Technical Documentation

Evaluation of the Rapid Onset Flooding Product

Draft v1.1

January 24, 2022

Katie van Werkhoven, Jason Regina, and Mike Smith

Version Control

| Version | Date | Comments |
| --- | --- | --- |
| Draft 1.0 | October 1, 2021 |  |
| Draft 1.1 | January 24, 2022 | Updated to reflect changes in ROF process since v1.0 |

# Contents

[**Contents**](#_tdx5iqz0pto3) **2**

[Background Information](#_wstwlahg6izn) **3**

[Overview of the ROF product](#_e09f74l79tst) 3

[“Observed ROF”](#_slcckns1rb9r) 5

[Overview of NWM Forecast Cycles](#_4hpbe946m2ki) 5

[Naming and Time Step Conventions](#_bdwlnf4fzzq5) 6

[Reference Time Selection (Evaluation Modes)](#_phdqqeewddw) 8

[Source of “Observed ROF” (AnA Data Selection)](#_o4g1kmkwhash) 8

[“ExtAnA" Approach](#_xnc67dvbtzix) 8

[“LatestAnA” Approach](#_vr534ok9zbjd) 11

[Evaluation Procedure](#_ca05tdetornb) **14**

[Configuration Definitions](#_c0vt2q9n950b) 14

[Summary of Steps](#_y88lpvqkeew7) 15

[Detailed Descriptions of Steps](#_ifc8h1dr56mm) 15

[Initialization Steps](#_hbi4hhbm0fs9) 15

[1 Determine Reference Time List](#_ssqstaeleeb0) 15

[2 Initialize Static Data](#_dd6brfxo1ojo) 16

[Execution Steps for each Reference Time](#_rh2nh9o46ol6) 16

[3 Get Channel Data](#_qwxdyphk8fql) 16

[4 Calculate Flow Metrics by Reach](#_5qjignj2bbiw) 16

[5 Aggregate Metrics by HUC](#_i15q70o6d8ap) 17

[6 Identify Regions of Interest](#_8b2efh6oyq7c) 17

[7 Calculate Evaluation Categories and Statistics](#_5udubn3aetp) 22

[8 Computation of Mean Areal Precipitation](#_yeyidc9y1cdm) 25

[9 Build Summary Tables and Write GeoJSON Files](#_ijx9wdpvld5w) 25

[10 Generate 15-Panel PNG Figure (Optional)](#_ep791imdcqju) 26

[Presentation and Interpretation of Results](#_i8sne94i51vn) **27**

[Known Limitations](#_sbix4m7oj2gl) **29**

[Future Work](#_poyktzs9j8ti) **29**

[References](#_vg1jsn705y5g) **30**

[Appendices](#_wt18q5tzzwzo) **31**

[Appendix A](#_b5thky50e82d) 31

[Appendix B](#_h76rb85ehtw2) 32

[Appendix C](#_k17y7w9g2dmq) 33

[Appendix D](#_vb155zer5ohb) 56

# **Background Information**

## Overview of the ROF product

The Rapid Onset Flooding (ROF) product (Figure 1 and [here](https://docs.google.com/document/d/1RbPTtV4VbIBFcG0WId1J9VZwZFFIK8lp-mfGoyHzJvc/edit)) is a derivative of the Short Range Forecast (SRF), one of the suite of National Water Model (NWM) forecasts. The ROF product indicates areas with a high potential to experience flash flooding within an 18hr forecast horizon. Two criteria must be met for a stream channel to be “in ROF.” First, the forecast flow in the reach must double within one time step (one hour). Second, the forecast flow in the reach must exceed bankfull flow within six hours of the flow doubling. ROF conditions can be met at any time within the 18hr SRF forecast horizon (see example depiction in Figure 2).

The ROF product is computed for each hourly SRF cycle. Bankfull flow is defined as the 1.5yr flow (1.67% annual exceedance probability, AEP), derived from an analysis of the NWM v2.0 multi-decade retrospective simulation. ROF is computed only on Strahler stream orders

of four or less. While ROF is computed on a NWM reach basis, it is summarized and presented on a Hydrologic Unit Code-10 (HUC10) spatial scale. The color of a HUC10 in Figure 1 indicates the total length of reaches (order four or less) in ROF conditions, divided by the total length of reaches of order 4 or less.

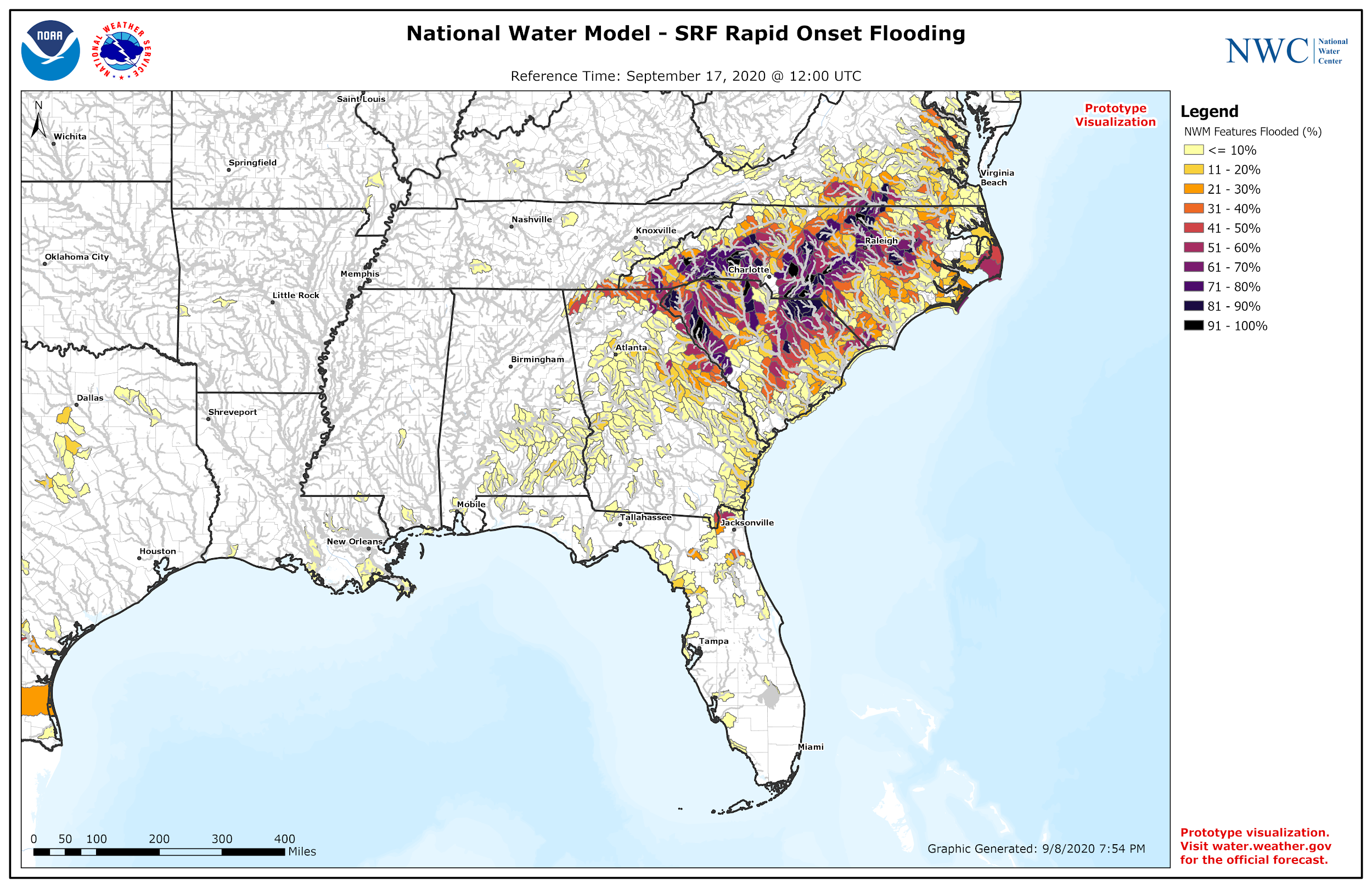


Figure 1. Example of SRF ROF product for reference time Sept 17, 2020 at 12z. Darker colors indicate HUC10 basins with a high percentage of reaches forecast to meet ROF criteria. Source: [NWC Visualization Services Version 1.4.1](https://docs.google.com/document/d/1RbPTtV4VbIBFcG0WId1J9VZwZFFIK8lp-mfGoyHzJvc/edit)

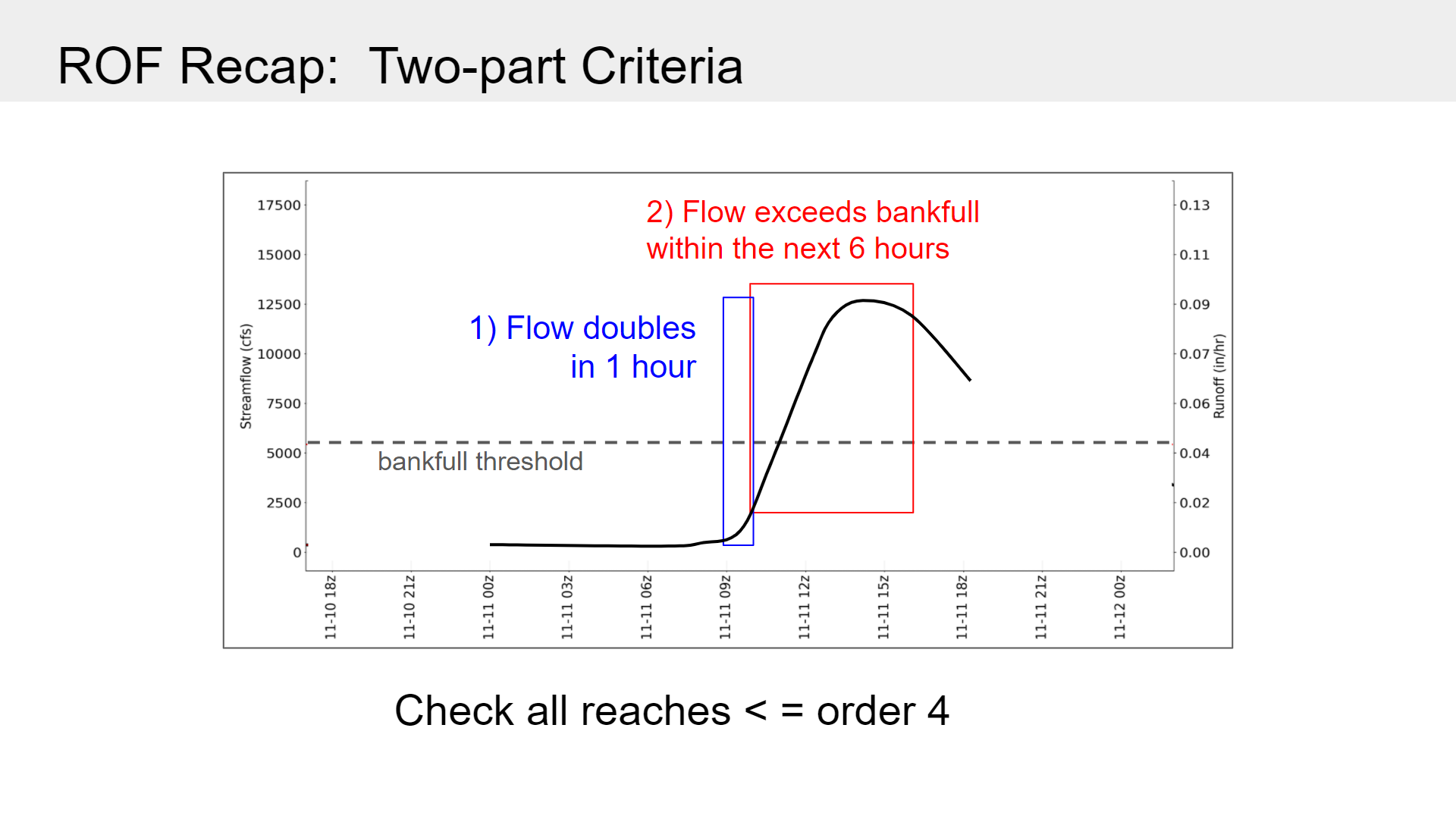


Figure 2. Example depiction of the two-part ROF criteria.

Figure 3 depicts the computation of the ROF percentage for a HUC10 basin. The reaches in red meet ROF criteria and total 3km in length. The total length of all reaches (stream order four or less) is 7km. Thus, the percentage of reaches in ROF in this HUC10 is 3/7 or 43%. The HUC10 is assigned a display color corresponding to the ROF percentage.

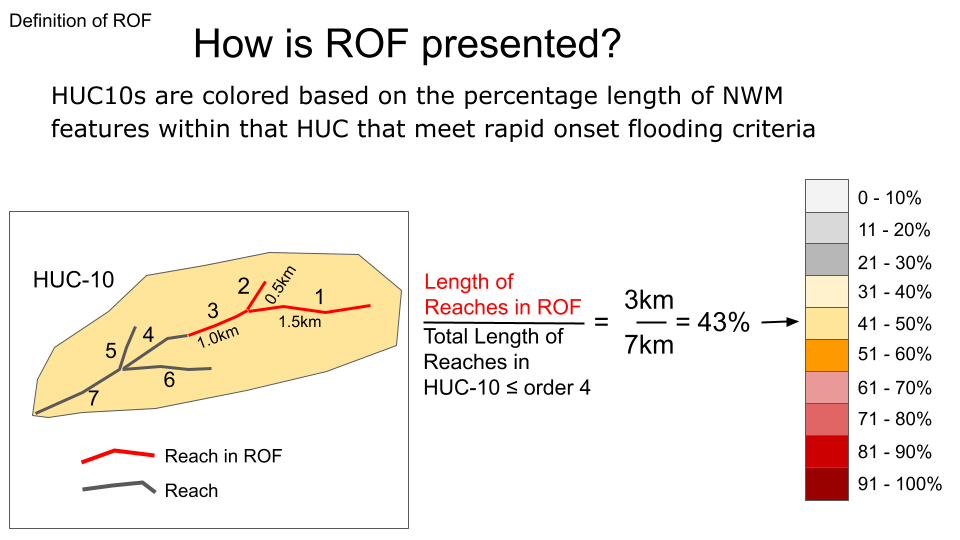


Figure 3. Example computation of ROF percentage in a HUC10 basin

## “Observed ROF”

There is no single complete set of observed data with which to evaluate the ROF product. In other words, there is no complete ‘observed ROF’ data to use for evaluation. USGS gage observations are limited in coverage. Local storm reports are available but have non-uniform spatial coverage and don’t necessarily indicate flow exceeding bankfull within six hours. Given this lack, a proxy for observed ROF is computed using the two analysis and assimilation (AnA) runs of the NWM. “Observed ROF” is computed using the NWM extended Analysis and Assimilation (extended AnA) runs or a combination of the extended AnA and standard AnA runs (see Source of “Observed ROF” below). It should be emphasized that comparing the SRF to AnA in this way provides an evaluation only of potential error in the ROF product due to precipitation forecasts. Since both the AnA (standard or extended) and the SRF are generated by the same hydrologic model, this approach ignores model error.

The procedures herein were developed using NWM 2.1 standard AnA and extended AnA simulations. However, they also work with the NWM 2.1 “open loop” standard and extended AnA simulations.

## Overview of NWM Forecast Cycles

A review of the NWM forecast cycles relevant to the ROF product is presented here as background. The spreadsheet [here](https://docs.google.com/spreadsheets/d/16vW438lAuDIANNjY48EqPv5WQfYugYEH/edit?usp=sharing&ouid=104594716320034215588&rtpof=true&sd=true) details the timing and interaction of the relevant cycles. SRF streamflow forecasts are generated every hour using precipitation forecasts derived primarily from the High Resolution Rapid Refresh (HRRR) model outputs. If these forcings are missing, precipitation forecasts are derived from the Rapid Refresh Model (RAP) outputs. Initial conditions for the SRF are derived from the standard AnA simulation. The standard AnA simulations at USGS gages ‘assimilate’ streamflow observations when available. Here, the simulation is ‘nudged’ to use the observed flow value. Observed precipitation for the standard AnA is derived mainly from the Multi-Radar Multi-Sensor (MRMS) system. The standard AnA runs are executed each hour and have a 3 hour look-back. The standard AnA runs are self-cycling, meaning that states from previous runs are used to initialize the next cycle of standard AnA. Standard AnA states at the end of the first hour of the 3-hour run are used to initialize the next cycle of the standard AnA. One exception will be discussed.

The extended AnA simulations are generated using the NCEP Stage IV precipitation estimates in an effort to have a NWM run with the same forcings used by the RFCs. The Stage IV values are also considered to be the best estimates of observed precipitation. Nelson et al. (2016) provide details of the Stage IV data. The extended AnA runs are executed every day at 16z and have a 28hr ‘look-back,’ thus starting from initial conditions at 12z the previous day. Like the standard AnA, the extended AnA simulations are self-cycling, meaning that each extended AnA simulation generates initial conditions for the subsequent extended AnA, in this case at 12z. The extended AnA simulation also drops out states at 16z each day to be used to initialize the standard AnA run starting at 16z. The extended AnA simulations also ‘assimilate’ observed streamflow values at USGS gage sites. Interested readers can find more details about the NWM configurations at https://water.noaa.gov/about/nwm and in [this document](https://docs.google.com/document/d/1EmFoSHjWx8m0nWVBlRXa2w3jogZ-0mHD26q-8oqSHR0/edit?usp=sharing).

## Naming and Time Step Conventions

The naming and time step numbering conventions in the SRF are different from those in Analysis and Assimilation runs (both extended AnA and standard AnA). Figure 4 shows that SRF forecasts (green bars) are denoted by their reference time, or the model time at which the model integrates forward in time from initial conditions. For example, the 18z SRF is the SRF forecast that uses initial states at 18z model time to integrate forward in time. The time step numbering begins at T0, or the time at which the initial conditions are valid. Time steps are numbered forward in time, such as F001, F002. On the other hand, the reference time and name for the analysis and assimilation simulations is denoted by the *end* of the run period. For example, the 19z standard AnA (blue) is that standard AnA simulation that ends at 19z and sets the initial conditions for the 19z SRF. Similarly, the 16z extended AnA (beige bar) is named for the end time of 16z, at which time it drops out initial states for the 19z standard AnA. The extended AnA drops out states at 12z each day to be used as initial conditions for the next day’s 16z extended AnA run. For both the standard AnA and extended AnA runs, time steps are numbered backward in time from T0 as in TM01, TM02, TM03, etc where TM refers to “time minus”.

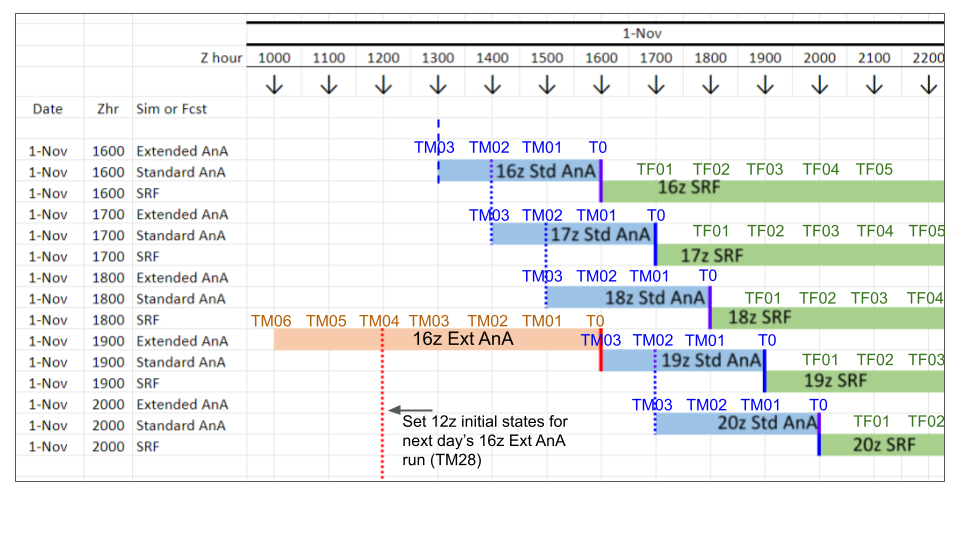


Figure 4. Naming conventions for SRF forecasts, AnA simulations, and time steps.

### 

## Reference Time Selection (Evaluation Modes)

The ROF evaluation can be run either in "current" mode or in "past" mode, which use different logic to select the forecast reference time(s) to include in the evaluation. The "past" mode requires a user to enter a date range to select one or more reference times in the past (e.g. for post-event evaluation). The "current" mode triggers a check for the most recent possible forecast(s) that can be evaluated. Hence "current" mode is used for automatic, near-real time execution. The timing and specifications of the most recent possible forecast(s) that can be evaluated in “current” mode (e.g. how far back in time and how many forecasts included) depend on which AnA version is being used to compare to the SRF for the evaluation (described below).

## Source of “Observed ROF” (AnA Data Selection)

For a given forecast reference time, the evaluation compares ROF percentages based on short range forecasts (SRFs) to those based on analysis and assimilation (AnA) output. The “LatestAnA” and “ExtAnA” options define which version of AnA is used in the evaluation, and thus the timing and frequency that the ROF evaluations can be updated in near-real time (i.e., due to the frequency and timing that new AnA data become available each day).

The “ExtAnA” option uses extended AnA data exclusively, which generates the best possible evaluation but limits updating frequency to once per day. The “LatestAnA” option uses a varying combination of standard and extended AnA simulations, based on the hour of the day, to enable more frequent updating. Currently, ISED is running both versions and generating two separate sets of output. The aspects of each AnA version are described in Table 1 and described in more detail below.

### “ExtAnA" Approach

The extended AnA configuration of the NWM runs once per day with a 16z reference time and a simulation “look-back” start time 28 hours prior (12z the previous day). The latency is close to 3 hours, thus updated output becomes available around 19z (clock time) every day.

Just after 19z, when the new model output is assumed to be available, the “Ext AnA" version of the ROF evaluation is executed. The short range forecasts included in the evaluation are those with at least one timestep that aligns with the newly available extended AnA output, but with no timesteps that extend beyond it (i.e., no timesteps with a valid time after 16z that day), since those forecasts require the next day’s extended AnA for complete evaluation of all 18 hours of the SRF.

Table 1. Characteristics of ‘Verifying’ Flow Data used in Evaluation

| Evaluation Mode → | “LatestAnA” | “ExtAnA” |
| --- | --- | --- |
| NWM flow data | Variable combination of standard AnA and extended AnA | extended AnA |
| Latency after reference time | Less than 2 hours after standard AnA reference time | Approximately 3 hours after extended AnA reference time |
| Look-back period | 3 hours for standard AnA | 28 hours; 4 hour overlap with previous day’s run |
| Forcing QPE | MRMS for standard AnA | NCEP Stage IV |
| NWM Initial conditions | From standard AnA except for 19z SRF, which uses extended AnA to initialize 16z standard AnA | From extended AnA at 12z previous day |
| Nudging to observed USGS data | Yes | Yes |

Note: -Evaluation updated only after 18-hr SRF forecast period has passed (plus any additional latency)

-NWM data posted to d-Store and Google Cloud 2 hours after run, so built-in 2-hour latency

Figure 5 depicts the logic to select the set of SRFs to be included in the “ExtAnA” evaluation for an example date (6/3). In the figure, the second column indicates the reference time, the first column is the clock time when the forecasts become available (typically 2 hrs after the reference time) and the top two rows indicate the valid time of each output time step. The darker green bar at the bottom represents the output of the current day’s (e.g. 6/3 16z) extended AnA and the dark blue bar represents the prior day’s (e.g. 6/2 16z) extended AnA output. The horizontal light blue, light green and gray bars are the SRFs. The green cells represent the SRF timesteps that align with the current day extended AnA and blue cells represent the timesteps that fall prior to the current extended AnA and thus align only with the prior day extended AnA (i.e., TM27-TM04)

The set of SRFs that are included in the current day ExtAnA ROF evaluation is indicated by the bold red polygon. The **first** (earliest)SRF included is the one for which the valid time of the *last* timestep (F018) aligns with the *first* timestep (TM27) of the current extended AnA. The **last** (latest) SRF included is the one for which the last timestep (F018) aligns with the *last* timestep (TM00) of the current extended AnA. Any forecasts after this require next day’s extended AnA output for complete evaluation.

This resulting set of SRFs included in a given ExtAnA evaluation are those with reference times beginning 19z two days prior and ending 22z one day prior. For example, the 6/3 ExtAnA evaluation includes SRFs with reference times from 6/1 19z through 6/2 22z. Note that the last four forecasts (19z-22z) use TM03-TM00 of the current forecast and thus are re-evaluated the next day using the overlapping (and assumed superior) data from TM27-TM24 of the next day’s extended AnA output.

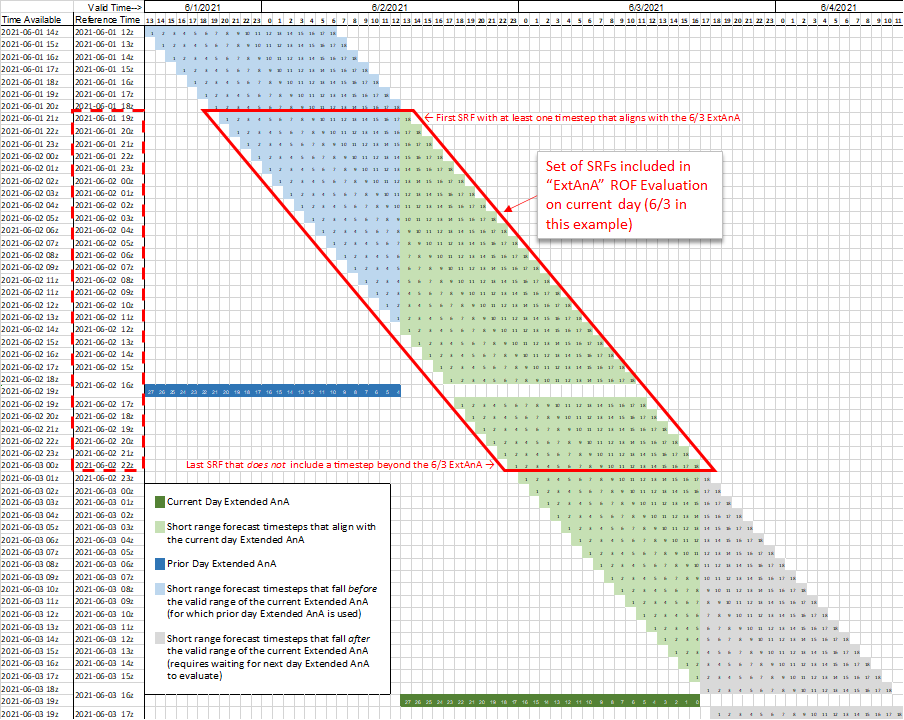


Figure 5. Depiction of the set of SRFs that are included in a given day’s “ExtAnA” version of the ROF evaluation

The extended AnA streamflow simulations are stitched together to create a continuous time series as shown in Fig. 6. The 28 hour look-back of the extended AnA results in a 4-hour overlap with the previous day’s extended AnA run. In the 4-hr overlap period, the most recent extended AnA run is selected as it is forced by more recent (and assumed better) Stage IV data (Figure 4). ROF computed from the SRF forecasts is evaluated against ‘observed’ ROF computed from the stitched-together extended AnA simulations.

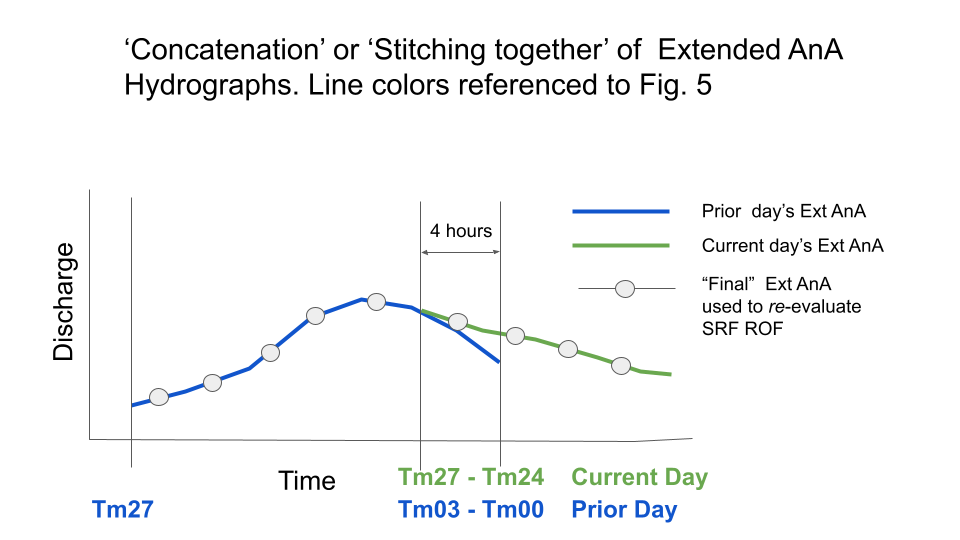


Figure 6. “Stitching together” extended AnA time series in ExtAnA mode

### “LatestAnA” Approach

In order to provide more frequent updates to the ROF evaluation, a second version was developed that uses standard AnA output for timesteps that fall after the valid range of the currently available extended AnA. For example, as shown in Figure 5, the short range forecast with a reference time of 6/2 23z (the first one below of the red polygon) includes one timestep (F018) that falls beyond the valid range of the 6/3 extended AnA. Rather than waiting for the next day’s extended AnA to become available to evaluate this forecast, the “LatestAnA” version uses standard AnA output for this timestep. Likewise, for the 6/3 00z forecast, the “LatestAnA” version uses the standard AnA for the last two timesteps (F017 and F018), and so on.

For this hourly-updated version of the ROF evaluation, the most recent SRF that can feasibly be evaluated is the reference time 20 hours prior to the current clock time, given the 18-hour forecast period and 2-hour latency time.

Figures 7 and 8 illustrate the combination of extended and standard AnA timesteps that are used to evaluate each reference time using the “LatestAnA” method for two examples (evaluating 01z and 02z SRFs). In these figures and consistent with Figure 5, the dark green cells represent the most recent extended AnA (e.g. 6/3 in the example), the light green cells are the SRF timesteps that align with the most recent extended AnA, and the gray cells are the SRF timesteps with valid times after the end of the extended AnA range. The yellow cells represent the standard AnA output for each reference time. The solid red box indicates the forecast being evaluated,, and the bold black boxes indicate the extended and standard AnA timesteps that are extracted and stitched together to create the ‘verifying’ or ‘comparison’ time series for that forecast’s “Latest AnA" ROF evaluation. Appendix C presents the remainder of the hours until the availability of the new extended AnA time series. Starting at 02z, the ‘stitching together’ of the standard AnA simulations encounters a 2-hour overlap analogous to Figure 6. The most recent standard AnA values in the overlap period are selected for ‘stitching together.’

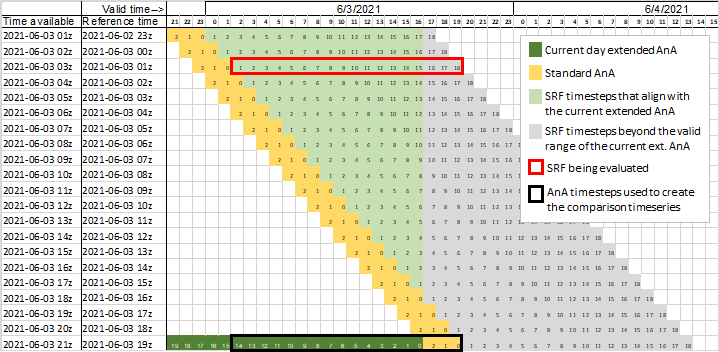


Figure 7. 01z SRF reference time (feasible to evaluate at 21z clock time the same day, when the 19z standard AnA, which was initialized by the current day’s extended AnA, becomes available. The 01z forecast is the reference time for which the greatest number of extended AnA timesteps are used for the “Latest AnA" method (since the new extended AnA becomes available at 21z clock time as well).

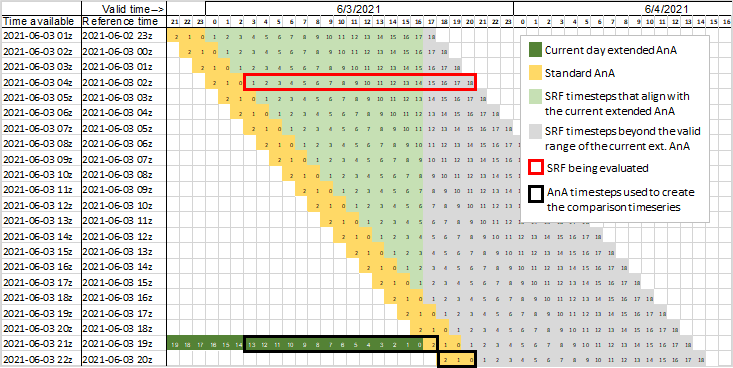


Figure 8. 02z SRF reference time (feasible to evaluate at 22z clock time the same day)

# Evaluation Procedure

The evaluation of the SRF ROF predictions involves several data processing steps and procedures, which are outlined further below with links to a more detailed spreadsheet of steps. Note that the ROF product is re-generated rather than accessing the ‘official’ ROF product generated by the National Water Center Water Prediction Operations Division (NWC WPOD). This is to streamline the computations. Testing has proven that the ROF product generated here is identical to the ‘official’ ROF product.

## Configuration Definitions

Prior to executing a ROF evaluation, multiple configuration variables must be defined. These relate to the geographic domain, local directory paths for input and output, evaluation specifications, mean areal precipitation specifications, and output specifications. Most of these settings will likely remain the same after the initial set up for execution in a given location. A complete list of the configuration definitions is provided in the “Config Definitions” tab of the procedure spreadsheet ([link](https://docs.google.com/spreadsheets/d/1ZcAxNvbZiukm9earmDlEVjka6OG_BJvv/edit?usp=sharing&ouid=112766245900592246085&rtpof=true&sd=trueQLzlngHAW/edit?usp=sharing&ouid=112766245900592246085&rtpof=true&sd=true)). Key configuration variables that are more likely to be altered are listed in Table 2.

Table 2 - Key configuration variables for a ROF evaluation

| **Variable name** | **Description** | **Possible values** |
| --- | --- | --- |
| eval\_timing | Evaluation timing ‘mode’, either most current feasible forecast(s) or specified past dates | "current" or "past" |
| ref\_start | If “past” mode, first reference time of the desired range | datetime(yyyy, m, d, h) |
| ref\_end | If “past” mode, last reference time of the desired range | datetime(yyyy, m, d, h) - set equal to ref\_start to run execute for a single reference time |
| verif\_config | AnA configuration to use as verifying observations | “analysis\_assim\_extend” or “latestana” |
| source\_list | List of data sources to access NWM output, in priority order | ''dstor”,”nomads”, and/or “google” |
| q\_thresh | Column header in feature input file of flow threshold for exceedance-based metrics (including ROF) | Any string, must match a column header in the input file. E.g. “1\_5Q” for 1.5 yr recurrence flow  “2Q” for 2 yr recurrence flows |
| event\_thresh | Spatial aggregation event threshold for a HUC-based “yes” event (e.g. 30% of reaches in ROF is an event) | Any value between 0-100 (%) |
| map\_method | MAP calculation method, either using on preprocessed spatial weights or rasterstats zonal\_statistics method | "weights" or "zonal" |

## Summary of Steps

The main steps of the ROF evaluation procedure are listed below. Brief descriptions of the steps are provided in the subsequent sections. Further details for all steps, including variable and method names, are included in a separate spreadsheet ([link](https://docs.google.com/spreadsheets/d/1ZcAxNvbZiukm9earmDlEVjka6OG_BJvv/edit?usp=sharing&ouid=112766245900592246085&rtpof=true&sd=trueQLzlngHAW/edit?usp=sharing&ouid=112766245900592246085&rtpof=true&sd=true)).

*Initialization:*

1. Build the list of SRF reference times to be evaluated
2. Initialize static data

*Repeat the following for each reference time in the list:*

1. Get NWM channel data for SRF and AnA
2. Calculate flow metrics (ROF criteria and others) for all reaches for SRF and AnA
3. Calculate % of reaches in all HUC10s that meet ROF criteria for SRF and AnA
4. Identify distinct 'objects' (clusters of HUC10s) that meet specified minimum size/magnitude criteria
5. Calculate evaluation categories and statistics for reaches and HUC10s within each ‘object’
6. Calculate Mean Areal Precipitation for HUC10s within the ‘objects’
7. Build summary tables and write GeoJSON output files
8. Generate 15-panel figure (png) (optional)

## Detailed Descriptions of Steps

### Initialization Steps

#### 1 Determine Reference Time List

Upon execution, the first step performed in the ROF evaluation is building the list of forecast reference times to be evaluated, which can include a single reference time or a range of reference times, depending on the configuration settings. The logic for selecting reference times was described previously (see “Reference Time Selection” section above).

#### 2 Initialize Static Data

The next initialization step is defining global settings that apply for all reference times and reading static input data. Global settings include the NWM version and specifications for generating PNG figures based on the domain (i.e., geographic extent) and version. Static input data include feature information (e.g. feature\_id, lat/lon of the midpoint, flow thresholds, associated HUC10, and stream length), HUC10 information (e.g. number of contained features, centroid lat/lon, area, and polygon boundary), and spatial weights relating NWM feature catchments to NWM forcing grid cells for MAP calculations (if map\_method = “weights”). A complete list of data included in the input files is provided in the procedure spreadsheet ([link](https://docs.google.com/spreadsheets/d/1ZcAxNvbZiukm9earmDlEVjka6OG_BJvv/edit?usp=sharing&ouid=112766245900592246085&rtpof=true&sd=trueQLzlngHAW/edit?usp=sharing&ouid=112766245900592246085&rtpof=true&sd=true)).

### Execution Steps for each Reference Time

After the initialization steps have completed, which should take just a few seconds, the evaluations are executed in series for each forecast reference time. Steps 3-11 described below are repeated for each reference time.

#### 3 Get Channel Data

In the current version of the ROF evaluation process, NWM channel output data are accessed by downloading NWM channel\_rt NetCDF files from a specified source to a temporary local drive and reading the necessary data from the files into memory. This step will be updated in the future to read data from the Water Resources Data Service (WRDS). To reduce storage space, only subsets of the NetCDF files are stored. The required SRF and AnA files are determined based on the reference time and selected version of AnA (“ExtAnA” or “LatestAnA” method). The logic for selecting the appropriate time steps from AnA simulations was described previously (“Source of “Observed ROF” (AnA Data Selection)”). The options for data source are currently NOMADS, DSTOR, and Google Cloud. The source(s) to use are defined by the configuration variable *source\_list*, listed in priority order.

Following download, data for all reaches are read for each timestep and stitched together into 18-hr time series. The time series data are stored in memory in separate dataframes for the forecast (SRF) and verifying observations (AnA) for use in evaluation calculations.

#### 4 Calculate Flow Metrics by Reach

Based on the 18-hr flow time series, a range of summary metrics are calculated for each reach (for SRF and AnA). Metrics calculated include the ROF criteria (i.e., if the reach meets the two-part criteria - yes/no) as well as the following intermediate flow metrics/characteristics:

* Maximum flow value
* Time step of maximum flow
* Threshold is exceeded in T0 (ongoing flood)
* Threshold exceeded in any timestep
* Threshold exceeded in all timesteps
* Number of timesteps exceeded
* Timestep of first exceedance (bankfull arrival time)
* Timestep of last exceedance
* Flow doubles at least once
* Timestep when first flow doubling occurs

#### 5 Aggregate Metrics by HUC

After reach metrics are compiled, the next step is to spatially aggregate the reach metrics by HUC polygons. The code is set up to be flexible and allow aggregation for any of the metrics above and for multiple HUC scales (HUC 8, 10 or 12). For the current version of the ROF evaluation, the aggregation is done by HUC10 and by calculating the percent of the total length of small reaches (order 4 and below) within a given HUC10 that meet the ROF criteria (see “Overview of the ROF Product” above). This is done for both the SRF and AnA, resulting in a table (dataframe) storing the percent of reaches in each HUC10 that is predicted to be in ROF conditions according to the SRF and the AnA.

#### 6 Identify Regions of Interest

In order to compare the SRF and AnA ROF results for a given event and perform a statistical evaluation, distinct regions of interest (ROI) for the evaluation must be defined. An ROI defines the geographic area over which results are spatially pooled for statistics calculations. As shown in Figure 9, a typical CONUS SRF ROF product, there are often widely scattered areas of high (darker) and low (white and yellow) ROF percentages. For evaluation, we are interested in comparing results between SRF and AnA across an area that is part of a single and reasonably significant meteorological event (i.e., spatially separate from other ongoing events). Thus a boundary must be drawn around each assumed, distinct meteorological event where *either* the SRF or AnA indicated a ROF signal. This process also removes isolated, inconsequential locations from the evaluation.

Several steps are used to resolve the data in Figure 9 into distinct ROIs (also referred to as ‘objects’). Concepts related to the Method for Objective Diagnostic Evaluation (MODE; Davis et al., 2006a,b; Davis et al., 2009; and Bullock et al., 2016) are employed to define the spatial objects. MODE attempts to delineate areas as the human eye perceives them. The steps to define the objects are described in detail below.

Note that in the current version, we are not comparing the shape and extent of ‘objects’ defined separately based on a forecast and observations. Rather, the goal here is to algorithmically find the total area to focus a HUC-10 and reach-based evaluation (i.e., can be thought of as the union where either a forecast or observed predicted a value above a threshold). Future work may involve delineating separate objects for SRF and AnA results and performing additional spatial (shape/size) comparisons. The steps performed to resolve the ROIs for the ROF evaluation are as follows:

1. Merge data (Figure 10) - Merge the SRF and AnA ROF results into a single data field containing the maximum ROF percentage for each HUC, represented as a point at the centroid location (Figure 9). This ensures that if a HUC is above a threshold % in either the SRF or AnA, it will be included in the final ROIs. The same result is generated if SRF and AnA ROIs are generated separately and the union is taken afterwards, however merging the data as the first step is more efficient.
2. Smooth data (Figure 11) - Apply a simple circular filter to smooth data by removing small scale variations and isolated outliers. The ‘radius’ R of this circular window is user-defined and currently set to 1 degree. The filter uses a moving spatial average window. The value for a given HUC is averaged with all other HUC10s’ values within distance R (centroid to centroid distance). This is repeated for all HUC10s, resulting in a smoothed data field. \*Note that the smoothed data are *not* used in the evaluation, only to define the ROI.
3. Thresholding (Figure 12) - Remove HUC10s from the data field with values below a threshold after smoothing. This removes isolated and inconsequential locations with low percentages. The user-defined value of this threshold is currently set to 1%. This step results in an initial binary ‘mask’ of HUC10s to include in ROIs.
4. Splitting (Figure 13) - Here, spatially distinct ROIs (or ‘objects’) are identified. A user-defined gap threshold is utilized to identify the distinct objects. The gap threshold is the maximum distance between two points for the two points to be considered part of the same object. The gap threshold is currently set to 1 degree (future versions of the ROF evaluation will update this threshold to units of distance (km) rather than degrees). An unique 'object number' is assigned to the HUC10s in the mask that are clustered together within a distance less than the gap threshold. As a further step to filter out inconsequential events, objects that include fewer than 5 HUC10s with a ROF % >= 10 are removed from the analysis. These criteria were defined by the WPOD based on the size of cluster that they typically consider to be important during operations.
5. Buffering (Figure 14) - This step draws a buffered boundary around the outer HUC10s included in each distinct object shown in Figure 13. This ensures that ‘true negatives’ are incorporated within and around the perimeter of the object. A true negative is the case when **both** the forecast and observed do **not** predict an event (i.e., definition of ‘event’ further discussed in Step 7). As such, true negatives are a valid result and should be accounted for in the computation of categorical metrics.

To draw the initial boundary, this step uses the concept of an ‘alpha shape’, which is similar to stretching a ‘rubber band’ around the outer extent of a given cluster of HUC centroid points. The parameter 𝛼 determines how closely the border follows the cluster edges and here is set to a value of 1. (More information: [https://en.wikipedia.org/wiki/ Alpha\_shape](https://en.wikipedia.org/wiki/Alpha_shape)). All HUC10s within the resulting boundary are included in the final region, thus interior true negatives (appearing as ‘holes’ in Figure 13) are filled. The boundary of the alpha shape is also expanded outward by a user-specified buffer distance (currently set to 0.5 degrees) using the Buffer method of the Shapely python package (see section “Constructive Methods” on page https://shapely.readthedocs.io/en/stable/manual.html). Note that by this approach, the boundary will not exactly correspond to edges of HUC10s. All HUC10s with a centroid that falls within the boundary are assumed to be included in the ROI. The final result is one or more filled ROI masks (Figure 14). (Note, a further area of research is to analyze the impact of the buffer distance on the number of true negatives and subsequently the values of categorical metrics).

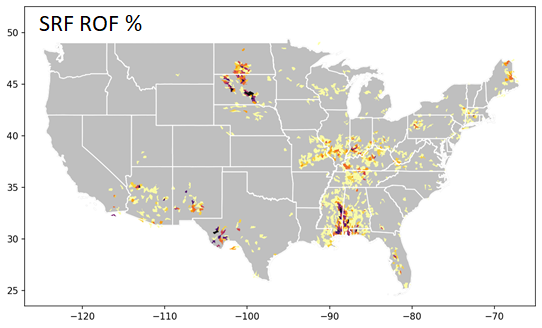


Figure 9. Example SRF ROF product. Note the widespread, scattered areas with low percentages (e.g., WI, MI)

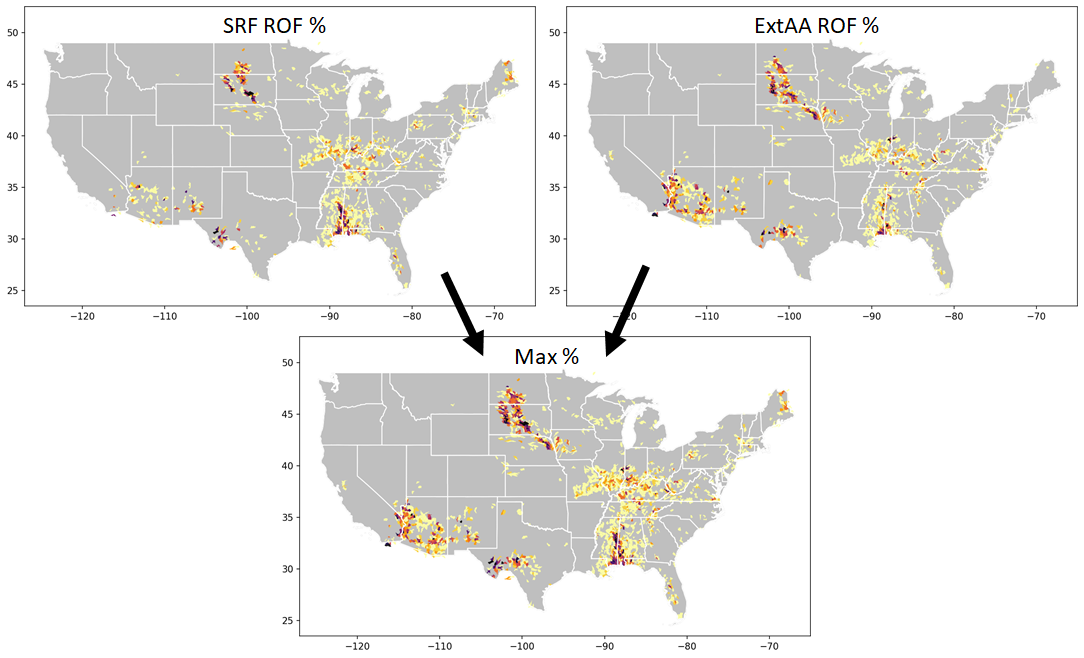


Figure 10. Depiction of Step 6A. For each HUC10, the maximum ROF % from either the SRF or the AnA is selected and merged into a single data field so that evaluation regions are identified where *either* the SRF or AnA indicated a ROF signal.

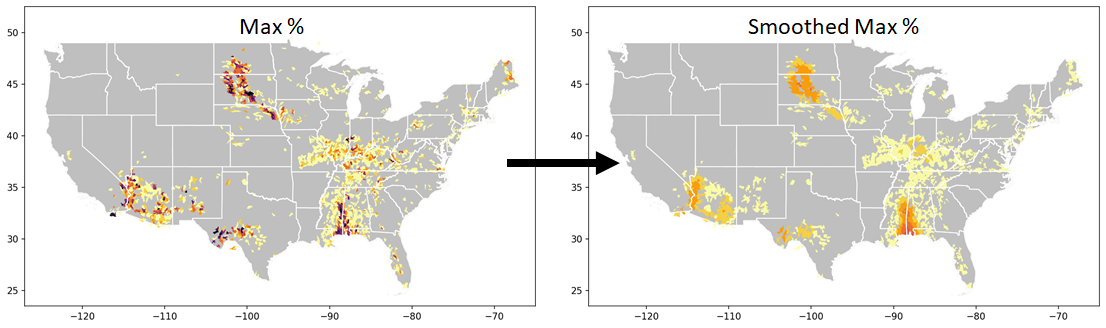


Figure 11. Depiction of Step 6B. Smoothing the maximum ROF % data field.

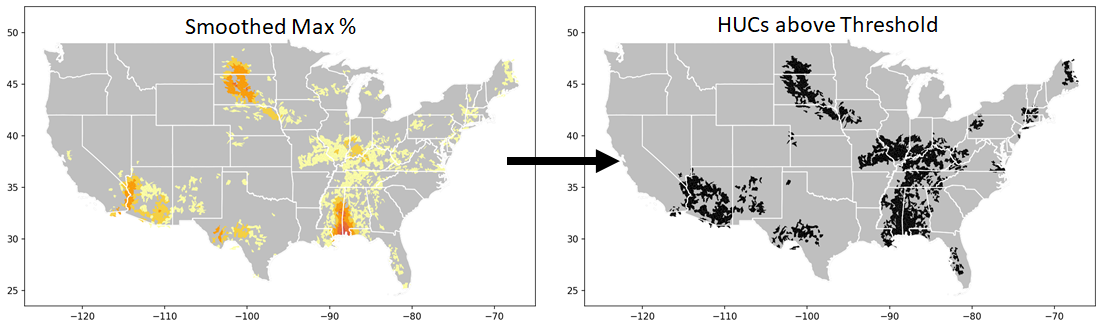


Figure 12. Depiction of Step 6C. Application of a threshold to remove HUC10s with low (inconsequential) ROF percentages after smoothing, e.g., the isolated locations in WI and MI drop out.

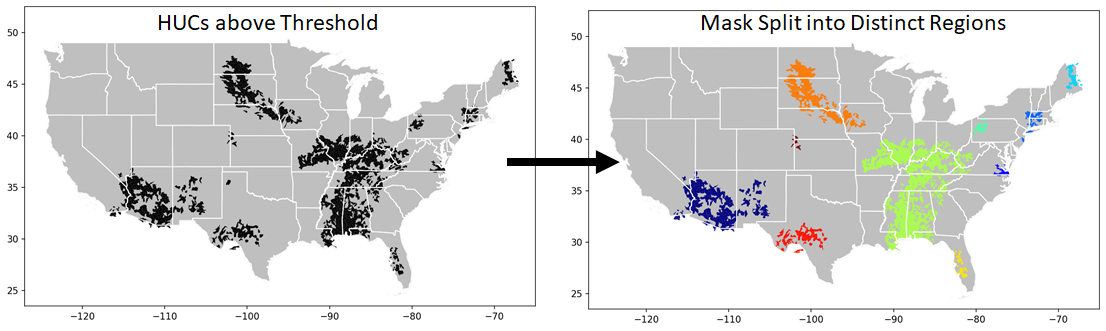


Figure 13. Depiction of Step 6D. Split the thresholded mask into distinct regions aiming to represent separate meteorological events, based on user-defined separation distance. Each color is a distinct region/object. ROIs that do not meet minimum criteria (at least 5 HUC10s greater than 10%) are dropped (e.g. the small cluster in the TX panhandle).

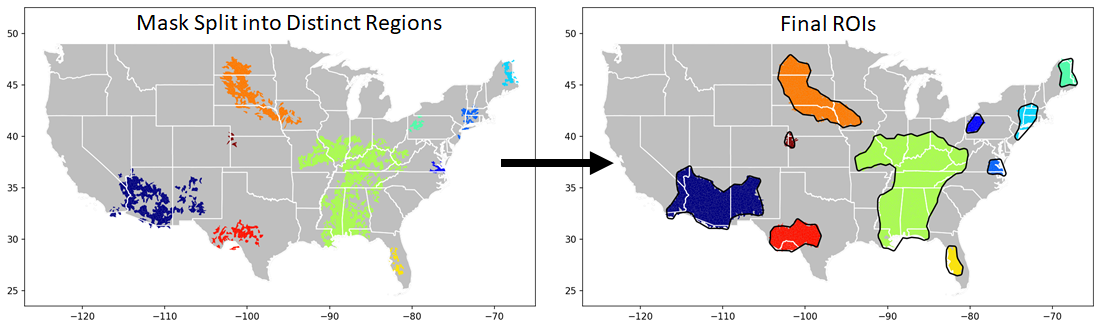


Figure 14. Depiction of Step 6E. Boundaries of the filled and buffered final ROIs.

As described throughout the sections above, a number of parameters and/or thresholds were necessary to define for the derivation of ROIs. Table 3 summarizes the thresholds used in the ROI derivation.

Table 3. Thresholds used in the derivation of the Region of Interest

| Threshold | Description | Value in Ver 1.1 |
| --- | --- | --- |
| Filtering Radius | Fill in holes; make area more contiguous | 1 degree |
| Filtering Threshold | Remove inconsequential HUC10s | 1 % |
| Gap Threshold | Identify spatially independent events | 1 degree |
| Minimum Cluster Size | Remove inconsequential objects | At least 5 HUC10s with >=10% |
| Alpha 𝛼 | Draw outer boundaries around clusters of HUC10s to define objects | 1 |
| Buffer Distance | Extend ROI to consider true negatives; fill holes | ½ degree |

#### 7 Calculate Evaluation Categories and Statistics

Once the ROI(s) has been defined for a given forecast reference time, the evaluation of the SRF ROF product within each ROI can proceed. Categorical evaluations are performed at both the reach level (does/does not meet ROF criteria) and at the HUC level (ROF percentages).

A. Defining “Yes” Events

As a first step, for each scale, the conditions that constitute a ‘yes’ event are defined. For reaches, a ‘yes’ event is simply meeting the ROF criteria (and ‘no’ if it does not). For the HUC-level evaluation, however, the percentage value that constitutes a ‘yes’ event is somewhat subjective. This is analogous to defining a high-flow ‘yes’ event for a standard streamflow forecast evaluation. In that case a ‘yes’ event is when the predicted and/or observed flow exceeds a predetermined threshold, say out-of-bank flow or a site-specific flood level. Similarly, a ROF value for a HUC10 is classified as a ‘yes’ event when the ROF percentage exceeds a selected threshold. In selecting this threshold value, questions arise such as: What level of ROF percentage in a HUC10 is important to forecasters? For the current version, WPOD selected a threshold of 30% to define a HUC10 as a ‘yes’ event. Future work may include testing the impact on the evaluation when higher or lower percentages are selected for this threshold. For quick reference, Table 4 highlights the definition of ‘yes’ events for reaches and HUC10s. Appendix D presents the impact of using thresholds of 20% and 50% to define a HUC10 ‘yes’ event.

Table 4. Definition of “yes” events for reach and HUC-level ROF evaluations

| Scale | Value in Ver 1.1 |
| --- | --- |
| Reach-level ‘yes’ event | Meets ROF criteria |
| HUC-level ‘yes’ event | ≥ 30% reach length in ROF conditions |

B. Comparing Predicted and Observed ROF

*HUC10 Analyses*

For each SRF reference time, the predicted event category (yes/no) for all HUC10s in the ROI are compared to the AnA categories, as shown in Figure 15. A ‘hit’ is defined when a HUC10 is a ‘yes’ event in both the SRF and AnA (green ellipses). A ‘miss’ is when a HUC10 is observed to be a ‘yes’ event but the SRF does not predict a ‘yes’ (red ellipses). A false alarm occurs when the extended AnA predicts a HUC10 to be a yes event, but it is not observed (yellow ellipses). True negatives are where the SRF and extended AnA are both ‘no’ and denoted by grey areas within the ROI.

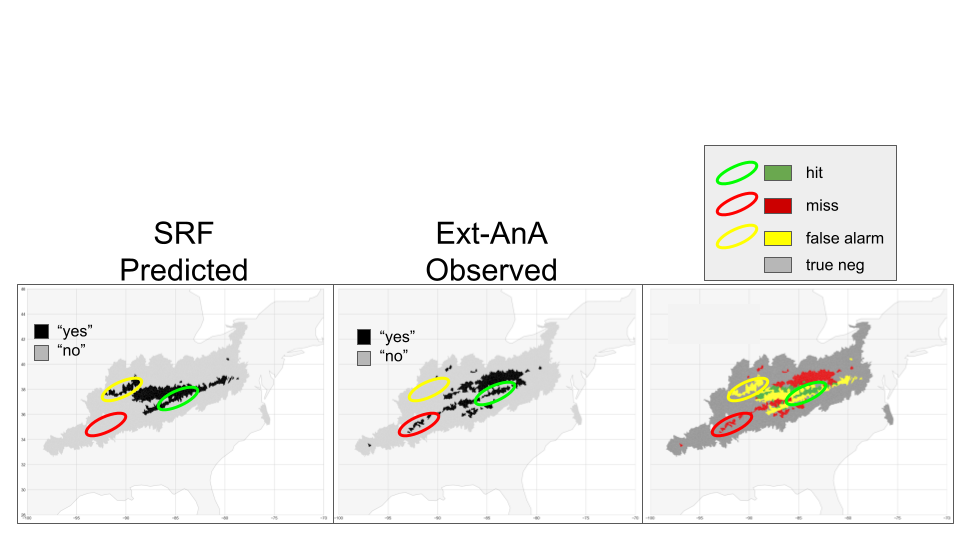
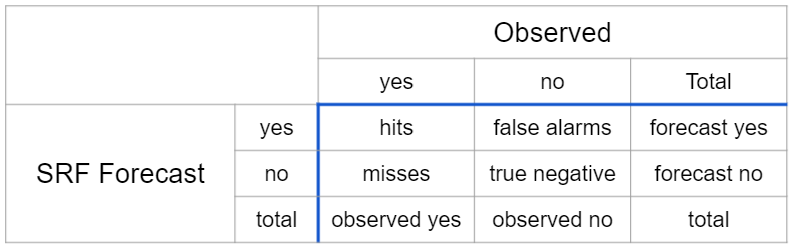


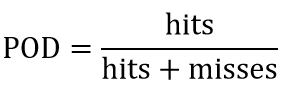
Figure 15. Comparing SRF yes/no HUC10s with ExtAnA yes/no HUC10s.

The hits, misses, false alarms, and true negatives can be tabulated in a standard 2x2 contingency table as shown in Table 5. From these, categorical metrics can be computed such as Probability of Detection (POD), False Alarm Ratio (FAR) and Critical Success Index (CSI) (source: <https://cawcr.gov.au/projects/verification/>). The categorical metrics are computed for all of the defined ROIs, however only the metrics for the largest ROI are shown on the multi-panel PNG figure (see Step 10).

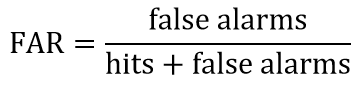
Table 5. 2x2 contingency table



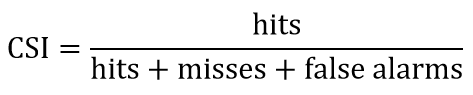
The Probability of Detection (POD) answers the question: What fraction of observed ‘yes’ events were correctly forecast? It is also called the ‘hit rate.’ POD ranges from 0 to 1, with a perfect score being 1.



The False Alarm Ratio (FAR) answers the question: What fraction of forecast ‘yes’ events did not occur (were false alarms)? FAR ranges from 0 to 1, with a perfect score being 0.



The Critical Success Index (CSI) answers the question: How well did the forecast ‘yes’ events correspond to the observed ‘yes’ events? It is the fraction of observed and/or forecast events that were correctly predicted. CSI (number of correct flood forecasts/(number of floods plus incorrect flood forecasts). CSI is also called the Threat Score. It ranges from 0 to 1 with a perfect score being 1. CSI can also be thought of as the ratio of correctly warned events out of all warnings issued and unwarned events



C. Reach-Level Evaluations

Evaluations are also performed at the scale of individual NWM reaches, some of which are also USGS gage locations. Note that since the AnA simulations are ‘nudged’ to USGS data at gage locations, the evaluation at those reaches effectively compares the SRF to gage data. Reaches that meet the ROF criteria in the SRF forecasts and AnA simulations are identified as ‘yes’ events. Note, as mentioned previously, in this case there is no need to prescribe a threshold as in the HUC10 analyses. In addition, USGS gages that meet the ROF criteria are identified. These locations are a subset of the reaches that meet the ROF criteria. Cross classification identifies hits, misses, false alarms, and true negatives at all reaches, including the USGS gage locations. The same contingency table metrics used at the HUC10 scale are computed for the reach and USGS gage analyses.

#### 8 Computation of Mean Areal Precipitation

Mean Areal Precipitation (MAP) values are computed as part of the ROF evaluation to enable viewing differences in SRF and AnA forcing alongside differences in ROF. This helps to understand the results of the ROF comparison since, as previously emphasized, the primary source of differences between the SRF and AnA is the precipitation forecasts. The SRF and AnA MAP values are computed for the total 18 hour span of the SRF. Recall that the SRF forecast precipitation is derived from the HRRR and RAP forecasts, while the AnA precipitation is computed using the NCEP-derived Stage IV data (Extended AnA) or MRMS data (Standard AnA). Note that for the ExtAnA evaluation mode, since only the extended AnA simulation output is used, the MAP values are computed only from the Stage IV for the 18 hour SRF span. For the LatestAnA mode, the source of the AnA MAP (standard or extended) for each time step matches the source of the channel data (see Figures 7, 8 and Appendix C). Thus, the 18-hr MAP for the LatestAnA mode is comprised of varying combinations of Stage IV values used to generate the extended AnA simulations and MRMS data used to force the standard AnA simulations.

#### 9 Build Summary Tables and Write GeoJSON Files

As a final processing step, the results from Steps 1-8 are compiled into summary geodatabases in memory and written to a local drive as GeoJSON files for subsequent visualization as desired. Two separate GeoJSON files are generated - one for reach-level output and one for HUC10-level output. The filename convention is as follows for reach and HUC10 output respectively:

* “ROF\_HUC10\_\_*domain\_yyyymmdd\_hhz\_AnAmethod*.geojson”
* “ROF\_REACH\_*domain\_yyyymmdd\_hhz\_AnAmethod*.geojson”

(*AnAmethod* is “ExtAnA” or “LatestAnA”)

The HUC10 files contain the following fields:

| **Field** | **Description** |
| --- | --- |
| HUC10 | HUC ID as string |
| tot\_feats | Number of features in the HUC |
| SRF\_MAP | SRF total accumulated precip |
| AnA\_MAP | AnA total accumulated precip |
| MAP\_DIFF | Difference between SRF and AnA precip |
| SRF\_pct | SRF ROF % |
| AnA\_pct | AnA ROF % |
| perc\_diff | Difference between SRF and AnA ROF % |
| SRF\_event | SRF 30% event exceedence (yes/no) |
| AnA\_event | AnA 30% event exceedence (yes/no) |
| matrix | Contingency category for 30% event (1=TP, 2=FP, 3=FN, 4=TN) |
| matrix\_x | Contingency category as a string (TP, FP, FN, TN) |
| geometry | Geometry of the polygon |

The reach files contain the following fields:

| **Field** | **Description** |
| --- | --- |
| feature\_id | Feature\_id of the NWM reach |
| order | Stream order |
| length\_m | Reach length in meters |
| gage\_str | Gage ID as a string (or ‘none’) |
| HUC10 | HUC10 within which the reach is located |
| srf\_rof | Reach meets the ROF criteria in the SRF (true/false) |
| exana\_rof | Reach meets the ROF criteria in the AnA (true/false) |
| obj | Object the reach is located within |
| matrix | Contingency category for 30% event (1=TP, 2=FP, 3=FN, 4=TN) |
| matrix\_x | Contingency category as a string (TP, FP, FN, TN) |
| geometry | Geometry of the polygon |

#### 10 Generate 15-Panel PNG Figure (Optional)

For quick visualization of results, the ROF evaluation includes the option to output a static PNG file containing a 15-panel figure for each reference time evaluated. The next section discusses the visualization approach in the PNG file as created for an example flood event. Note in this example the figure is shown as two separate figures and is zoomed in to a specific location. The default extent for the PNG file is the full domain (e.g. CONUS), however zoom specifications can be specified as configuration variables (see *xcent, ycent, xext* in Sheet “Config Definitions” in the procedure spreadsheet [[link](https://docs.google.com/spreadsheets/d/1ZcAxNvbZiukm9earmDlEVjka6OG_BJvv/edit?usp=sharing&ouid=112766245900592246085&rtpof=true&sd=trueQLzlngHAW/edit?usp=sharing&ouid=112766245900592246085&rtpof=true&sd=true)]).

# Presentation and Interpretation of Results

An important aspect of the ROF evaluations is to display the results in a manner that facilitates interpretation and use by forecasters. Initial ideas are presented in Figures 16 and 17, which were developed to analyze flooding in Renssalaer County, NY. Since the evaluation was performed days after the event had passed, the ExtAnA output was used. The county boundary is shown in white. The forecast is for the SRF 18 hour horizon beginning with the reference time of July 18 at 8z.

Forecast and observed precipitation are plotted in the top row of Figure 16. The left panel shows mean areal precipitation (MAP) values derived for each HUC10, computed using the SRF forecast precipitation. The MAP values are the sum over the 18 hour period of the SRF forecast horizon. Similarly, the middle panel depicts the ExtAnA precipitation based on the NCEP Stage-IV product. The ExtAnA precipitation is also expressed as MAP for each HUC10, summed over the same 18 hours spanning the SRF forecast horizon. The right panel shows the difference in the MAP values (predicted minus ‘observed’).

The source of precipitation (SRF forecast or extended AnA) is the only factor driving the difference between the SRF ROF and the extended AnA ROF values. Thus, it is important to understand the agreement between the two precipitation forcings. If the SRF forecast precipitation does not align spatially or agree in amount with the observed precipitation, then of course the ROF evaluation metrics will be lower. In this example, the predicted precipitation is less than the observed by about 1 inch, especially in the southern and southwestern part of Rensselaer County.

The middle row of Figure 16 depicts the SRF predicted and extended AnA observed ROF values by HUC10 for the 18 hour forecast horizon. The predicted ROF percentages are low compared to the extended AnA observed values, especially along the southern border of the county, reflecting the difference in precipitation. The right panel in this row shows the difference in HUC10 percentages; in this case an underprediction within the county. A slight overpredication of ROF can be seen in the northeast portion of the county.

The bottom row of Figure 16 shows the ‘yes’ and ‘no’ events on a HUC10 basis. Recall that the threshold for a yes event is 30%. The black areas depict the HUC10s with ROF values greater than or equal to 30%. The lower right panel depicts the cross classification of the yes and no ROF HUC10s. Green, red, and yellow colors denote hits, misses, and false alarms, respectively.

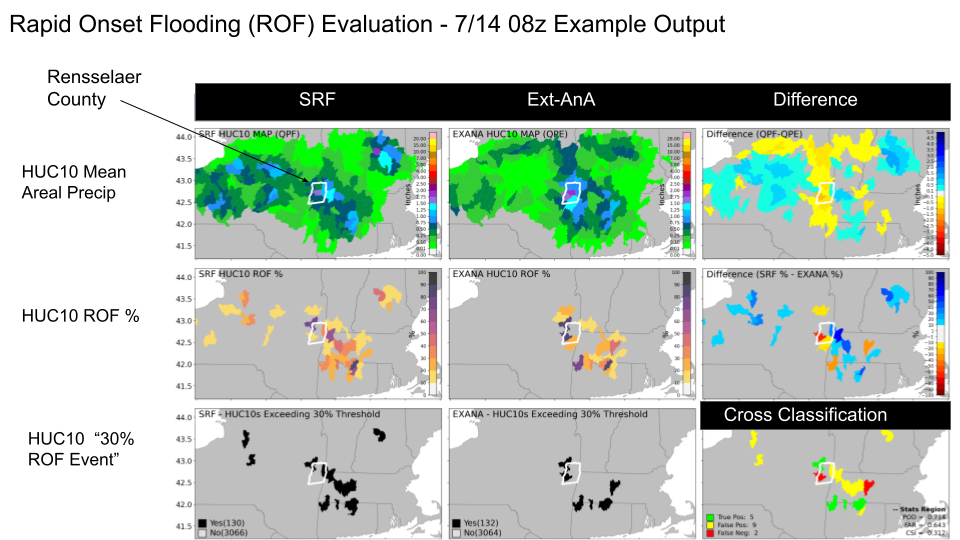


Figure 16. ROF evaluation results on a HUC10 basis.



Figure 17. ROF evaluations at the NWM reach-level (top row) and at USGS gage locations (bottom row)

Reach-level evaluation can be presented as shown in Figure 17. The top row depicts the individual reaches in the ROI that meet the ROF criteria in the SRF and the extended AnA. The reaches are depicted as dots marking the midpoint of the reach feature. The top right panel shows corresponding categorical metrics. The bottom row shows USGS gages that meet the ROF criteria. The USGS gages are a subset of the reaches in the top row. In this example, there are very few USGS gages in the ROI, and none of them recorded hits for this 18hr SRF forecast.

Figures 16 and 17 present the analyses for one SRF reference time. Such analyses and plots are generated for each SRF reference time in a storm period to provide a complete picture of the storm progression. The slides [here](https://docs.google.com/presentation/d/1yKxTX1EEXdP_zHn_NVqjEguDUMTJ7X30QyhKYX9ubcs/edit?usp=sharing) show the analyses for each reference time in the Renssalaer flood.

Interpretation Cautions

The evaluation of ROF for each SRF reference time in a storm can only answer the question: “How well did the SRF ROF predictions align with the analysis for this reference time?” Viewing the sequence of evaluations across multiple reference times can answer the question: “Did any of the ROF forecasts for this event provide any prior indication of potential flooding?” The single-storm analyses described above cannot be used to infer future performance of ROF predictions. Evaluations using large sample sizes are needed to compute statistics that can be used to infer future ROF performance.

# Known Limitations

It is worth noting several aspects of the forecast process and their impact on the ROF evaluations. First, the ROF product does not recognize when the SRF flow at T0 is already above bank full; it will only identify when the flow doubles in one time step and subsequently exceeded bank full within 6 hours.

At USGS or other streamflow gages, the standard and extended AnA simulations are ‘nudged’ to equal the observed discharge. This nudging may result in sudden hydrograph rises when the simulated flow is less than the observed flow. Consequently, nudging may result in a false alarm of ROF conditions in a reach.

# Future Work

The approach to defining the ROI uses the ‘tunable’ parameters in Table 3. The sensitivity of the ROF evaluations to the values of these parameters should be investigated. As an example, Bullock et al., (2016) provide a brief illustration of the impact of changing the filtering radius R and the threshold T in the analysis of precipitation fields. In general, they found that increasing R resulted in larger and smoother objects. Increasing T led to smaller areas of more intense rainfall.

It may be possible to expand the matching of predicted and observed ROF areas to include a timing aspect as in Davis et al. (2006b). They defined forecasted and observed rain objects by centroid coordinates at a specific time. If the forecasted object had its centroid in the correct location but occurred at some other time (within a defined time span), then the error was one of timing and not location. This allowed for the analysis of rain systems over the life of the event. Such analyses may not be directly amenable to the ROF evaluation since the ROF product specifies that flooding can occur at any time within the 18-hour forecast horizon of the SRF.

Evaluations of ROF based on large sample sizes are needed to generate inferential statistics. Such statistics could be used by WPOD to place a measure of confidence in their guidance products.

# References

Bullock, R., Brown, B., and Fowler, T., 2016. Method for Object-based Diagnostic Evaluation.

NCAR Technical Note NCAR/TN-532+STR 2016-11

Davis, C., Brown, B., and Bullock, R., 2006a. Object-Based Verification of Precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. Monthly Weather Review, Vol. 134 (7), 1772-1784. <https://doi.org/10.1175/MWR3145.1>

Davis, C., Brown, B., and Bullock, R., 2006b. Object-Based Verification of Precipitation Forecasts. Part II: Application to Convective Rain Systems. Monthly Weather Review, Vol. 134 (7), 1785-1795, <https://doi.org/10.1175/MWR3146.1>

Davis, C.A., Brown, B.B., Bullock, R., and Halley-Gotway, J., 2009. The Method for Object-Based Diagnostic Evaluation (MODE) Applied to Numerical Forecasts from the 2005 NSSL/SPC Spring Program, Weather and Forecasting, Vol 24, 1252-1267. doi:10.1175/2009WAF2222241.1

Nelson, B.R., Prat, O.P., Seo, D.-J., and Habib, E., 2016. Assessment and implication of NCEP Stage IV Quantitative Precipitation Estimates for product intercomparisons. Weather and Forecasting, Vol. 31, 371- 394, doi: 10.1175/WAF-D-14-00112.1

# Appendices

## Appendix A

Acronyms

1. AEP Annual Exceedance Probability
2. AnA Analysis and Assimilation
3. CONUS Contiguous United States
4. CSI Critical Success Index
5. ExtAnA Extended analysis and assimilation ROF evaluation mode
6. FAR False Alarm Rate
7. HRRR High Resolution Rapid Refresh
8. HUC10 Hydrologic Unit Code 10
9. LatestAnA Latest analysis and assimilation ROF evaluation mode
10. MAP Mean Areal Precipitation
11. MRMS Multi-Radar Multi-Sensor
12. NCEP National Centers for Environmental Prediction
13. NWC National Water Center
14. NWM National Water Model
15. POFD Probability of False Detection
16. POD Probability of Detection
17. QPE Quantitative precipitation estimate
18. QPF Quantitative precipitation forecast
19. RAP Rapid Refresh Model
20. ROC Relative Operating Characteristic
21. ROF Rapid Onset Flood
22. SRF Short Range Forecast
23. StandardAnA Standard Analysis and Assimilation
24. TF Time forward
25. TM Time minus
26. T0 Time Zero
27. USGS United States Geological Survey
28. WPOD Water Prediction Operations Division
29. WRDS Water Resources Data Service
30. Z Zulu time, same as Universal Time Coordinated

## Appendix B

Time Definitions

1. Available time: The clock time at which the forecast or simulation becomes available for use
2. Forecast horizon: the time span from the reference time to the last valid time
3. Latency: the time span from reference time to the available time. For example, the 16z SRF is generated at 16z but there is a 2-hour delay to the time when the forecast becomes available on NOMADS.
4. Reference time:
   1. Forecast. the model time at which the model starts to integrate forward in time. Also called “T zero” or T0.
   2. AnA. The last valid time of the simulation. For example, the reference time for the 16z extended AnA is 16z of the present day, even though the computations began at 12z the previous day.
5. Valid time: The time at which a simulated or forecast value is expected to occur.

# 

# 

# 

# 

## Appendix C

Progression of SRF forecasts and stitching together of standard AnA and extended AnA streamflow simulations for the LatestAnA evaluation mode. These figures continue Figures 7 and 8 to describe the combination of standardAnA and ExtAnA simulations from which to compute the proxy ‘observed’ ROF for the LatestAnA mode of evaluation.

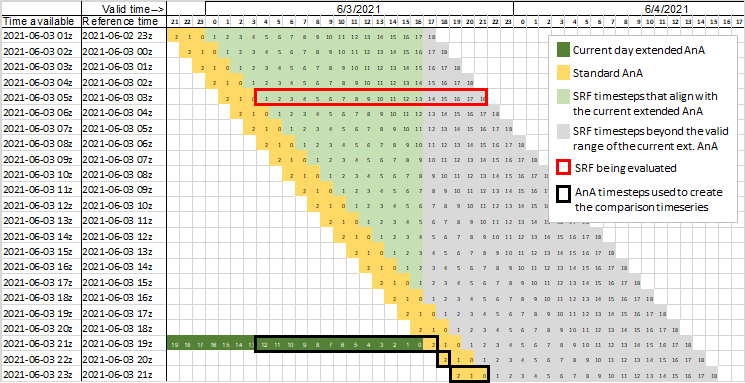


Figure C-1. 03z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 23z the same day)

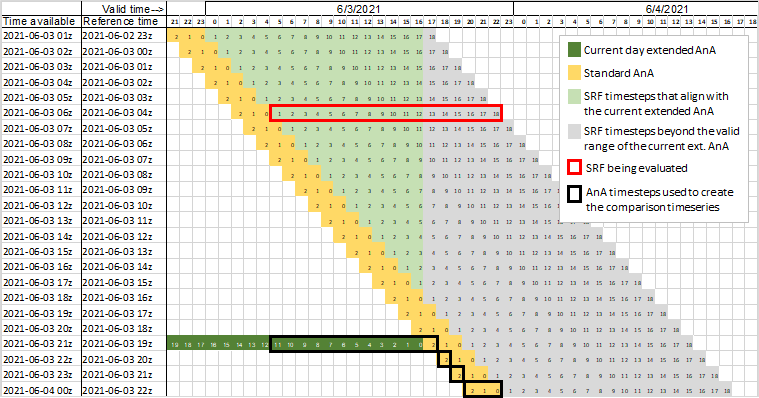


Figure C-2. 04z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 00z the next day)

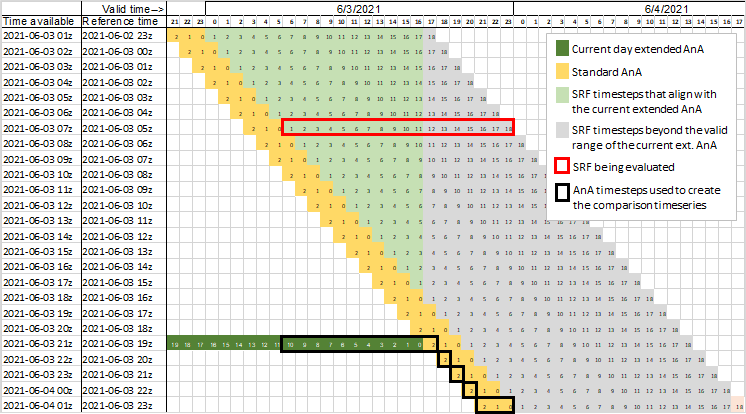


Figure C-3. 05z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 01z the next day)



Figure C-4. 06z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 02z the next day)

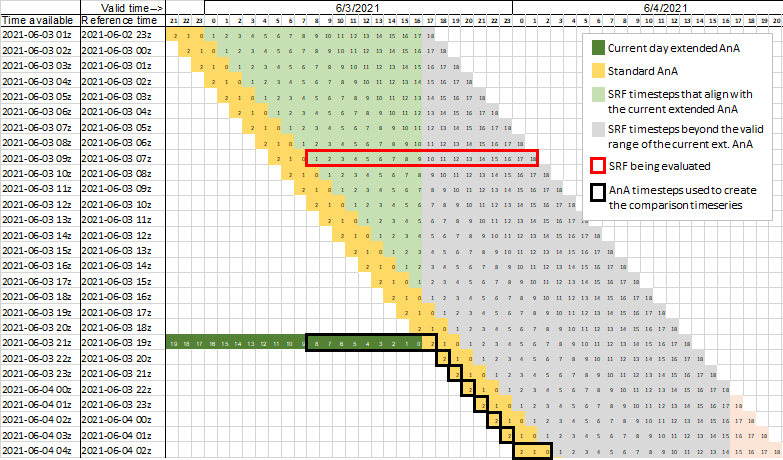


Figure C-5. 07z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 03z the next day)

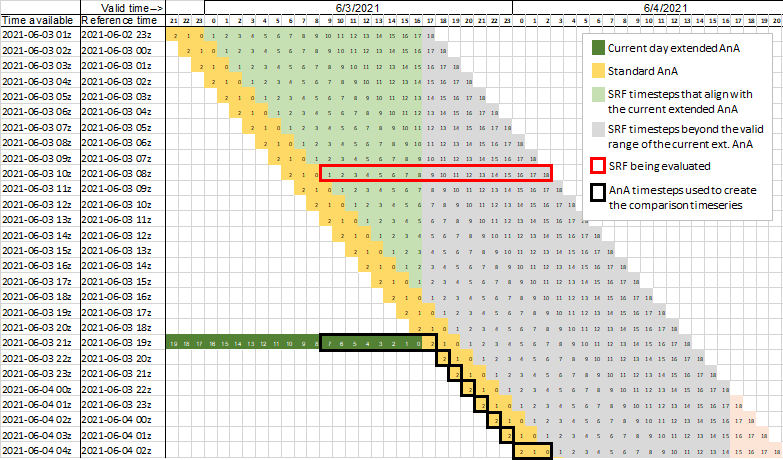


Figure C-6. 08z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 04z the next day)

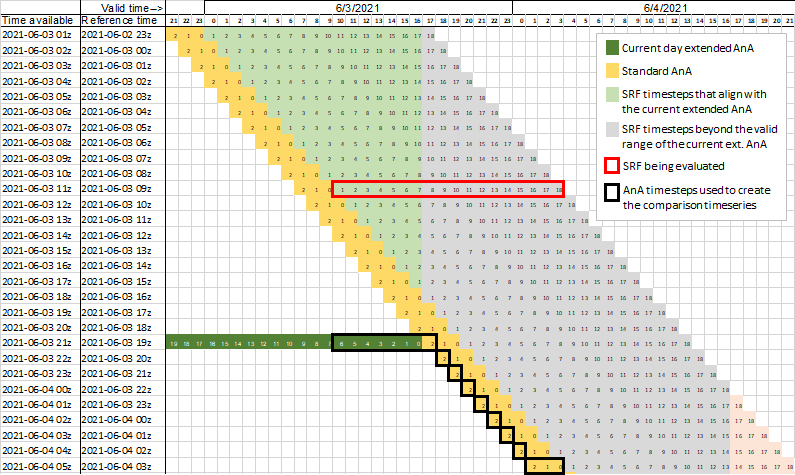
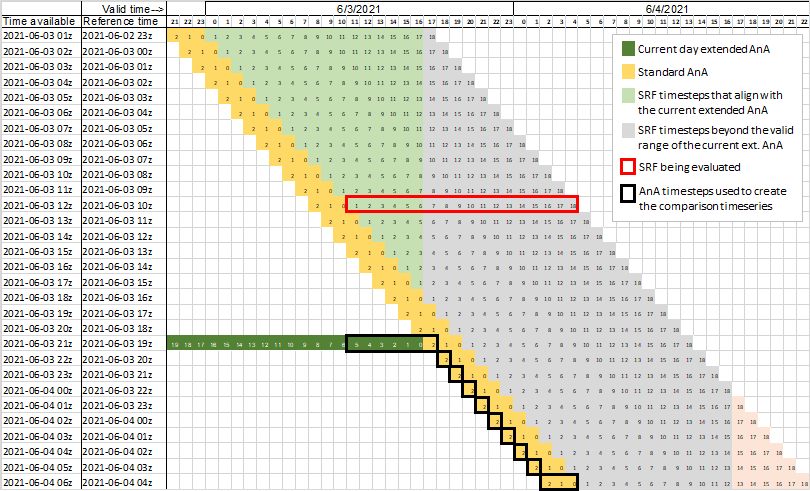


Figure C-7. 09z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 05z the next day)

Figure C-8. 10z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 06z the next day)

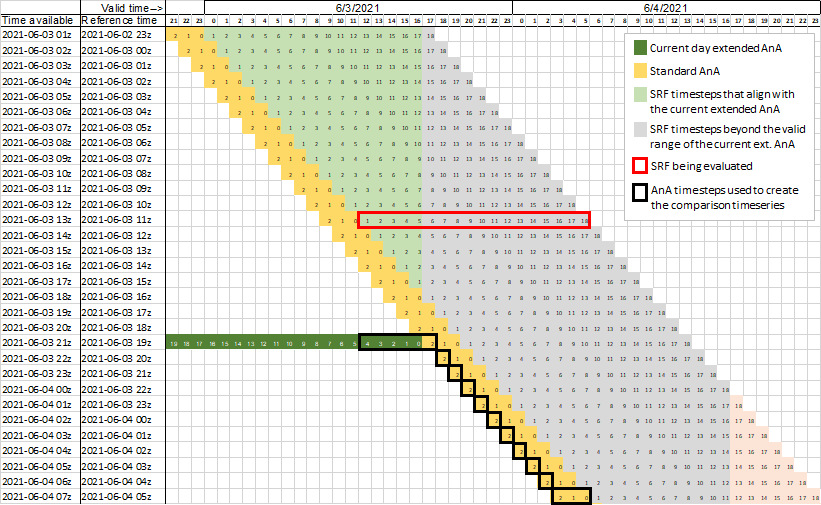


Figure C-9. 11z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 07z the next day)

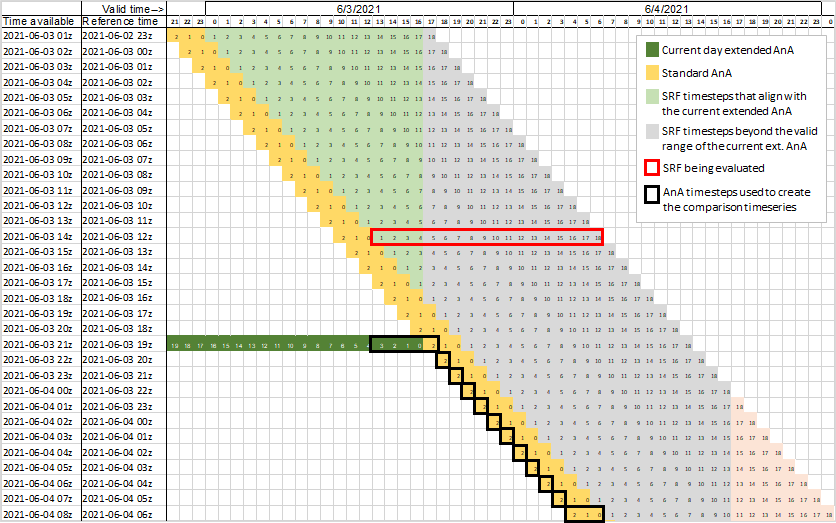


Figure C-10. 12z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 08z the next day)

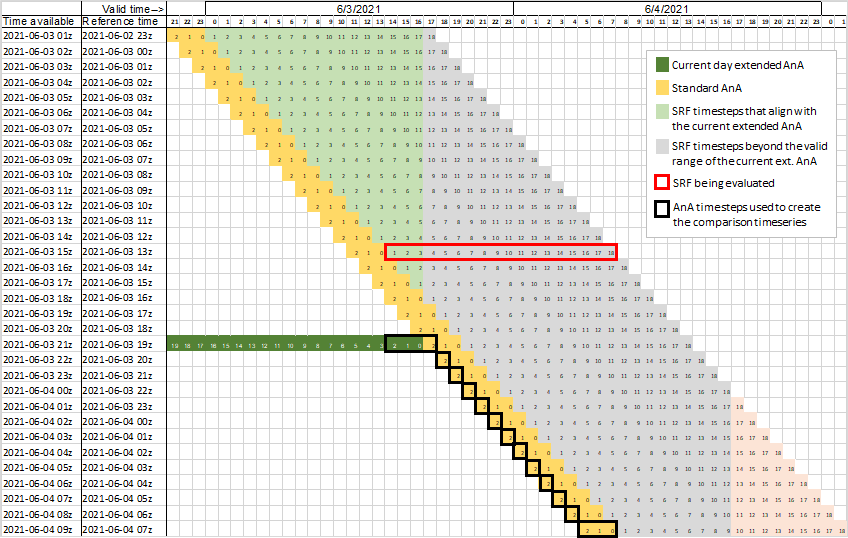


Figure C-11. 13z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 09z the next day)

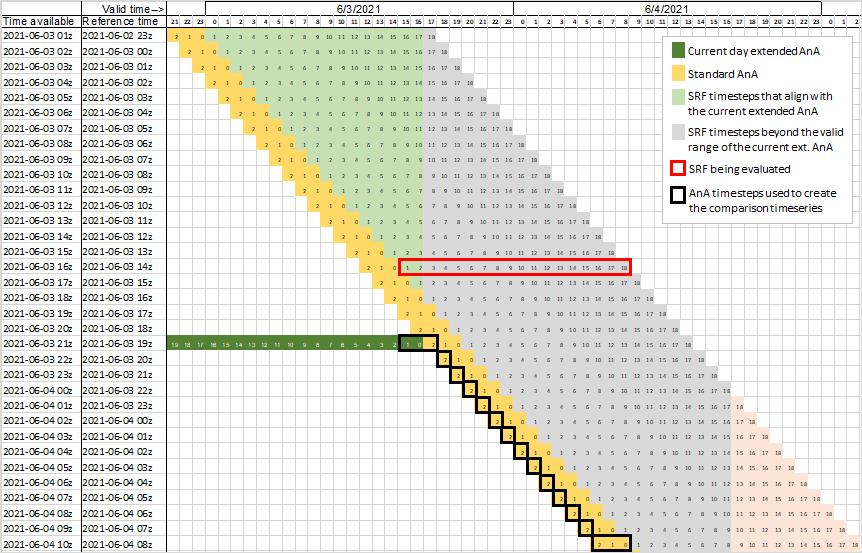


Figure C-12. 14z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 10z the next day)

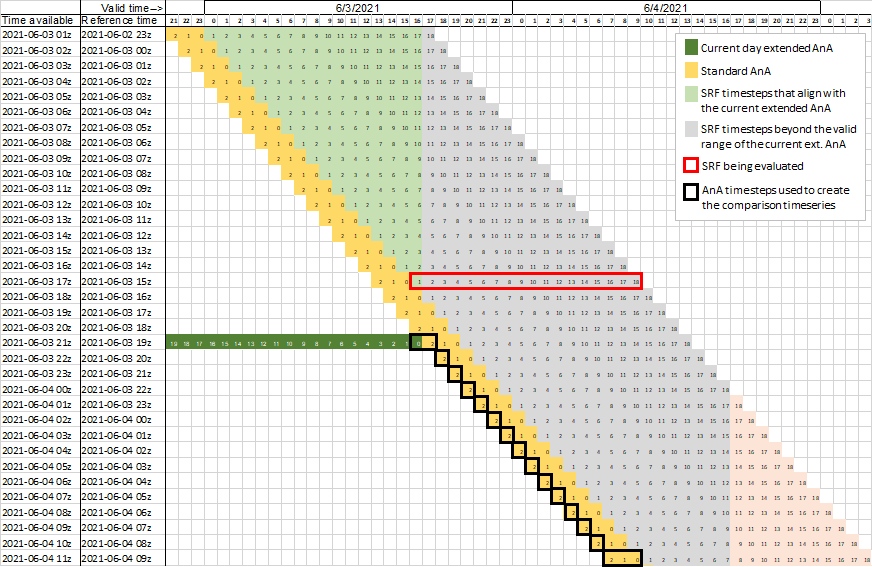


Figure C-14. 15z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 11z the next day)

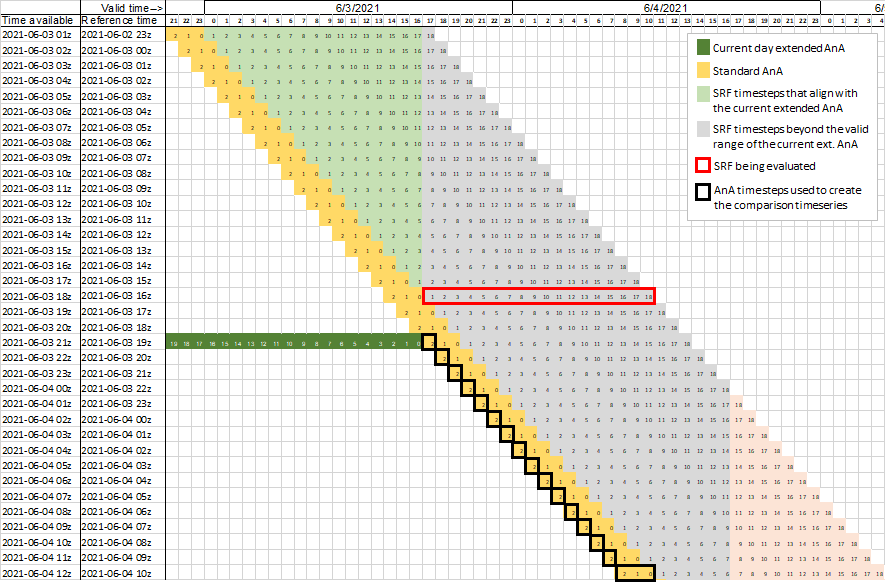


Figure C-15. 16z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 12z the next day)

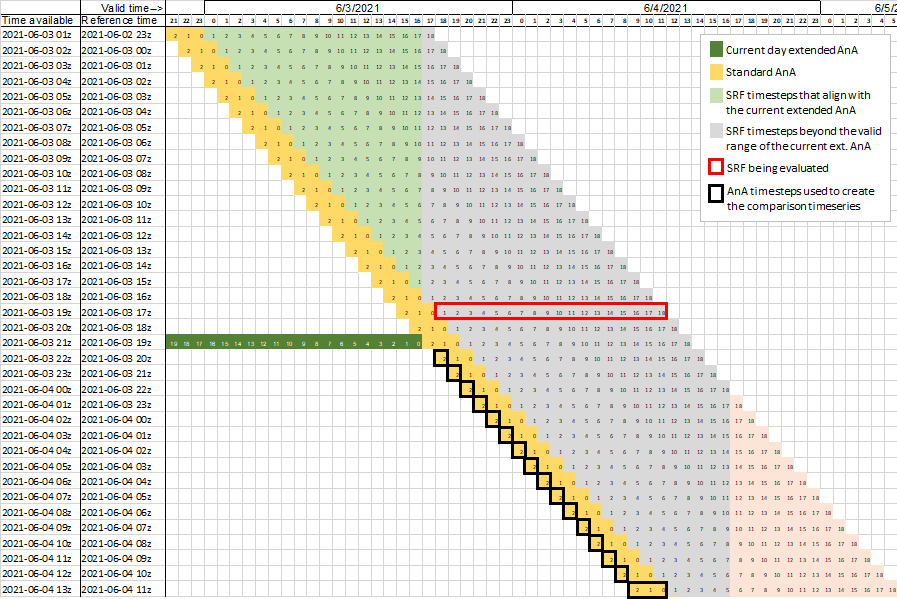


Figure C-16. 17z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 13z the next day, uses only standard AnA data)

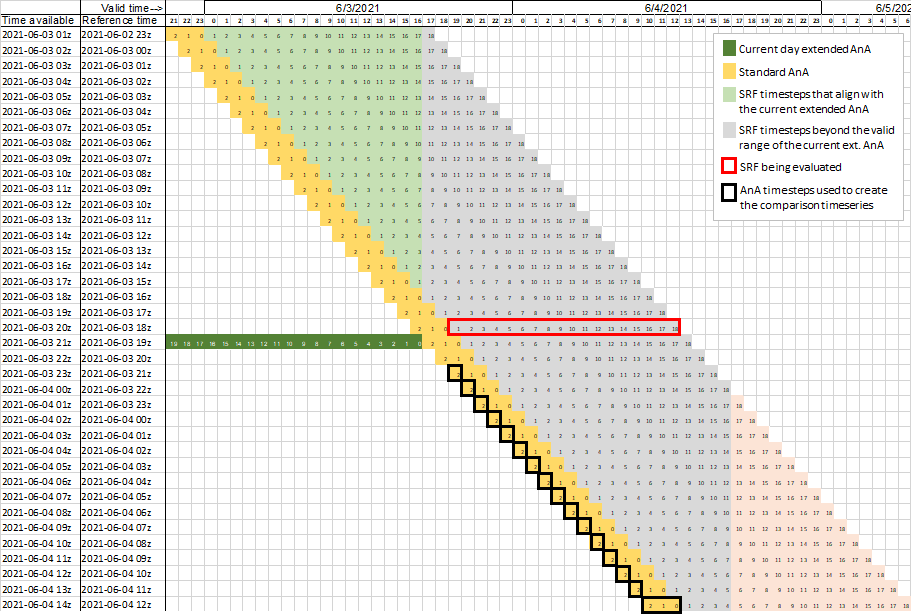


Figure C-17. 18z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 14z the next day, uses only standard AnA data)

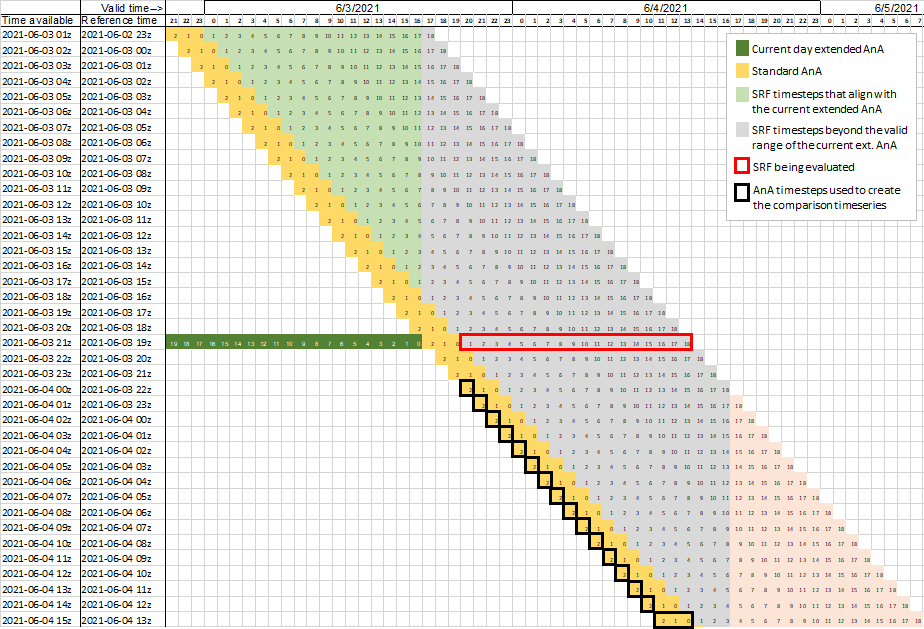


Figure C-18. 19z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 15z the next day, uses only standard AnA data)

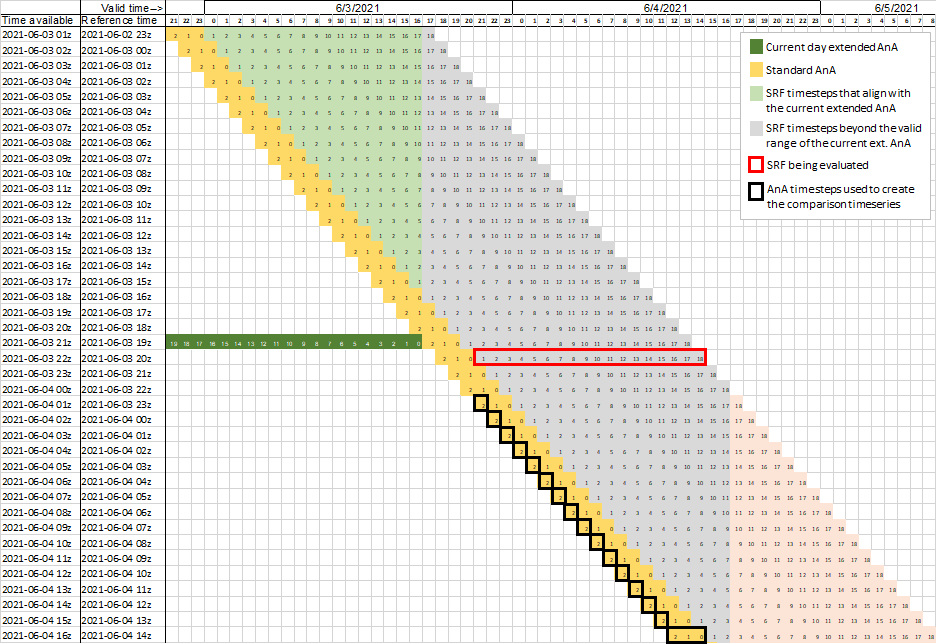


Figure C-19. 20z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 16z the next day, uses only standard AnA data)

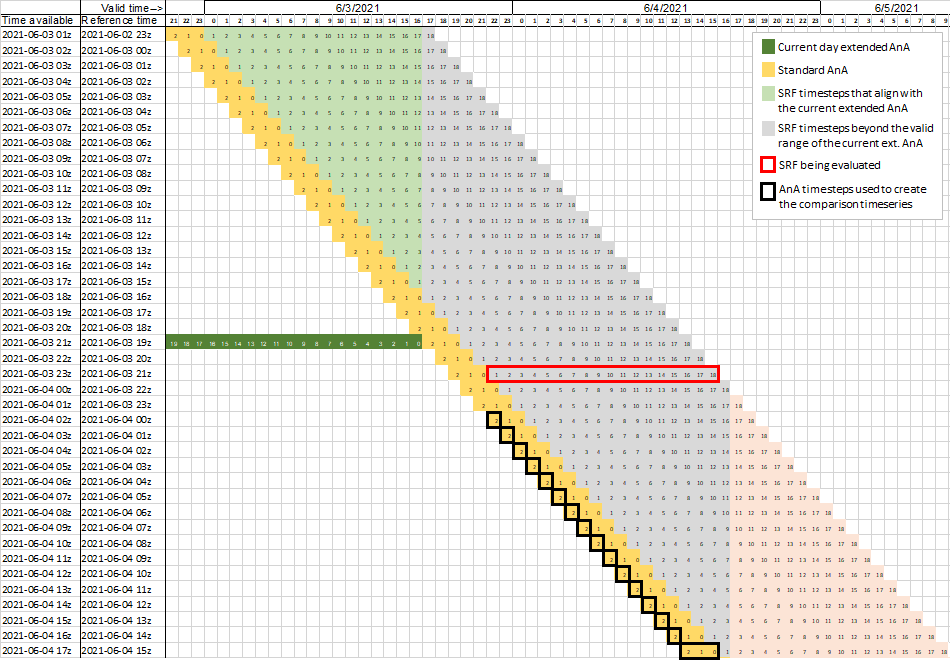


Figure C-20. 21z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 17z the next day, uses only standard AnA data)

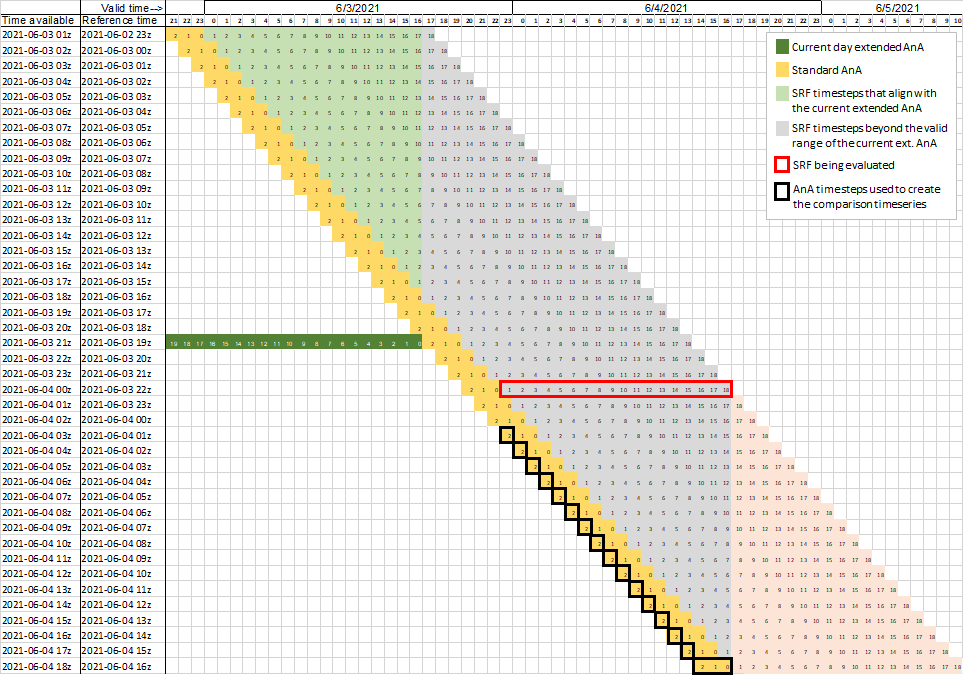


Figure C-21. 22z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 18z the next day, uses only standard AnA data)

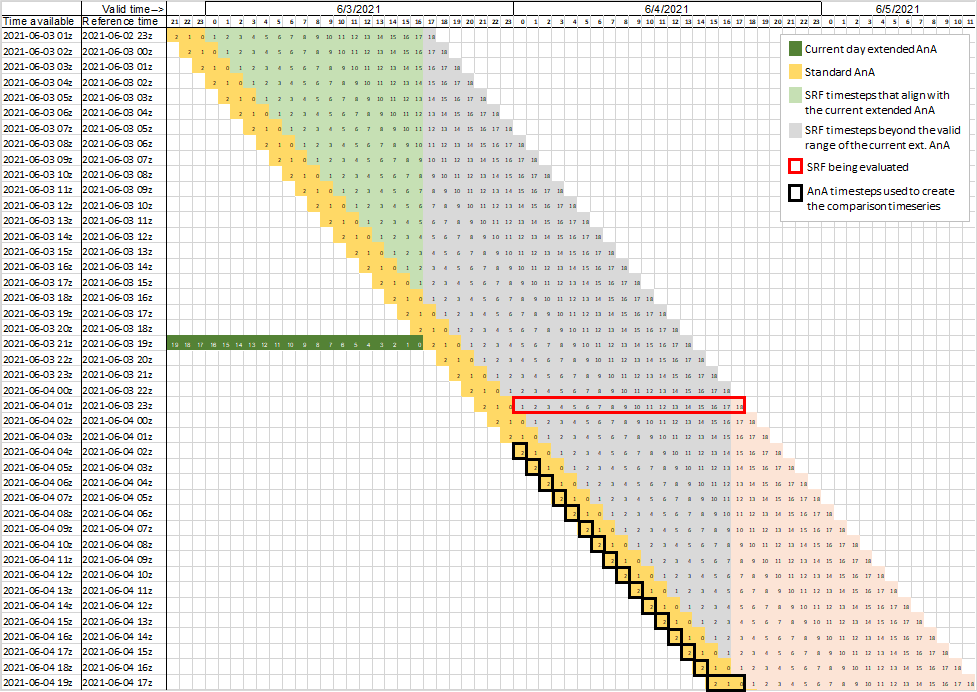


Figure C-22. 23z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 19z the next day, uses only standard AnA data)

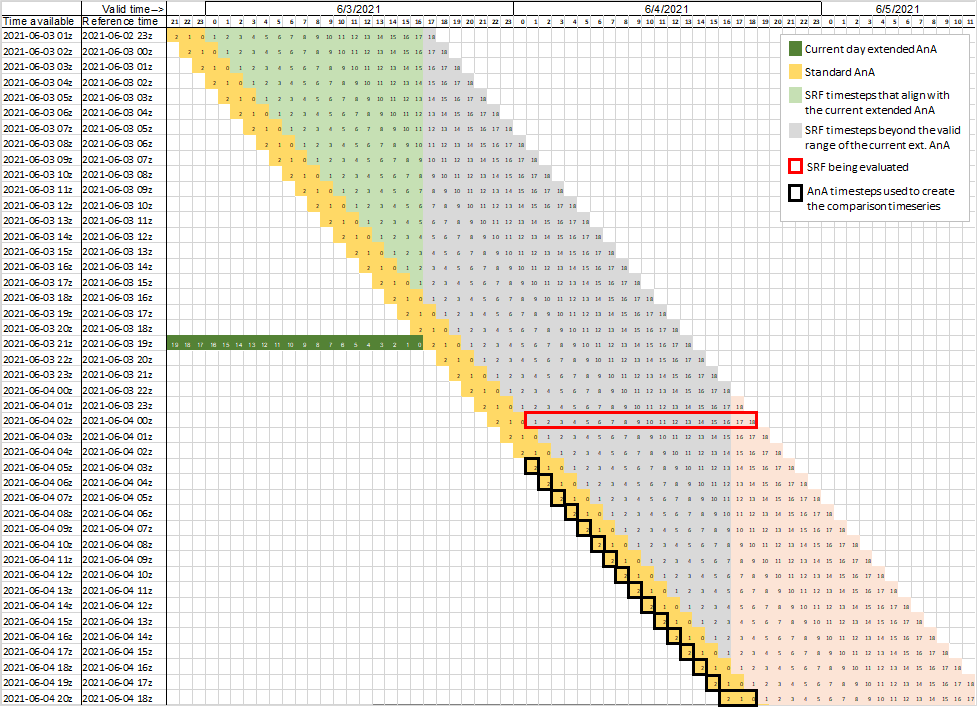


Figure C-23. 00z SRF reference time and stitching together of extended AnA and standard AnA simulations. (feasible to evaluate at 20z the same day, uses only standard AnA data)



Figure C-24. 01z SRF reference time (feasible to evaluate at 21z the same day, similar to Figure 7 above to clearly show the full cycle, uses the newly updated extended AnA)

## Appendix D

Use of Different Thresholds to Define ‘Yes Events’

The impact of this threshold choice can be seen in Figures D-1 and D-2. This example is taken from the February 2021 flood event in Nashville, TN. In Figure D-1, a HUC10 is classified as a ‘yes’ when its ROF % is greater than or equal to 20%. The top row displays the HUC10 ROF percentages derived from the SRF and the ExtAnA. The bottom row displays the extent of the ROI and the HUC10s classified as ‘yes’ events (ROF ≥ 20%) in black and the ‘no’ HUC10s as grey. The lower right panel shows the agreement between the SRF and extended AnA yes events. A ‘hit’ is defined when a HUC10 is a ‘yes’ event in both the SRF and ExtAnA. A ‘miss’ is when a HUC10 is observed to be a ‘yes’ event but the SRF does not predict a ‘yes.’ A false alarm occurs then the ExtAnA predicts a HUC10 to be a yes event, but it is not observed. The resulting Probability of Detection (POD) is 0.2886, the False Alarm Rate (FAR) is 0.6802, and the Probability of False Detection (POFD) is 0.0993. Figure D-2 depicts the use of a 50% threshold to define a HUC10 as a ‘yes’ event. The POD, FAR, and POFD are 0.1416, 0.8699, and 0.0647, respectively. Comparison of Figures D-1 and D-2 shows that the 50% threshold reduces the number of yes events and alters the number of hits, misses, false alarms, and true negatives. Note: the metric POFD was used early in the development of the ROF evaluation strategy to compute the Relative Operating Characteristic (ROC) curve. POFD was replaced by the Critical Success Index (CSI).

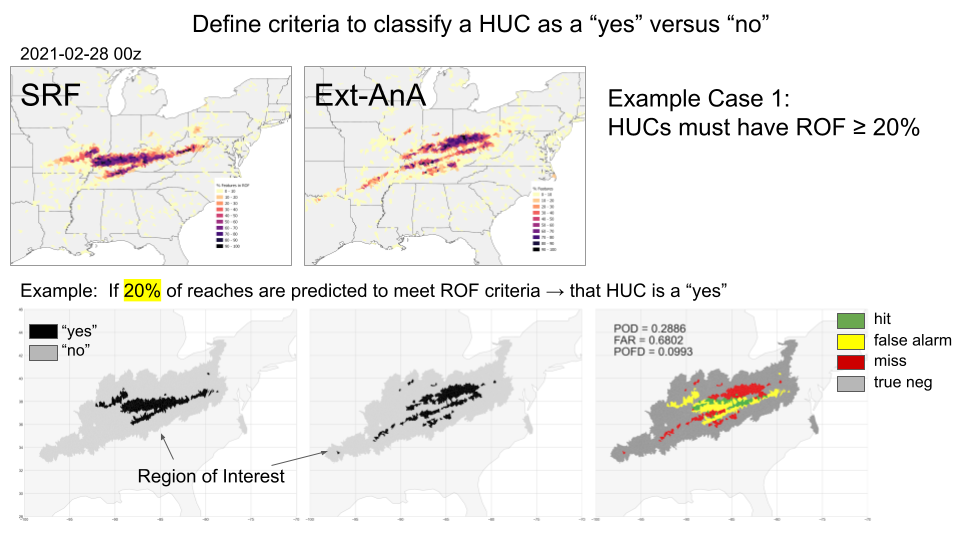


Figure D-1. Analysis of predicted and observed ROF using a threshold of 20% ROF to define a yes event.

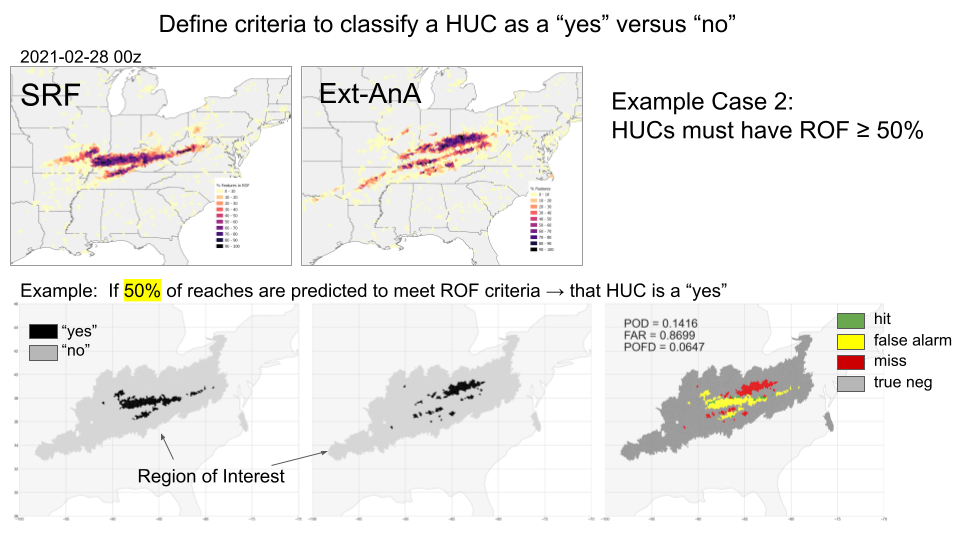


Figure D-2. Analysis of predicted and observed ROF using a threshold of 50% ROF to define a yes event.