

Information on JIFs for Gauge QC

Last Modified: 9 Nov 2007

1. Package to be JIF'd

1.1. Run time

Daily gauge QC runs are to be run at 03:40 UTC, right after hourly and daily gauge data are sent to /com/hourly/prod/gaugeqc [RC 20070404 suggested 5:00 UTC, but since the gauge data are copied over at 03:40, we can do this right after that.]

1.2. Flowchart

```
JIF=dew:/u/wx22y1/gaugeqc

/nwprod/jobs/JGAUGE_QC.sms.prod1
|
| Put daily and hourly gauge data on /com/hourly/prod/gaugeqc:
|_The existing /nwprod/jobs/JGAUGE_QC.sms.prod1
|   |_/nwprod/ush/gaugeqc_load_qual_prcp.pl1
|   |_$_JIF/scripts/exgaugeqc_qc.sh
|
| Cold start:
|   |_$_JIF/ush/gaugeqc_run_qc_cold.pl
|   | go back and loop for 33 days:
|   |   |_$_JIF/ush/gaugeqc_precip_qc.pl
|   |     |_$_JIF/ush/gaugeqc_refnwsli.pl
|   |     |   |_$_JIF/ush/gaugeqc_NWSLIreformat_EMG.pl
|   |     |   |_$_JIF/ush/gaugeqc_getneigh.pl
|   |     |   |_$_JIF/ush/gaugeqc_qc_precip_pre.pl
|   |     |   |_$_JIF/ush/gaugeqc_qc_precip.pl
|   |     |   |_$_JIF/ush/gaugeqc_qc_latlon.pl
|   |     |   |_$_JIF/ush/gaugeqc_pcp_r2l.pl
|   |     |   |_$_JIF/ush/gaugeqc_daily_sum.pl
|   |     |   |_$_wrkdir/pcp_qc_gz.scr
|
| Warm start:
|   |_$_JIF/ush/gaugeqc_precip_qc.pl
|   The rest is the same as for cold start.
```

1.3. List of files for the JIF

Files are located in dew:/u/wx22yl/gaugeqc

Under sub-directory ***scripts***:

exgaugeqc_qc.sh

Under sub-directory ***ush***:

gaugeqc_NWSLIreformat_EMG.pl
gaugeqc_daily_sum.pl
gaugeqc_getneigh.pl
gaugeqc_pcp_r2l.pl
gaugeqc_precip_qc.pl
gaugeqc_qc_latlon.pl
gaugeqc_qc_precip.pl
gaugeqc_qc_precip_pre.pl
gaugeqc_refnwsli.pl

Under sub-directory ***parm***:

gaugeqc_fileloc.ref

Under sub-directory ***fix***:

gaugeqc_correlate.dat
gaugeqc_normals.dat

1.4. Test runs

Cold start: "gaugeqc_run_qc_cold.pl 20070928":

Began at Mon Oct 1 10:49:31 EDT 2007

Ending at Mon Oct 1 18:59:05 EDT 2007

Run time: 8hr 10 min.

Output files covered 25 Aug – 26 Sep (33 days)

20070928- 33days = 20070826

From gaugeqc_run_qc_cold.pl, the following 33 calls were made:

gaugeqc_precip_qc.pl 20070826
gaugeqc_precip_qc.pl 20070827
....
gaugeqc_precip_qc.pl 20070927

It appears that when argument \$day is given to gaugeqc_precip_qc.pl, output (*.eval, *.neigh etc.) are done for \$daym1.

Cold start: "gaugeqc_run_qc_cold.pl 20071009":

Begin at Fri Oct 12 11:04:58 EDT 2007

Ending at Fri Oct 12 17:59:45 EDT 2007

Run time: ~7 hours (small changes have been made to the scripts today to avoid small waste. Additional variation in run time is probably related to the state of the machine).

Output files covered 5 Sep – 7 Oct (33 days)

Warm starts

20071008: run time: 22 min.
20071009: run time: 22 min.
20071010: run time: 22 min.
20071011: run time: 26 min.
20071012: run time: 22 min.
20071013: run time: 22 min.
20071013: run time: 23 min.

2. Changes to the original GSD package

2.1. Note on operational requirements

To have the job run in NCEP operations, the working directory \$WRKDIR is to be a scratch directory under /tmpnwprd. This directory will be removed at the completion of each run. To save the 'history' files, they are copied to /com/hourly/prod/gaugeqc. The directories under '/com' and '/dcom' should only be used to copy files into/out of (*e.g.* should not do gzip/gunzip there).

2.2. Changes made to the “Control scripts”

This set of scripts are kept in dew:/u/wx22yl/pcp_qc. I try to keep them as close as possible to Randy's scripts in /u/wx20rc/pcp_qc, except change 'wx20rc' to 'wx22yl', with the following exceptions:

1. The original getneigh_cold.pl points to “/public/data/precip/stations/all_dcp_defs.txt”, which does not exist on the CCS. Changed it to “\$ref{dcploc}/all_dcp_defs.txt”. This change was also made to “my” original script, *e.g.*, Randy's script with all 'wx20rc' changed to 'wx22yl', for test/comparison purposes. Oddly this change did not affect the results of runs.
2. The original qc_precip_cold.pl has this line 'open(INP,"\$ref{nwsliloc}/all_dcp_defs.txt)'. This might have worked if all_dcp_defs.txt occupied the same directory as the processed NWSLI file, but it would not have worked correctly even in Randy's set up at NCEP. I found this when I had trouble reconciling my run and the 'control run' – my run had fewer stations not found in NWSLI/DCP lists (in yyyyymmdd anl) and the 'extra' not-found stations in the control run were actually in all_dcp_defs.txt.
3. In precip_qc.pl, commented out “system('source /home/mab/collande/.cshrc);”.

2.3. Changes made to the pre-implementation scripts

1. Files needed for future runs (*.eval, &.good, *.precip and *.neigh), along with the gzip'd daily tar files, are kept in /com/hourly/prod/gaugeqc. We can send files to this

- directory, or copy from it, but we are not supposed to doing things like 'gzip/gunzip' in /com. So if we need to do anything to the files, copy them over to a working directory.
2. Use a shell script to invoke coldstart.pl or precip_qc.pl. The shell script copies over all the *.eval, *.good, *.precip and *.neigh to the working directory ('testloc').
 3. In getneigh_cold.pl, if \$today.eval or \$today.good or \$today.precip doesn't exist, we copy z\$today.tar.gz from /com/hourly/prod/gaugeqc to the working directory before gunzip it. Same with getneigh.pl.
 4. Similar logic as the above for qc_precip_cold.pl, though in this case the original code has the gunzip of the tar files in \$ref{dataloc} (that's our /com/hourly/prod/gaugeqc). Change this to what's done in getneigh_cold.pl.
 5. Same thing for qc_precip.pl
 6. In coldstart.pl and precip_qc.pl's write_scr that creates pcp_qc_gz.scr, copy the *.eval, *evalH, *.good, *.goodH, *.precip and *.neigh files to /com/hourly/prod/gaugeqc. Create daily tar files and copy them to /com/hourly/prod/gaugeqc. No need to do the "rm -f \${curdt}/*.*)" then un-tar to retrieve the few important files, since \$ref{outloc} is just a working directory that'll be scrapped later
 7. refnwsli.pl: re-formatted NWSLI file copied over to /com/hourly/prod/gaugeqc.
 8. In qc_precip_cold.pl's subroutine 'readrec', open input files in \$ref{testloc}, rather than in \$ref{dataloc}. Two occurrences of this.

3. Hourly and daily gauge data to 40-day rotating archive

This set of scripts are run each day at 03:40 UTC.

3.1. About the scripts:

1. JGAUGE_QC.sms.prod: SMS script written by Luke.
2. exgaugeqc_gauges2dcom.sh: my script, to
 1. copy and re-name the hourly precip data to 40-day rotating archive
 2. call the Michaud Perl script for the daily gauge data
3. exgaugeqc_gauges2dcom.sh: Perl script written by David Michaud to emulate Sid Katz's processing for the daily gauge data.

3.2. Notes on the Michaud Perl script

Sid's original JIF, submitted on 23 Feb 2007, had the following implementation instruction:

```
At 3:40 UTC, copy the following file on the CCS --
/com/ingest/prod/shf.yyyyymmdd/hrly.prcp.day.mmdd to
/dcom/us007003/gaugeqc/hrly.prcp.day.yyyyymmdd
```

```
Also copy the following file from the RZDM to the CCS --
/home/ftp/emc/mmb/gcp/precip/katz/usa-dlyprcp-yyyyymmdd to
```

```
/dcom/us007003/gaugeqc  
yyyyymmdd is the previous UTC day, not current UTC day.
```

etc..

David Michaud wrote a script to simulate the part of the Katz processing on the CCS (20070420). The (sub)script is at

```
dew:/u/wx22y1/gauge-michaud/load-qual-prcp.pl
```

Need to combine three different SHEF input before running the above script. The run script is at

```
dew:/u/wx22y1/gauge-michaud/run.sh (to run: run.sh $yyyyymmdd)
```

Dave's original scripts are at

```
dew:/nco/pmb/wx11dm/shef
```

Comparison of Dave's daily gauge file to Sid's, for 20070419, shows that Sid's listing has 9466 reports, while Dave's has 9783. 9438 reports are present in both listings. 28 reports are in Sid's listing but not in Dave's, all 28 are in Hawaii: the Hawaii SHEF reports are on Sid's workstation but as of now they are not on CCS (but since we are not concerned with gauge QC over Hawaii right now, it's OK). 345 reports are in Dave's listing but not in Sid's – likely because Dave's script uses Jeff Ator's station directory, which is more up-to-date than Sid's.

3.3. About the 40-day rotating archive

We had originally intended to have the entire 40-day rotating archive for gauge QC in

```
/dcom/us007003/gaugeqc
```

Which was established in early 2007 to hold the NWSLI and DCP station directories by Paula Freeman. When `exgaugeqc_gauges2dcom.sh` and `gaugeqc_load_qual_prcp.pl` were JIF'd in 15 May 2007, I noticed that the hourly and daily gauge data were placed in

```
PRODCCS:/dcom/us007003/gaugeqc
```

but not in DEVCCS. Turns out (from Luke and Brent) that data under `/dcom` are strictly imported datasets – PROD and DEV machines each imports datasets on their own (with ssh, e.g.), hence `/dcom` on PROD and DEV often does not match exactly. To have data mirrored between PROD and DEV, they need to be on `/com`. Luke changed the scripts to have the non-imported-data part of the 40-day rotating archive on

```
/com/hourly/prod/gaugeqc
```

I originally named the script `gaugeqc_gauges2dcom.sh` (Luke added the 'ex') because I intended to send the data to `/dcom`.

4. References/web links

GSD gauge QC page: <http://www-frd.fsl.noaa.gov/mab/sdb/diagnostic.cgi>

Randall Collander's README file (from RC, 30 Oct 2007)

http://www.emc.ncep.noaa.gov/mmb/ylin/parvae_cartae/collander_readme.doc

Tollerud, E., R. S. Collander, Y. Lin, and A. Loughe, 2005: On the performance, impact, and liabilities of automated precipitation gage screening algorithms. *Preprints, 21st Conf. on Weather Analysis and Forecasting/17th Conf. on Numerical Weather Prediction*, American Meteorological Society, Washington DC, 1-5 August 2005, Paper P1.42.

5. Log of latest events

2007/10/12: Found last bug (I hope). Began routine runs.

2007/10/17: JIF submitted

2007/10/29: Xiaoxue began testing gaugeqc package

2007/10/30: Xiaoxue moved gaugeqc to para testing (starting tomorrow)

2007/11/07: checked my gaugeqc run's 20071107.evalH (on dew) against that in `mist:/com/hourly/para/gaugeqc/20071107.evalH` and found them identical.