United States Department of the Interior
U.S. GEOLOGICAL SURVEY
Reston, Virginia 20192

In Reply Refer To:
Mail Stop 415

September 29, 2017

Memorandum

OFFICE OF SURFACE WATER TECHNICAL MEMORANDUM 2017.14

SUBJECT: Peak Flow File: Status and plans for resolving cleanup effort and an opportunity to preview a new report: The U.S. Geological Survey Peak-Flow File Data Verification Project, 2008–2016

This Office of Surface Water (OSW) Memo:

a. provides a pre-publication copy of a new U.S. Geological Survey (USGS) scientific investigations report documenting USGS Water Science Center efforts to "cleanup" the USGS peak-flow file (PFF);
b. suggests that WSCs review the materials;
c. and requests that WSCs provide additional information on peak flow data that the WSCs have previously identified as "questionable".

Background

USGS peak-flow data play an outsized role in modern American life. Analysis of these data by Federal, state, and local agencies and private engineers are the basis for the design of the bridges, dams, and levees, quantify the vulnerability of irrigation, water- and wastewater treatment systems to flood damages, and enable the delineation of floodplains that inform land-use codes that broadly determine where Americans live, work, shop, and play. Even a single USGS flood peak, particularly the highest and near-highest for a streamflow record, can dramatically impact the size and expense of individual infrastructure projects or lead to floodplain delineations and related zoning decisions that can nullify the building or expansion of some communities.

Data of such influence should be completely reliable. However, concerns about some USGS peak-flow data led OSW to develop a series of screening scripts and tools (Ryberg, 2008; Ryberg and Nielsen, 2014) that detect outliers and inconsistent data, and to ask Water Science Centers (WSCs) to apply the tools in a years-long effort to verify or correct data served publically via the USGS peak-flow file (PFF) (U.S. Geological Survey Office of Surface Water, 2008; URL: *https://water.usgs.gov/admin/memo/SW/sw09.01.pdf*).

Documenting the PFF Cleanup

OSW has documented the ongoing PFF clean-up effort in a forthcoming U.S. Geological Survey Scientific Investigations Report and associated data release. The report is titled *The U.S. Geological Survey Peak-Flow File Data Verification Project, 2008–2016* and was written Karen Ryberg, Burl Goree, Tara Williams-Sether, and Robert Mason. The report is based on two snapshots of the PFF (available in the data release) dated November 19, 2008, and November 14, 2016.

The report provides some history of the PFF, documents the peak-streamflow and gage-height qualification codes, and reviews the checks and changes of the PFF that have been performed by WSCs since 2008. It provides some numerical summaries of changes that have been made to the PFF, and graphically depicts many of the changes on maps. Documentation of the history, codes, checks, and revisions is useful given the planned transfer of the PFF to the new Aquarius functionality in that it summarizes the state of the PFF just before the move to the new database. The report, and the cleanup of the PFF, is especially timely given the expected release of new guidelines for flood-frequency analysis ("Bulletin 17C") and the expanded opportunities for analysis and use of the data that the guidelines herald.

This report has received three colleague reviews and is currently undergoing editorial review. The version of the report that has been colleague reviewed (but not yet approved) is being made available to WSCs now as a courtesy to provide advance notice in case cooperators or members of the public have questions about the data when the report is published. (Questions about specific peaks will be directed to the WSCs.) The draft report may be accessed at this Google Drive link (*https://drive.google.com/open?id=0B82xnkl_TO_eN0pTbmprZWpyM1U*). The data release is available at another link (*https://www.sciencebase.gov/catalog/item/58756aeae4b0a829a3276075/tokenEntry/8c431905-4e31-4324-8f4a-d7b6364ee622-15e72809e48*).

This report and the accompanying data release represent snapshots of the PFF at specific times. Doubtlessly, WSCs have made additional changes to the PFF since November 14, 2016; however, neither the new changes, nor any future corrections, will be reflected in either the report or the data release. The PFF (*https://nwis.waterdata.usgs.gov/usa/nwis/peak*), not the data release, continues as the official repository of USGS peak-flow observations and data.

Finally, an internal log of changes made to the PFF can be accessed at *http://wwwrustla.er.usgs.gov/pkentry/reports/site_list.asp?nwis_id=nwisla*, where the nwisld at the end of the URL can be replaced with the appropriated NWIS database, such as nwisnd (for North Dakota).

Resolving "Questionable Peak Flows"

The "Ryberg scripts" identified outlier or inconsistent PFF data. OSW Technical Memorandum (TM) 2009.01 requested that WSCs verify or correct the flagged data where the needed correction was clear or could be ascertained from published information (annual data reports, water supply papers, etc.) or internal files. OSW TM 09.01 also asked WSCs to label incomplete or suspect flagged data that either could not be verified as correct or that could not be corrected, as "questionable". These "questionable data" need to be addressed.

The tools that OSW provided enable WSCs to flag as questionable the entire peak-flow record; it did not enable WSCs to appropriately identify specific questionable record elements (flow, stage, date, code) or combinations of them. Nor did the tool permit WSCs to record their efforts undertaken to resolve the questionable information or recommend follow-on actions. However, it did provide a substantial comment field for the WSCs to document that kind of information. Regrettably, the use, extent, and clarity of that comment field varied greatly among the WSCs. Indeed, from the review of many of the comments, it is not clear whether some WSCs corrected or resolved questionable data, but then left the data flagged as "questionable" regardless.

An internal report of these questionable peaks, presented as a spreadsheet, is available at this Google Drive link (*https://docs.google.com/spreadsheets/d/1QyuItbVHN-Knuox_zo96ucS3V-TtcQ_IzfvQzA0VzdY/edit#gid=1720320250*).

Included in the spreadsheet are the station identification number, station name, state, district (water science center), water year, and a listing of the peak-flow data in 2008 and in 2016 (as there are cases where one or more elements were changed between 2008 and 2016, but the peak remained flagged as questionable), and comments previously provided by the WSCs.

By June 30, 2018, WSCs are asked to review the information and for each record element (date, discharge, stage, and code), and take the following actions:

1. Clarify the status of the data ("questionable", "verified as correct", "resolved and changed") in the online spreadsheet (Column R).
   a. A peak may be considered no longer questionable when it has been verified correct or has been edited based on additional sources of information such as the Annual Data Report, a Water Supply Paper (WSP), or WSC internal files, or other sources of information, such as those that might reside with a cooperating agency. In the 1960's the USGS compiled virtually all of the peak-flow data then available and published them in a series of WSPs by basin. These documents have more detail than typically provided in previous WSPs or the PFF alone and could be an especially valuable source of validation data. They are all listed in the reference section of the PFF cleanup SIR (URL link provided above).
   b. Analytical tools, such as correlation with other peaks at nearby gages may also serve as verification.
   c. Note that codes assigned to individual peaks may also provide useful information.

2. If the questionable peak is resolved, update the PFF through PKEntry. Comments entered into PKEntry should include a statement that the peak was either:
   a. Verified as correct.
   b. Resolved and changed.

   The comments should also include basic information supporting the judgment to verify or revise the peak. While verification of a peak can rely on correlation with other gages in an area, changes to peaks should generally be based on at-site information.

3. If a peak is left questionable, record in the google spreadsheet the steps taken to resolve the data issues:
    - reviewed ADR
    - reviewed WSP (list the WSP number and page(s))
    - reviewed internal files
    - reviewed Federal archive files
    - compared with nearby correlated gages
    - other - explain in comments

    Each of the first 5 actions is required for questionable peaks that, in the analyst's judgment, are important to resolve. The degree to which information should be sought to verify a peak requires judgment, as peaks that are extremely unusual may suggest a need for additional verification steps. The type of information that is missing is also relevant. The largest concern is whether the magnitude of the peak flow is accurate, followed by the qualification codes (which affect how peaks are used in flood-frequency analysis or may indicate an uncertain date). Verification of date or stage is of lesser importance.

4. If the actions above do not result in resolution, provide in the spreadsheet any recommendations for follow-on action, such as "investigate historical records", "review precipitation records", or "consult with other agencies". Additional investigations are warranted if the nature of the questionable peak is such that it may be one of the largest peaks in the record at the site and thus substantially affect flood frequency analysis. These additional studies can be done immediately or as part of a subsequently "funded" the next flood frequency project.

5. Additional options for resolution of a "questionable" peak will be provided in mid-FY2018.
    a. A new PFF will be made available by mid-FY2018 to accommodate information required for flood frequency analysis using the Expected Moments Algorithm, as described in Bulletin 17C and for use with the new database. The new PFF will allow assignment of flow intervals to individual peaks.
        - Flow intervals can be used to replace a specific questionable peak if there is confidence that the flow was within the specified range. Flow intervals are typically based on at-site information that constrains the possible flow magnitude.
        - Additional guidance will be provided on assignment of suitable flow intervals for use with EMA.
    b. In conjunction with the release of the new PFF, clarification will be provided of existing PFF qualification codes. In addition, a small number of new qualification codes will be available.
    c. Provision may be made to publically flag a peak as uncertain when steps have been taken to resolve the peak but questions remain. Further guidance will be developed in FY2018.
    d. Pending further guidance, removal from database may be a last resort and only when there is demonstrable evidence that the peak is incorrect and there is no information available to assign a credible flow interval. A peak will not be removed only because it remains questionable. If the peak is removed, records of the original peak should still be

retained, along with the evidence that it is incorrect and reasons a flow interval could not be substituted. The procedures for this drastic step will be further developed in FY2018.

WSCs are asked to complete this process by June 30, 2018. We will provide periodic updates on the status of this review of questionable peaks in the PFF.

Please contact Robert Mason (rrmason@usgs.gov) or Karen Ryberg (kryberg@usgs.gov) with concerns or questions about the SIR report before **October 10, 2017**. Please direct questions about the review and documentation of questionable peaks to Julie Kiang (jkiang@usgs.gov).

U.S. Geological Survey Office of Surface Water, 2008, Review and correction of the peak flow file: OSW Technical Memorandum 09.01, 6 p., *http://water.usgs.gov/admin/memo/SW/sw09.01.pdf*.

//signed//

Robert R. Mason, Jr.
Chief, Office of Surface Water

Distribution: GS-W All

References Cited

Ryberg, K.R., 2008, PFReports: A program for systematic checking of annual peaks in NWISWeb: U.S. Geological Survey Open-File Report 2008–1284, 17 p., *http://pubs.usgs.gov/of/2008/1284/*.

Ryberg, K.R., and Nielsen, J.P, 2014, Improvements to peak flow file validation tests and reports: Office of Surface Water Informational and Technical Note 2015.03.

U.S. Geological Survey Office of Surface Water, 2008, Review and correction of the peak flow file: OSW Technical Memorandum 09.01, 6 p., *http://water.usgs.gov/admin/memo/SW/sw09.01.pdf*.