially incorrect) data in order to evaluate the extent of the problem should contact me.

</blockquote>

The last entry in that narrative is interesting.

(end comment 7)