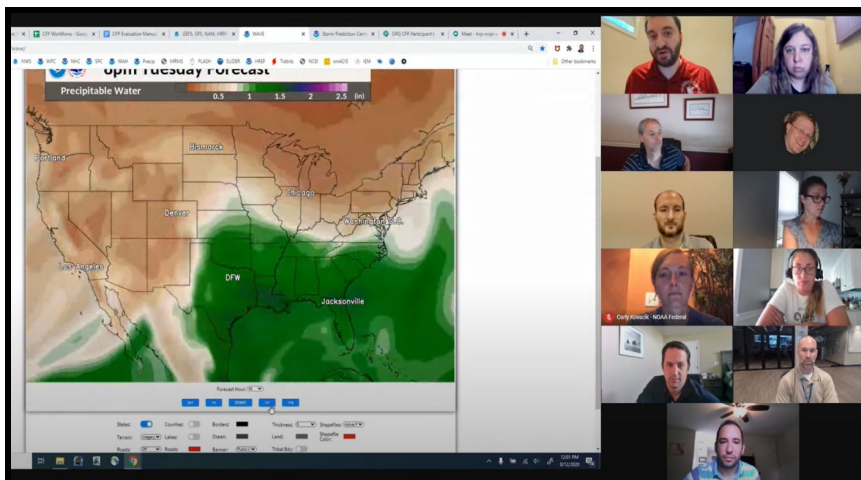


NWS Operations Proving Ground

Operational Evaluation Report

The Collaborative Forecast Process for QPF

An OPG-PMO Virtual Experiment



NWS Operations Proving Ground
7220 NW 101st Terrace
Kansas City, MO 64153

<u>1. EXECUTIVE SUMMARY</u>	<u>4</u>
<u>2. STRUCTURE OF REPORT</u>	<u>5</u>
<u>3. BACKGROUND</u>	<u>6</u>
<u>4. EXERCISE PREREQUISITE TRAINING</u>	<u>8</u>
<u>5. EVALUATION METHODS</u>	<u>10</u>
<u>6. OBJECTIVE 1: COLLABORATION TOOLS AND STRUCTURE</u>	<u>11</u>
<u>6.1 ANALYSIS OF CHAT IN COLLABORATION</u>	<u>11</u>
<u>6.2 ANALYSIS OF VIDEO CALLS IN COLLABORATION</u>	<u>15</u>
<u>6.3 ON USING THE PHONE IN COLLABORATION</u>	<u>17</u>
<u>6.4 IMPACT OF GIS TOOLS ON COLLABORATION</u>	<u>17</u>
<u>6.5 EFFICACY OF THE VARIOUS COLLABORATION TOOLS</u>	<u>21</u>
<u>7. OBJECTIVE 2: THE HUMAN FACTORS OF COLLABORATION</u>	<u>23</u>
<u>7.1 COMMUNICATION CHALLENGES AND SOLUTIONS</u>	<u>23</u>
<u>7.2 EVALUATE COMMUNICATION PRINCIPLES FROM CREW RESOURCE MANAGEMENT CONCEPT</u>	<u>25</u>
<u>8. OBJECTIVE 3: PROBABILISTIC DATA IN THE FORECAST PROCESS</u>	<u>26</u>
<u>8.1 DOES PROBABILISTIC DATA IMPROVE THE FORECAST PROCESS</u>	<u>27</u>
<u>8.2 DETERMINE FORECASTER PREFERENCES FOR QPF ANALYSIS</u>	<u>29</u>
<u>9. COLLABORATIVE FORECAST PROCESS WORKFLOW EVALUATION</u>	<u>29</u>
<u>9.1 THE COLLABORATION FLOW (TALK FIRST, EDIT SECOND)</u>	<u>29</u>
<u>9.2 THE VALUE OF A COMMON STARTING POINT</u>	<u>31</u>
<u>9.3 THE CONTENT OF COLLABORATION</u>	<u>32</u>
<u>9.4 The Participants on Collaboration Calls</u>	<u>33</u>
<u>10. NATURE OF FORECASTING - INFLUENCERS TO THE CFP</u>	<u>34</u>
<u>10.1 FUTURE CONSIDERATIONS FOR FORECAST METHODS</u>	<u>34</u>
<u>10.2 FUTURE CONSIDERATIONS FOR NDFD</u>	<u>35</u>
<u>11. TRAINING NEEDS</u>	<u>36</u>
<u>12. SUMMARY FINDINGS AND RECOMMENDATIONS</u>	<u>38</u>

<u>13. NOTE OF APPRECIATION</u>	<u>41</u>
<u>APPENDIX 1 - CHAT ANALYSIS CATEGORY NAMES AND DESCRIPTIONS</u>	<u>42</u>
<u>APPENDIX 2 - EXAMPLE ANALYSIS OF CHAT STATEMENTS</u>	<u>43</u>
<u>APPENDIX 3 - CFP EXERCISE DEVELOPERS</u>	<u>44</u>
<u>APPENDIX 4 - CFP EXPERIMENT AGENDA</u>	<u>45</u>
<u>APPENDIX 5 - END-OF-WEEK ANONYMOUS PARTICIPANT SURVEY QUESTIONS</u>	<u>46</u>
<u>APPENDIX 6 - END-OF-WEEK ANONYMOUS OBSERVER SURVEY QUESTIONS</u>	<u>47</u>
<u>APPENDIX 7 - END-OF-WEEK FINAL DEBRIEF SESSION QUESTIONS</u>	<u>48</u>

1. EXECUTIVE SUMMARY

Main Findings:

- The National Weather Service (NWS) can enter into a National Collaborative Forecast Process (CFP) Demonstration for Quantitative Precipitation Forecasts (QPF) with the collaboration tools at our disposal today without significant risks to operations.
 - Having all collaborating parties on the Advanced Weather Interactive Processing System (AWIPS) will maximize the effectiveness and quality of the collaboration process.
- There are some relatively simple challenges that can be resolved today with minimal training, and without significant cost, or developing new tools
- There are unanswered questions that became apparent during the CFP exercise that require further evaluation, but do NOT prevent the CFP demonstration from moving forward

The Operations Proving Ground (OPG) and Program Management Office (PMO) co-developed and hosted a CFP exercise in August 2020. As a result of this exercise, the OPG feels strongly that the statements above can be considered “validated” and “true.” This report provides reasoning and justification for each statement above and presents findings and recommendations focused on the current planned demonstration of a CFP for QPF, and future CFP initiatives.

Originally planned as an in-person exercise at the OPG in April 2020, the virtual CFP exercise took place in August 2020 after a delay caused by the COVID-19 pandemic. Participants representing four Weather Forecast Offices (WFOs) from the four CONUS Regions, one River Forecast Center (RFC), two Regional Operations Centers (ROCs), one regional headquarters, the Weather Prediction Center (WPC), and the National Water Center (NWC) joined each other from across the Continental US for a three-day heavy rainfall simulation. Most of the participants joined each day from their own homes, in four different time zones, leveraging the Google suite of products (Calendar, Chat, Meet) as their primary mode of interaction. The OPG provided the participants with archived data from a multi-day heavy rainfall event that impacted Oklahoma, Texas, Arkansas, and Missouri in early 2020 through a Virtual Lab (VLAB) page designed specifically for this exercise.

The OPG and members of the PMO’s CFP team co-developed the exercise based on the Executive Council’s approved CFP process for the eventual national demonstration. Due to the virtual nature of the exercise, the OPG and PMO decided to focus on the human aspect of collaboration and the flow of collaboration within the CFP plan more so than any technical conditions. For example, participants did not utilize AWIPS, nor did they create, transmit, or otherwise modify QPF grids. However, the participants were exposed to AWIPS in the Cloud during each morning’s weather briefing - a rather exciting and potentially transformative moment for both the OPG and the NWS.

In focusing on the human factors of collaboration, the OPG and PMO designed a scenario that varied the collaboration methods, tools, and expectations each day of the exercise. By altering the conditions of collaboration each day, the OPG gathered evidence on the pros and cons of using text chat, phone calls, video calls, and geospatial software within the context of a complex rainfall event. Further, the OPG

requested the help of a team of social scientists who brought a wealth of expertise and experience in analyzing the human factors of the collaborative process.

The OPG recognizes that a single exercise with only ten participants cannot provide enough evidence to justify major changes to NWS operations. However, results from this exercise combined with results from ongoing or prior regional experiments, past OPG experiments, vetted social science research, and data gathered during NWS service assessments and after action reports, collectively justify the findings and recommendations presented in this report.

As such, the OPG feels the most important lesson learned from our experiment is that our agency CAN and WILL collaborate effectively if provided with clear roles, responsibilities, and expectations along with a well-designed structure, quality tools, and common goals. Most of these factors are within our grasp today which justify moving forward with a national CFP Demonstration.

However, it is the opinion of the OPG that the Common Operating Picture along with new collaboration tools and training will be required to fully maximize the operational collaborative experience - especially when we expand to additional program areas.

2. STRUCTURE OF REPORT

The structure and flow of this report will follow the thematic objectives detailed within the [Virtual CFP for QPF Experiment Plan](#). Within the plan for the experiment, the OPG and PMO outlined three broad themes that guided specific objectives. Sections 6, 7, and 8 of this report address the objectives defined in our experiment plan. As a reminder, the objectives are noted here:

- **Objective 1 - Collaboration Tools and Process**
 - Determine if text-based, phone-based, video-based, or graphic-based collaboration methods provide the most effective and efficient environment.
 - Determine how geospatial collaboration tools impact the flow and focus of the CFP.
 - Evaluate the time required to collaborate - did participants require more/less time than allotted. Did they have enough time to reach a decision?

- **Objective 2 - Human Factors**
 - Determine if communication challenges happened, what caused them, and how we might fix them.
 - Determine if there is value in applying principles from the Crew Resource Management approach as a structure for human to human collaboration and communication.

- **Objective 3 - Probabilistic Data**
 - Determine if existing operational or experimental ensemble-based probabilistic information supports a forecaster's decision-making process when deciding whether the National Blend of Models (NBM) starting point is representative, or if high-impact events are reasonably reflected in the starting point.
 - Determine if forecasters find value in describing desired changes to the NBM starting point relative to percentiles versus raw QPF.

After describing the results from each of the objectives above, this report will describe findings for the following seven CFP related topics:

- The flow of the CFP
- The value of a common starting point
- The content of collaboration
- The participants of collaboration
- The nature of forecasting
- The training needs for a CFP

Finally, the findings and recommendations for the experiment will be listed at the end of the report. The OPG attempted to capture broad findings and recommendations that require an understanding of the entire report for proper context.

3. BACKGROUND

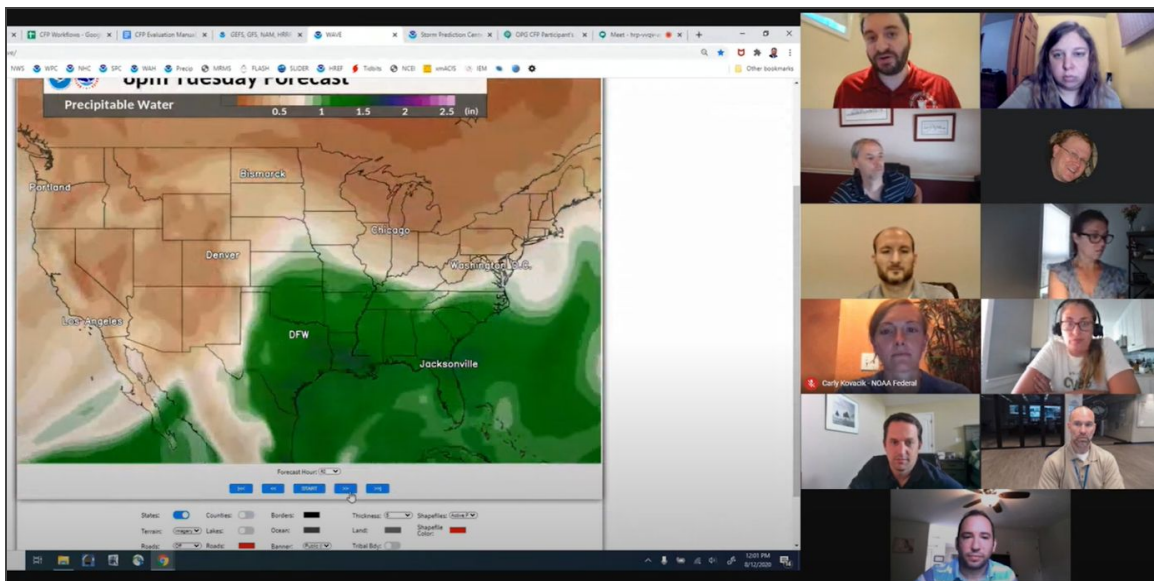


Figure 1: Forecasters and Hydrologists from across the CONUS participate in a collaboration video meeting where WPC discusses the large scale features contributing to an anticipated heavy rain event.

Changing the operational forecast process to become more collaborative in nature represents a significant turning point in the evolution of weather, water, and climate forecasting. Since the advent of weather prediction in the 19th century, forecasting has primarily been an individual effort. Certainly there are exceptions to the rule, but for the most part, an individual reviews data, applies that data to their understanding of conceptual models, and then converts their knowledge into usable information for the public and partners.

Collaboration in the NWS is most common within an office. Forecasters reviewing data simply turn to the person sitting near them and discuss a particular forecast challenge. Collaboration is even more

common within an office during significant weather events. These conversations are typically ad hoc, organic, and unstructured (opinion based on anecdotal evidence). Indeed, in some offices these conversations do not take place at all because of forecaster preferences, or even personality conflicts (observed and anecdotal evidence).

Since the introduction of chat rooms (within AWIPS and using external applications), conversations among multiple offices, including national centers, have become more common. During significant weather events, collaboration phone calls often take place and primarily focus on the potential impacts or headlines expected for the event.

To date however, the OPG is not aware of any formal studies focusing on the effectiveness or efficiency of NWS collaboration. Robust survey results from this OPG exercise, an informal poll conducted during the 2020 National Weather Association Annual Meeting (Brost, John, "The 2020 Collaborative Forecast Process Risk Reduction Experiment in the NWS", Sept. 2020, National Weather Association Annual Meeting), along with anecdotal evidence suggest NWS forecasters and hydrologists have mixed feelings about the effectiveness of NWS collaboration efforts.

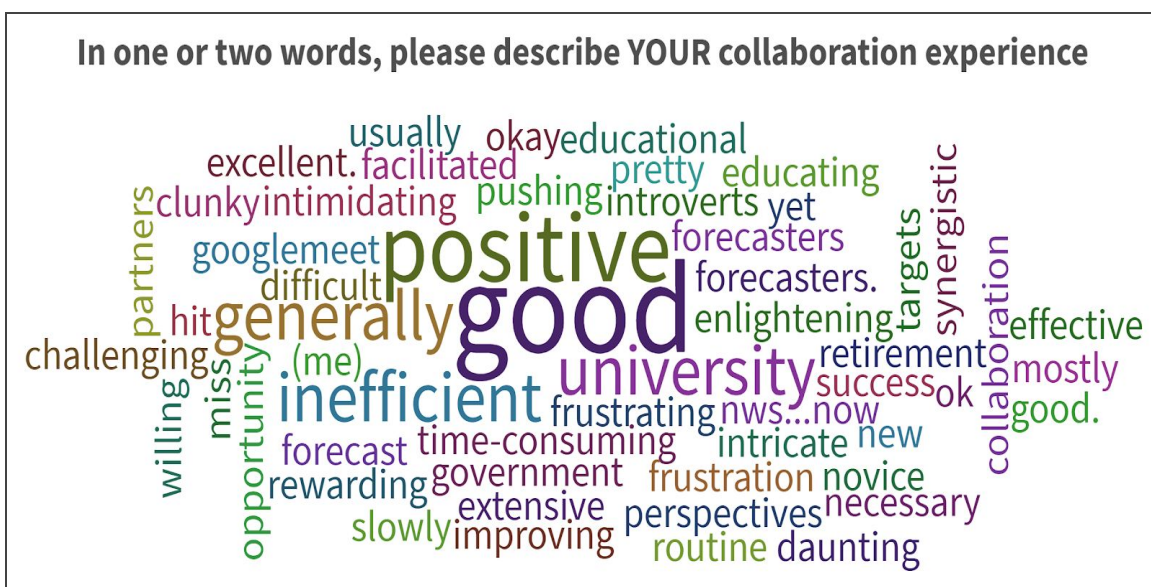


Figure 2: Responses to a poll question asked during the OPG’s NWA Presentation in September, 2020. Of the 38 relevant responses, 14 were positive words (e.g. Good, Improving, Effective), 13 were neutral words (e.g. Ok, Educational, Necessary), and 11 words were negative (e.g. Daunting, Inefficient, Frustrating).

Central Region Headquarters produced [a comprehensive report](#) following the conclusion of the first phase of the Central Region (CR) CFP with the Weather Prediction Center (WPC) conducted in 2019/20.

The report did not specifically address the effectiveness of collaboration, but did gather feedback from forecasters on ease of collaboration. Forecasters were asked: 1) “What is your perceived impact of using WPC QPF on collaboration specifically with WPC?” and 2) “What is your perceived impact of

using WPC QPF on collaboration specifically between WFOs?” Results showed that overall, the CFP demonstration had a positive impact on WFO collaboration with WPC and between WFOs. However, “while 67.8% of respondents said collaboration between WFOs was ‘easier’ or ‘much easier’ with the CFP demonstration, only 53.3% of CR meteorologists said collaboration with WPC was ‘easier’ or ‘much easier.’ This difference likely reflects on the basic overall benefit of a common starting point for QPF for WFOs (independent of the role WPC plays), and also the fact that collaboration guidelines with WPC have not been fully developed or implemented. Expectations and protocols to support effective collaboration with WPC need to be addressed prior to the start of Phase 2 of the CR CFP demonstration, and/or the planned national CFP demonstration.”

Transitioning to a collaborative forecast process is thus filled with questions and challenges. Since early 2017, the PMO has worked to address many of those questions and challenges through various teams composed of a broad spectrum of NWS employees (including the National Weather Service Employees Organization and other field representatives). These teams have produced reports for senior leadership describing various aspects of a CFP including (but not limited to): What should be used as a common starting point and why? When should the process begin and end? How does the process vary depending on event type and severity? What are the possible workload changes within a CFP? How will we evaluate success or failure and what should we do if the CFP poses major challenges in operations?

In 2019, the PMO provided enough evidence to Senior Leadership to justify a national demonstration of the CFP. One of the conditions for moving forward with a national demonstration is conducting risk reduction activities at the OPG. Our main function was to simulate the CFP and provide as much objective evidence as possible to help Senior Leadership determine if a national demonstration should proceed.

The OPG understands that one experiment with ten total participants can not possibly answer all of the questions surrounding a CFP. A single experiment can answer some questions though, and lead us to informed recommendations for additional tests or research. In addition, when the OPG exercise is put into context with other research, anecdotal evidence, and experienced participants, we can produce informed and reliable opinions regarding the CFP.

4. EXERCISE PREREQUISITE TRAINING

Prior to the exercise, the OPG and a few subject matter experts, conducted three Google Meet based video conversations, which served as an opportunity for participants to get an overview of the exercise, test technical requirements, and build familiarity and camaraderie with one another. In addition, there were two webinars focusing on concepts that the OPG and PMO felt were important for the participants to understand.

For the first webinar, the OPG invited Dr. Trevor Alcott (Earth Systems Research Laboratory and the Global Systems Laboratory) and Matt Jeglum (Western Region Headquarters Science and Services Division) to discuss the use of probabilistic guidance in the forecast process. These two experts focused on tools and techniques for identifying “Targets of Opportunity,” which were loosely defined as

anomalous events in which a single deterministic forecast was not representative of the real potential. Our goal was to prepare the participants to leverage probabilistic information during the exercise in their effort to identify both unrepresentative forecasts by the NBM and events which required enhanced messaging (for example, low probability and high impact events).

Eight out of ten participants found the Probabilistic Data prerequisite training to be at least "Moderately Valuable," and two deemed it "Extremely Valuable." One participant did not respond to the question. All ten participants **would** recommend the presentation be offered NWS-wide via the CLC or national webinar(s).

A week later, the OPG invited Tim Oram from the Operations Service Division at Southern Region Headquarters to provide an overview of a concept called "[Crew Resource Management](#)" (CRM). This overview on CRM would serve as a baseline for collaboration in advance of the exercise. Prior to Mr. Oram's presentation, the OPG asked all participants to review a document describing CRM from a fire industry perspective.

The OPG sought out Mr. Oram because of his unique background: having served in the United State Air Force in weather operations, with NASA's Spaceflight Meteorology Group during the shuttle missions as weather support, and now with NWS Southern Region Headquarters. Mr. Oram brought practical experience of using the CRM concepts in his prior positions and has completed extensive research on how the methods are utilized by various industries. In short, CRM teaches methods for communicating effectively in a team environment during high stress, time-sensitive, and uncertain events.

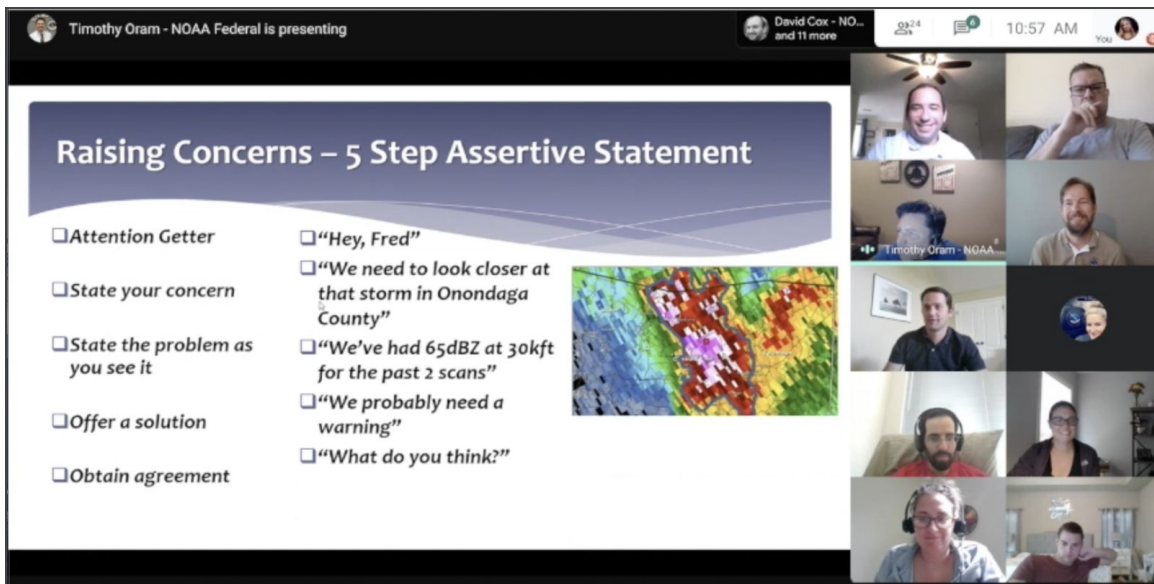


Figure 3: Screen capture from the Crew Resource Management presentation by Tim Oram describing the five steps of making an assertive statement.

Nine out of ten participants found the Crew Resource Management prerequisite training to be at least "Moderately Valuable," while three deemed it "Extremely Valuable." Nine of the ten participants **would**

recommend the presentation be offered NWS-wide via the CLC or national webinar(s). The one dissenting participant stated, *"The only reason I wouldn't recommend CRM for CLC is we need to make this more focused on NWS ops. There were obviously a lot of similarities, however."*

Finally, all ten participants felt the pre-work helped prepare them for the exercise, but they also indicated that the virtual nature of the exercise created challenges. One individual suffered from a poor network connection at times causing them to miss information in the pre-exercise webinars.

"[The OPG did a good job of preparing me for the experiment], and a lot of the material presented ahead of time I think really shaped the experiment afterward. I heard a lot more usage of things from the ESAT table and NBM percentiles. I did my best to incorporate ESAT stuff too in chat :) The CRM module talked about how important body language is for communication and collaboration, which is another reason why I feel Google Meet Video chat is so essential to the collaborative forecast process."
Participant #6

"I thought all prerequisite assignments and webinars were relevant and the quality was excellent. I...had very little idea about what the key takeaways were going to be from the experiment. This helped mitigate the risk of 'listening with a preconceived notion' (a category of receiver error identified in the CRM)." Participant #9

5. EVALUATION METHODS

Prior to the exercise, the OPG held a webinar with the group of observers (roughly 20 individuals from the PMO, the Office of the Chief Learning Officer, the National Center for Environmental Prediction, regional headquarters, and NWS Headquarters) to explain their roles and expectations. The OPG set up a Google Chat chat room where these individuals could interact with each other and the OPG team without disrupting the participants. The OPG then captured themes, questions, or concerns from this chat room and discussed these items during the daily debriefings with participants. We asked the observers to seek out certain behaviors during the exercise and share thoughts in the chat room in real time. For example, our observers documented use of probabilistic data, points of confusion, and effective collaboration techniques.

The OPG also invited a team of social scientists affiliated with the NWS Headquarters and the Office of Atmospheric Research (OAR) to provide feedback on the human factors involved in the process of collaboration. This team was given access to all of the various communication channels used by the participants (i.e. the chat rooms, Google Meet, etc.), so they could monitor the dialogue in real time.

Exercise participants were asked to complete a survey upon the conclusion of the experiment. The OPG used results from the survey to help evaluate exercise design, objectives, and results. Further, the social science team and observer groups notes were included in analyzing the participant behaviors.

6. OBJECTIVE 1: COLLABORATION TOOLS AND STRUCTURE

On each day of the exercise, participants had roughly an hour and 45 minutes per day to analyze data, collaborate via chat, and have a collaboration call if needed. The method of collaboration was varied each day to evaluate the efficacy of each type. On the first day, participants were allowed to use only text-based chat (as a proxy for AWIPS collaboration chat) and phone calls with the intent to simulate typical collaboration methods currently used in operations. On the second day, we added video conferencing and the option to share images in chat. On the third day, we added the capability to draw and share an area of proposed collaboration in an interactive, online Geographic Information System (GIS) platform.

6.1 ANALYSIS OF CHAT IN COLLABORATION

It is clear from the participant feedback that the use of chat for QPF collaboration in the NWS is important. 60% of participants felt that Google Chat was “moderately effective” for QPF collaboration (Image X). When asked to describe their “ideal collaboration experience,” 8 of 10 participants noted chat as a positive utility and Participant 5 explicitly stated *“that having the majority of collaboration through chat, and augmented by graphical tools similar to the AGOL one we tested, is the way to go.”* The fact that chat was so heavily utilized during the exercise provides further evidence of the essential role that chat fills in the collaborative process.

However, chat also had its disadvantages. Participant 9 summarized some of the pitfalls of chat:

“There is a point where chat yields diminishing returns in the CFP if (1) there are a lot of different conversations going on among a larger group (key information may become buried among the noise in a chatroom), (2) conflict arises (can be more easily resolved if it was handled by phone or video so intent can be better explained and tone, body language, or other visual/verbal cues can be expressed), and (3) tight deadlines are quickly approaching (reading and writing text in chat can become time consuming {at least based on personal experience} and could even distract one from completing a short-term task).”

In order to better understand the nature of collaboration using the chat interface, the OPG performed a thorough, sentence-by-sentence, analysis of each of the daily chat logs. We developed a list of 14 categories that attempted to describe the nature of each sentence. Each category and description can be found in Appendix 1 (also see Appendix 2 for an example of our analysis). Some of the sentences or phrases we analyzed required multiple classifications because their content could not be described by a single category or they included more than one unique thought.

The chat logs consistently ended up with the assignment of 220 to 230 classifications, or labels, each day. The logs produced anywhere from nine to 14 pages of content each day as well. The day with the most pages of content (Day 3) was the day that images were readily shared in chat; image sharing added to the number of pages due to the size of the images shared. Each text

chat session lasted about one and a half to two hours, but there was a conference call or video call during that time frame each day.

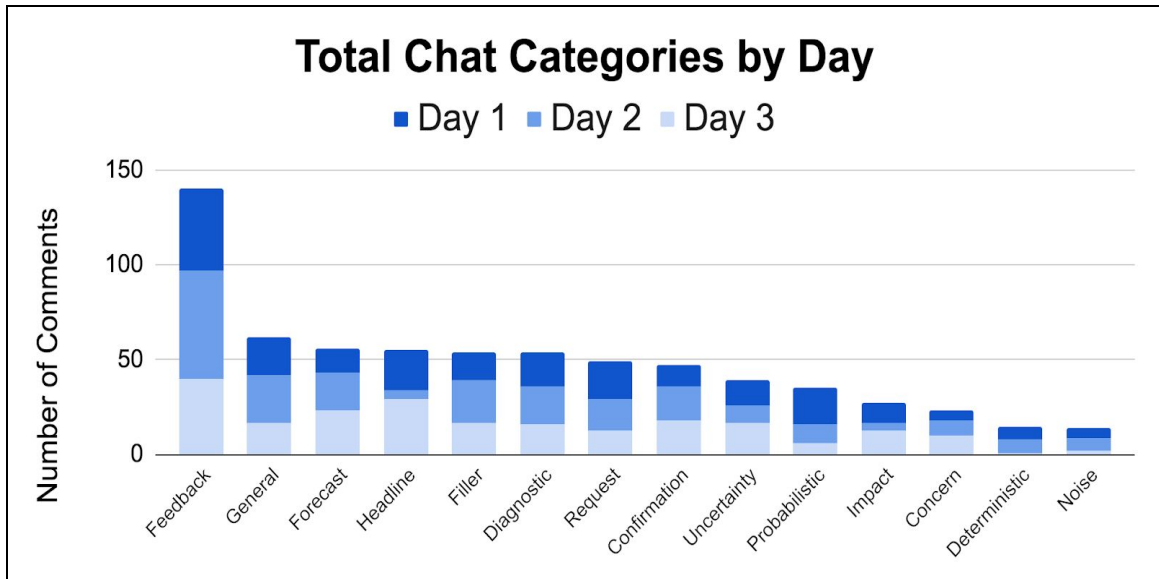


Figure 4: Chart showing the total number of statements with the applied label/category.

“Feedback” was the most common form of communication in the chat rooms, while “Noise” was the least common. “General Conversation” and “Filler” were both responsible for a little over 50 comments each. Most of the comments labeled as Filler and Feedback were short - on the order of 4 words or less. However, numerous statements that simply read, “I agree” or “I will look at that next” add up over the length of the chat log and can ultimately decrease the effectiveness of chat for collaboration. To that point, Participant 8 stated, *“Seems like the google chat room comments leading up to the call were just a running stream of consciousness instead of just a nice summary of thoughts/concerns. ... There was so much back and forth in chat it was hard to keep up.”*

Human Interaction	Analysis	Decision Making	Gut/Experience
General Conversation	Deterministic	Headlines	Uncertainty
Filler	Probabilistic	Forecast	Concern
Confirmation	Diagnostics	Feedback Request	
Noise	Impact/Hazard	Feedback Given	

Table 1: The 14 individual chat categories respectively grouped into one of four broad themes.

In order to describe the flow of the conversations in chat with respect to time, the OPG organized the 14 categories into four broad groups (Table 1). Then each individual label was

assigned to its broad theme, and then the running total for each of these themes was plotted against time for each day (Figures 5 - 7).

There are clearly some differences in the flow of conversations in each chat log. However, a few noticeable themes stand out.

- The first third of the conversation, especially on days one and three, is a combination of several different topics occurring at the same time. Human interaction appears to be the most discussed topic, but clearly the conversations are mixed.
- The middle third of the log shows a transition from a combination of topics, possibly focused on analysis, to decision making.
- The final third of each chat log is dominated by decision making.

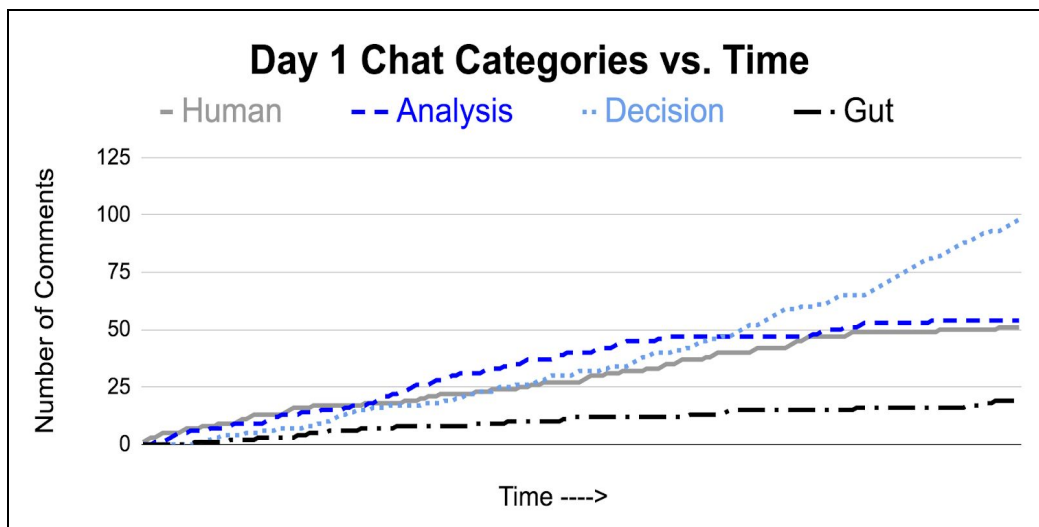


Figure 5: The flow of chat conversations over time on the first day of the exercise based on the four broad thematic categories.

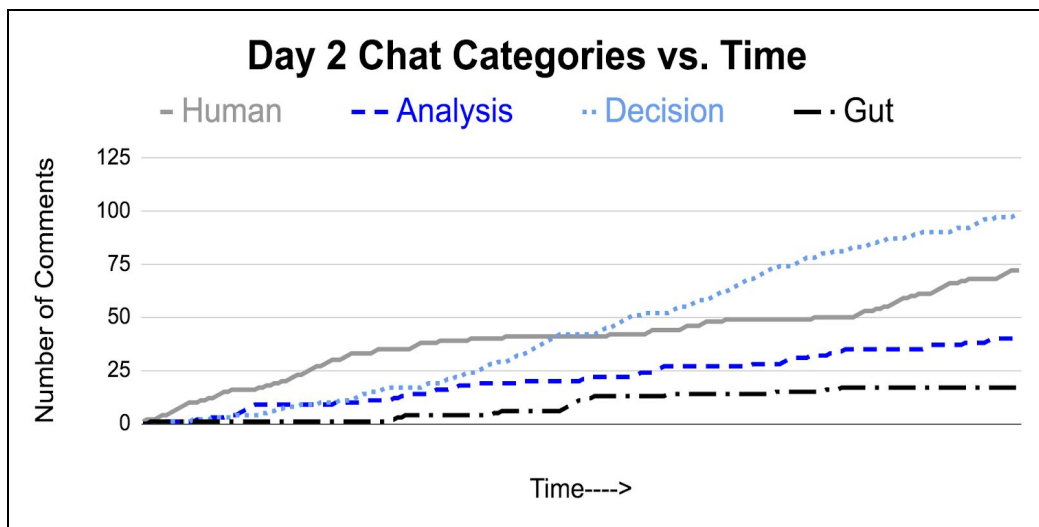


Figure 6: Same as Figure 5, but for Day 2.

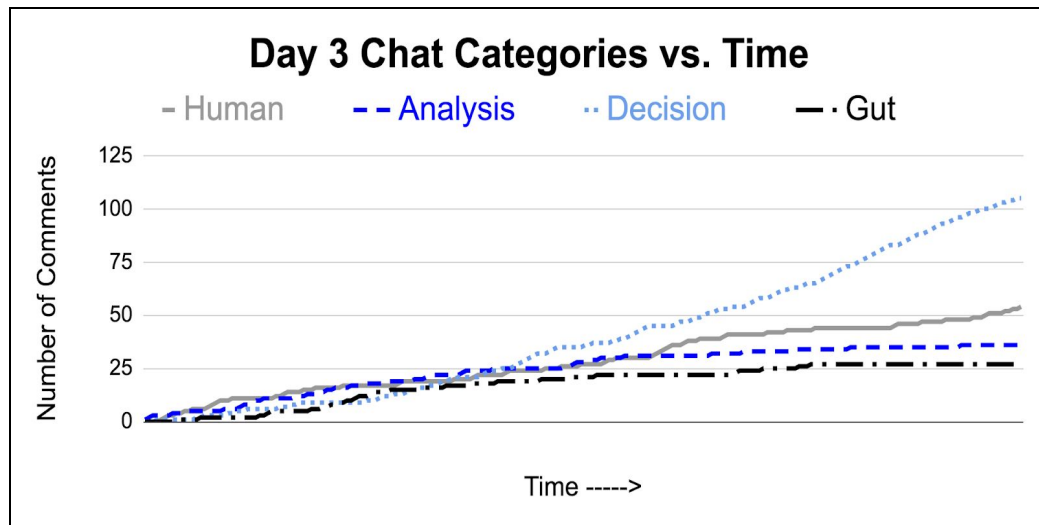


Figure 7: Same as Figure 5, but for Day 3.

Ultimately this analysis reveals that for roughly the first half of the conversation in chat, there are a variety of types of statements being made. At least qualitatively, the results here provide evidence to the chat conversation’s lack of focus. Moreover, the conversation might be described as having **“several different conversations going on among a larger group”** which Participant 9 noted as one of the situations where chat *“yields diminishing returns in the CFP,”* in this case due to information being lost in the broader chat conversation. During the Day 2 debriefing, one participant noted that **“[t]here’s not really a coherent discussion focused on one thing or the other”** and that the conversation could be *“a little confusing”* and that it *“does kind of get jumbled.”*

Based on the survey results, there is some desire to create a more streamlined approach to collaboration via chat. Participant 8 said, **“The process has GOT to be more streamlined/focused! The chats/calls in this exercise were way too long and unstructured ... and this only involved 1 weather event with 5 offices!”** Participant 5 noted the need for a better process and offered some solutions, *“It becomes a challenge when there is a high volume of collaboration and/or the collaboration is highly unfocused. To that end, we need tools that help streamline the process in an efficient way -- methods to quickly and graphically provide feedback to the national center, sorting conversations in a logical way, an understanding among all offices about what needs to be collaborated both in chat and via calls.”*

Participants offered a variety of solutions to improve the effectiveness of collaboration in chat.

- Breakout, targeted chat rooms: *“I’d find it advantageous to somehow be able to break up into a couple different chat rooms and to say, ‘...you and me are going to work on this problem, and my other forecast partner and this person’s gonna work on the other forecast problem.’ Just to try and maintain the threads a little more effectively.”* From Day 2 Debrief

- Ability to sort chat: “...you notice that the chat rooms get really full, and we're only talking QPF. So, you know, it'd be cool if a chat feature had tags or something like that, that you could actually...sort it.” From Day 2 Debrief
- Use of emojis in chat: “...emojis helped the communication flow in chat. They didn't clog up the chatroom with more text to read (Yes, I agree.). For example, you can quickly acknowledge that you read a chat meant for you by responding with an emoji. You can also give a thumbs up emoji if you agree with their statement.” Participant #9
- Polling and image sharing: “I think leaning heavily on chat is the way to go, but the chat should have the ability to integrate things like image uploads, polling, and other graphical tools to aid in communication.” Participant #5

Although not specifically revealed in the analysis of chat, one individual alluded to the difficulty of monitoring multiple chat rooms in the survey.

*“Right now we have AWIPS collaboration chat, which is largely a source for internal collaboration, but **there has been an explosion in the number of NWS Chat rooms.** Even a WFO is now expected to monitor multiple rooms for their own office, their respective ROC, special event rooms for things like tropical collaboration, etc. From a national level the number of NWS Chat rooms is enormous.”*

In this exercise, one chat room in Google Chat was established prior to the exercise, so while the issue described was not pertinent to the OPG exercise, it is worth considering with respect to the national implementation of the Collaborative Forecast Process. The same person added, *“...we need a single approach for the agency that has been thought through in detail. **We cannot have internal collaboration occurring on two different chat platforms and in many different rooms.**”*

6.2 ANALYSIS OF VIDEO CALLS IN COLLABORATION

Google Video was the only method (as compared to Google Chat and phone conferencing) to be considered “extremely effective” for QPF collaboration by any of the participants in the survey (30% considered Google Video Chat as “extremely effective”), and it was considered the most preferred collaboration medium by six of ten participants (See figures 11 and 12 in section 6.5 below). The ability to see body language and use of the screen sharing capability were commonly noted by the participants as attributes that made video calls an effective collaboration tool. For example, Participant 1 said video calls *“would clearly be the most effective way to communicate, taking advantage of both audio and visual language cues.”* Moreover, Participant 10 commented, *“**Video allows for the sharing of screens, allows for a better representative of body language and tone, and prevents awkward phone silences or people accidentally beginning to talk over one another.**”* In the end-of-week survey, 40% considered being able to see other people while collaborating “extremely valuable.” The remaining 60% considered it “moderately valuable.” Collectively, screen sharing during collaboration was considered even more valuable; 70% considered screen sharing “extremely valuable” while the remaining 30% considered it “moderately valuable.”

Participants also noted the value of video chat in specific scenarios such as when collaboration took place in a large group or when the impact, uncertainty, or complexity of a particular event warranted it. Participant 9 ***“preferred the video chat when making a collective decision among a large group”*** and *“when trying to communicate complex information.”* Participant 5 was particularly specific on the occasions which called for collaboration via a video call, stating they should *“be engaged when there is coordination to be done on any of three things: hazard products like watches and warnings; national-level outlooks like the ERO given their importance for messaging; and IDSS messaging for complex weather situations.”*

Despite the many positive uses of video chat for collaboration, this exercise revealed that the video calls are not without their challenges. Most notably, the calls were noted by several participants as being too long. However, this challenge is more broadly attributed to calls in general within the CFP as the concern also applied to the conference phone call on Day 1 of the exercise. The calls on Day 1 (by phone) and Day 3 (by video) lasted around 15 to 20 minutes; the call on Day 2 (by video) was a little over 12 minutes in length.

The length of the calls was a frequent topic in the debrief discussions. On Day 1, one individual stated the call *“went way too long, especially for the limited number of participants on the call.”* Another person said, *“The thing that stuck out to me is [the call] was way too long.”* On Day 3 there was a similar sentiment as participants said *“I think I went a little too long today”* and *“[the call] was way too long.”*

During the Day 1 debrief one person addressed the problem with the length of the calls when they said, ***“the longer these calls draw out, they just...take away from the actual time you get to spend in the forecast piece, and your services as well.”*** Participant 10 made a related comment in the survey, stating *“Perhaps during larger events involving a handful of WFO/entities, sometimes the main focus can drift off topic which can drag things on and interfere with time needed for other important duties (like DSS).”* This speaks to the broader concern of the time requirements needed to adopt the CFP across the agency.

“There is certainly an increase in the amount of communication, and thus time, required [for the collaborative process]. ...Overall, QPF collaboration, as we did in this exercise, should be rather quick (<20 minutes outside of data analysis/forecast time).” Participant #2

One participant addressed time concerns that a national center would have if multiple collaboration calls become more common.

“There is a limit to the number of phone or video calls you could have, particularly if they all involve several national offices. In other words, WPC, NWC, etc. would be participating (and helping to lead) all of the calls. So having more than a couple per shift could quickly become unsustainable.”

To add to the national center perspective, Alex Lamers, the WPC participant on Day 3 stated, *“...having multiple calls in a day for one area is a lot. And not to consider the fact that there might be a few other regions in the country that also would want to call if that's kind of*

standard practice. So, you almost need to set a pretty high bar for actually getting on a call.” The national center workload, particularly that of WPC, reasonably seems as though it could increase with a national implementation of the CFP based on results from the exercise.

6.3 ON USING THE PHONE IN COLLABORATION

The use of phone conferencing was limited only to a collaboration call on Day 1 during this exercise and was primarily included in order to be comparable to collaboration calls commonly used in operations today. Thus, although not the focus of this exercise on collaboration, there were some survey results that provide valuable insight. While phone conferencing was the least preferred collaboration method as compared to chat and video calls, four participants still noted positive attributes of a phone call when describing their ideal collaboration experience. Participant 1 noted that *“phone calls can be useful at times when a quick call is needed but a Google Video Chat is not accessible...”* and Participant 7 stated that *“the phone could be used for just a 2 WFO chat.”*

6.4 IMPACT OF GIS TOOLS ON COLLABORATION

Finally, on the third day participants were given the opportunity to draw a polygon over an area where they felt the NBM starting point was not representative using ArcGIS Online (AGOL). AGOL is a cloud-based Geographic Information System (GIS) platform that is part of the NWS ESRI ArcGIS license. The OPG leveraged AGOL and ArcGIS Survey123 as a tool to allow participants the opportunity to highlight locations and times which they felt needed QPF collaboration and provide feedback on their reasoning. To do this, the OPG created an AGOL “Dashboard” that included two side by side windows. The left side included a survey and the right presented the survey results on a shared map (Figure 8).

The survey was used as a device to streamline the data entry process. Participants were asked to state their simulated office name and then use an interactive map to draw a polygon representing an area they wanted to further discuss. Then they entered a date in the forecast cycle that they wanted to discuss based on their perception of the unrepresentativeness of the NBM, or the potential for impacts. Participants then had to provide an assessment of the NBM starting point by selecting one of the following options:

- The NBM QPF is MUCH too HIGH
- The NBM QPF is too HIGH
- The NBM QPF seems representative
- The NBM QPF is too LOW
- The NBM QPF is MUCH too LOW

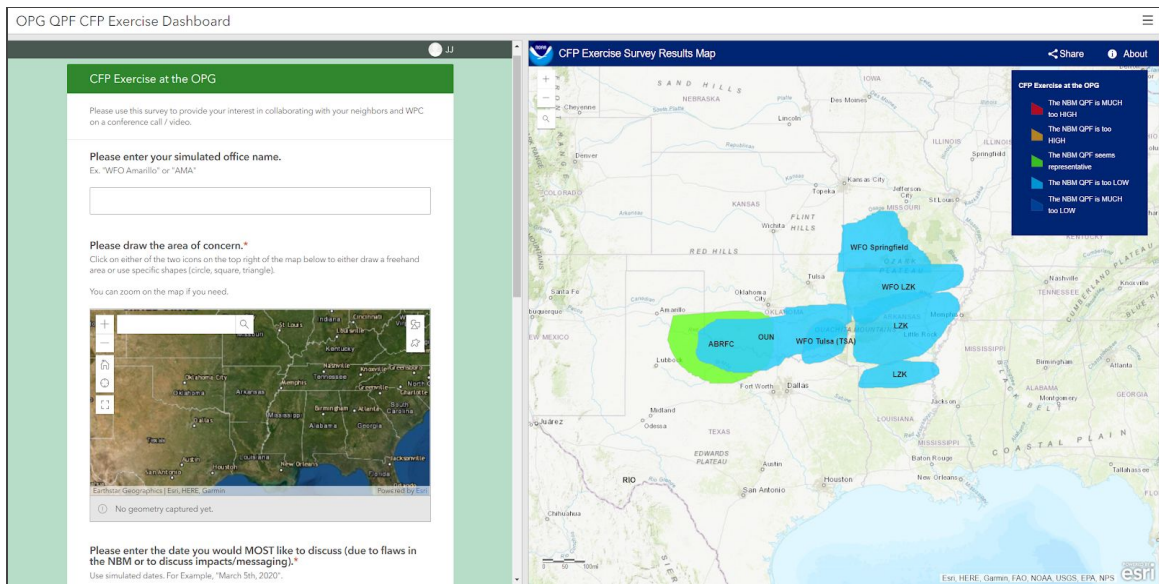


Figure 8: ArcGIS Online Dashboard with a survey on the left and results of the survey displayed on the right. The polygon colors on the right side map were based on the participants' interpretation of the representativeness of the NBM QPF starting point. The cyan/light blue colors correspond to, “The NBM QPF is Too Low,” while the green color corresponds to, “The NBM QPF seems representative.”

Finally, the participants had the option to add a brief narrative describing their reasoning on their evaluation of the NBM’s representativeness, or on the need to collaborate in general. They could also add another date of concern.

Once a participant's response was submitted, it appeared on the map on the right side of the dashboard and was viewable by all participants. The resulting polygons on the map were labeled by the office which they represented and were color coded based on the participant’s selection of the NBM’s representativeness. The reasoning associated with the responses could be accessed by clicking on each respective polygon (Figure 9).

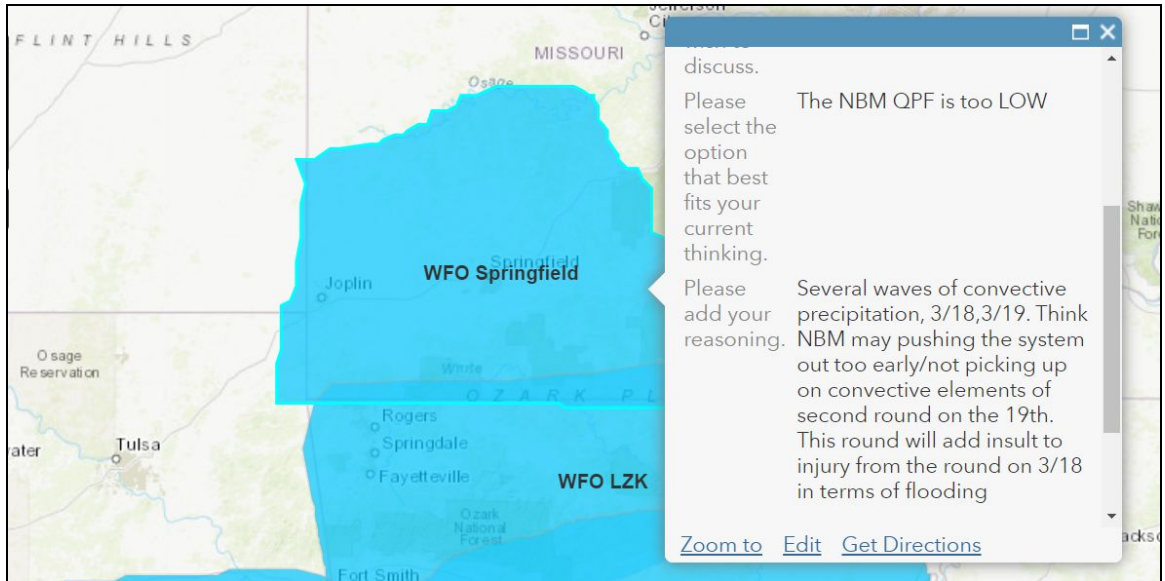


Figure 9: Survey123 response related to the polygon drawn by WFO Springfield during the exercise. Note the reasoning in the response not only addressed the concern over the QPF amount, but also indicated a concern for increased impacts due to additional rainfall.

In the survey, we asked participants about the value of conveying information geospatially using AGOL. Four out of the ten participants said using AGOL was “extremely valuable,” five participants said it was “moderately valuable,” and only one said it was “slightly valuable” (Figure 10).

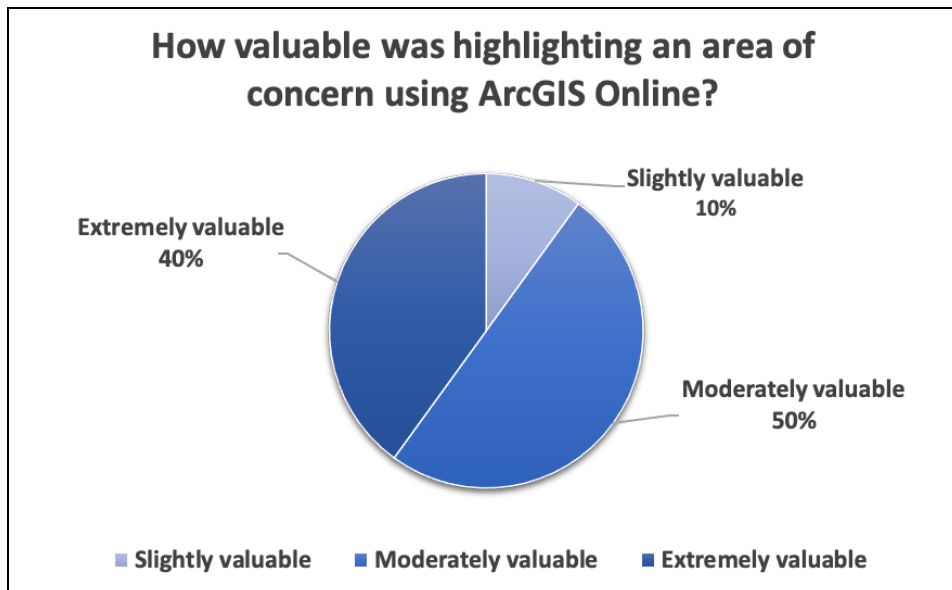


Figure 10: How valuable participants felt highlighting an area of concern using ArcGIS was during collaboration.

Interestingly, several participants brought up AGOL in the short answer questions even when the question was inherently biased against AGOL. We asked, "Now that you have used chat, phone calls, and video chat for collaboration, please describe what your ideal collaboration experience would be." This question fails to mention AGOL, and yet three of the participants brought up AGOL as a valuable tool for collaboration and one other alluded to an AGOL like platform:

"...something like Google Chat with an interactive viewer to specify potential targets of opportunity, such as the NBM is too low on QPF, ERO adjustments & areas of flash flooding concern watches, between offices would be useful." Participant #1

"I think having the majority of collaboration through chat, and augmented by graphical tools similar to the AGOL one we tested, is the way to go." Participant #5

Several individuals also discussed the positive attributes of AGOL in the Day 3 debriefing with a particular focus on the tool's ability to easily receive an overview of each individual's perspective in a spatial context. One participant said, *"it was good just to get kind of a first glance of what everybody's thinking drawn out on a map."* Another stated, *"...I thought it worked pretty well. ...it's an easy tool to be able to...signify where you see some of those discrepancies, or where you might be able to add value...target that area, and you're able to get a conglomerate of all the offices and...the river forecast center. The different areas where we need to put our primary focus on."*

Of particular note was the utility AGOL served to our WPC participant on this day, Alex Lamers. Given WPC's role in overseeing the entire nation, receiving concise feedback in a simple format seemed especially appealing. Even one of the social scientists observing the exercise noted in real time that *"AGOL may have helped focus the conversation."*

"I think the feedback map (AGOL) is something potentially nice to pair with just using chat for deterministic QPF collaboration. It was really useful for us to synthesize. From my perspective, I could easily look at the map and say, 'Well, everybody wants the QPF to be higher on day 2 so you have to be higher on day two.' And it was just graphically depicted and really easy for me to see." WPC from Day 3 Debrief

While there were a number of positive comments, there were comments of either concern in using the tool or suggestions for its improvement in the Day 3 debriefing.

"As we know, there's a limited amount of tools available to be able to kind of get an idea if you thought NBM was too low or too high. I'm just wondering if you have the same specific area and you had all the forecasters circle that area, and what would they say? I bet you'd get some people say too high, some people too low. ...It would be interesting, I think you get a lot of different opinions on it." Day 3 debriefing.

"I think it would be nice to maybe have tabs for individual days, because ABRFC provided feedback for day one, and that kind of overlapped with where Norman drew for day two. And

that's not necessarily an issue but if you do all seven days for a lot of offices that could become overwhelming." Day 3 debriefing.

6.5 EFFICACY OF THE VARIOUS COLLABORATION TOOLS

The end-of-week survey revealed that most (90%) of the participants felt that using Google Video Chat was either "moderately" or "extremely effective" (Figure 11). Moreover, video was ranked as the most preferred collaboration medium by over half of the participants (Figure 12). This is supported by the overwhelmingly positive feedback from offices using the video communications during recent tropical seasons (observed and anecdotal evidence). Meanwhile, chat was ranked by the majority of participants as the second most preferred collaboration medium, followed by the least preferred method of using the phone (Figure 12). 80% of participants considered using phone conferencing as only "slightly effective." No one stated that any one of the three methods considered in the survey was "not effective."

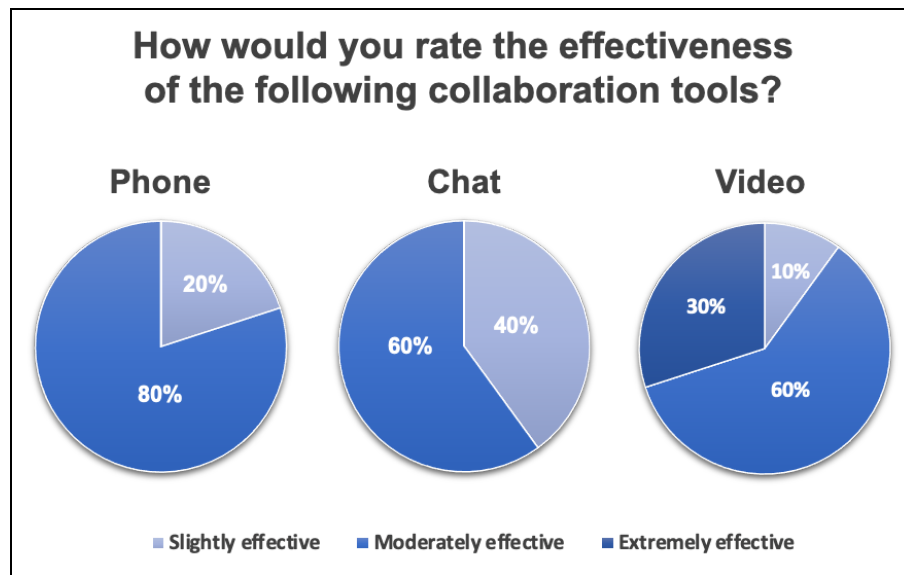


Figure 11: The participant's survey responses on the effectiveness of phone conferencing (left), Google Chat (middle), and Google Video Chat (right). Possible responses were "not effective," "slightly effective," "moderately effective," and "extremely effective."

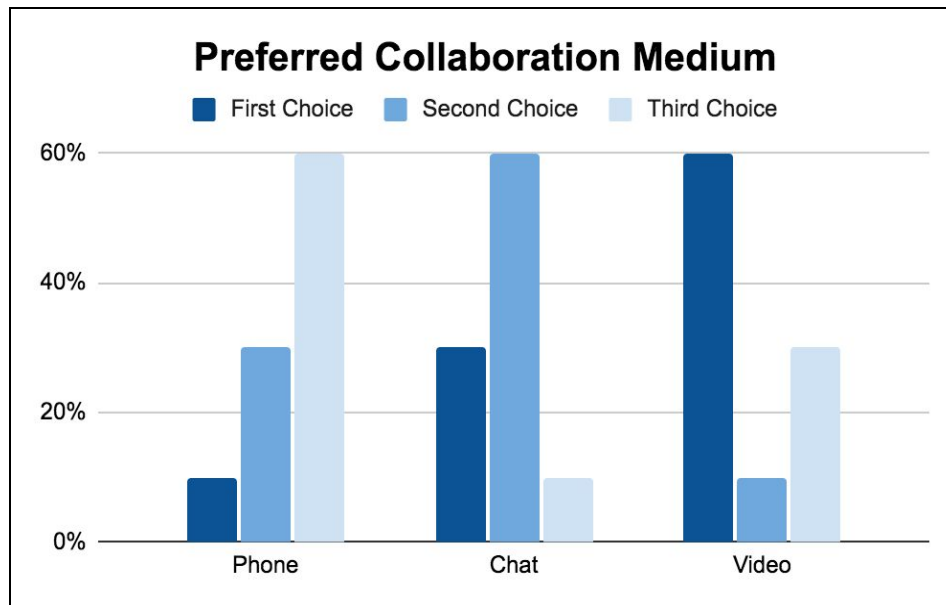


Figure 12: Participant responses to the statement: “Rank your preferred collaboration medium. With 1 being your first choice, 2 being your second choice, and 3 being your third choice.”

If the quantitative responses are viewed without context, one might consider the results fairly mundane. However, the short answer responses provided greater nuance and valuable insights regarding the conditions that suit each collaboration tool. For example, participant #9 stated, ***“There is no one-size fits all solution. I think [chat, phone calls, and video calls] are necessary depending on the situation.”*** Several other participants expressed the value of using multiple collaboration methods, many suggesting specific uses depending on the scenario. In fact, in the survey, nine out of ten participants specifically noted the utility of two or more collaboration methods being used in open response questions.

The most commonly described combination of collaboration methods, noted in the survey by seven of ten participants, was using at least a combination of chat and video (often mentioned alongside one or more other tools as well). Half of the participants outlined a collaborative approach where discussion first takes place in chat, and then, if needed, a video call would take place. Participant 2 stated, *“Doing most collaboration via chat seems reasonable, but then ironing out any necessary or outstanding issues via video calls should work most effectively in most instances.”* Participant 10 offered similar thoughts, *“My ideal collaboration experience would be to utilize a chat room to spark initial discussion regarding the forecast. From there, it would depend on the scale of the weather event being collaborated, but in most cases, then using video for discussion or some type of GIS platform (for visual collaboration) seems ideal if further collaboration appears necessary.”* Many participants also stated the initial chat should be focused primarily on the “meteorology or hydrology,” and the video call would be focused on making a decision on the points raised in chat and would ideally focus to some degree on hazards, impacts, and messaging (See section 9.4 for more information on the content suitable for specific collaboration tools).

Other commonly mentioned combinations of tools for ideal collaboration were the use of chat, phone, and video (40%) and the use of AGOL with chat (40%).

7. OBJECTIVE 2: THE HUMAN FACTORS OF COLLABORATION

It should be self-evident that there are a multitude of human factors that can have a profound influence on the collaboration process. The OPG, with the assistance of the social science team, attempted to capture some of these factors during the exercise. Many of these findings can be found in other sections, but a few noteworthy examples are discussed below.

There are several modules in the Commerce Learning Center that discuss communication, however, the OPG felt the need to introduce our participants to a concept that focuses on communication during high stress, time-sensitive, and uncertainty filled situations. As noted in Section 4, we introduced the participants to Crew Resource Management prior to the exercise. To be clear, our intent was simply to introduce important collaboration concepts to our participants because we were unsure if they had received adequate training prior to the exercise. We are not advocating that the agency invest in, nor adopt CRM as a standard practice, however, we do feel strongly that this concept could greatly benefit the collaborative process.

Our social science team was critical in helping the OPG assess the human factors of collaboration during the experiment. We focused on the potential for conflict, methods of resolving conflict, communication challenges, and of course best practices.

7.1 COMMUNICATION CHALLENGES AND SOLUTIONS

“With regard to the conference call, I would just have to say there’s simply too much going on. We were all over the place: short term, long term, back to short term. So we need to, even if it is a complex scenario we need to break this more up into sections like short term.” From the Day 1 Debrief.

One of the important communication lessons from the CRM materials involved “Sender Errors.” Sender errors can result in a misinterpretation of information and lead to poor decision making. Examples of sender errors include using vague or ambiguous words, failing to establish a common frame of reference, lack of body language, and disrespectful statements. Thankfully, we did not observe any disrespectful communication during the exercise, however, we did make note of vague statements in chat and our participants pointed out the importance of body language during the video calls.

The other important lesson involved “Receiver Errors.” Examples include poor preparation, listening with a preconceived notion, missing non-verbal cues, and failing to ask for clarification. Fortunately, our participants all came to the collaboration table prepared to discuss QPF and we did not notice any issues with preconceived notions. During the collaboration sessions, our observers did occasionally point out vague or ambiguous statements made by participants that never elicited a clarifying statement.

In short, we noticed our participants using some of the CRM techniques within the chat logs, although it is unclear if their behavior was directly a result from reading the CRM materials or not. We also noticed a few cases where the conversations may have become unclear or suffered from some of the errors noted above.

The OPG staff and the social science team paid close attention when the potential for conflict arose. During the final debriefing, and through the participant survey, the participants noted that very little conflict occurred during the exercise. In fact, nine out of the ten participants said they did not notice any conflict during the exercise. However, during the final debriefing, participants did share anecdotes regarding conflict during previous collaboration experiences in operations, such as:

*“I guess it depends on the situation but I would say in most cases **there might be some initial attempt to establish collaboration but again people seem to get defensive really quickly.** And then it just leads to [the feeling] that the [neighboring] office is always out to lunch or something like that. And so then that ends up being the end of collaboration because nobody wants to take the conversation any further because there's just complete disagreement.”* The comment above represents a pervasive sentiment within the agency. The quintessential pre-conceived notion (one of the CRM defined receiver errors) that the neighboring office or National Center is “out to lunch” or “is always wrong” or has some other inherent bias. The OPG believes that this sentiment was not observed during the exercise because our WFO participants played the role of different offices and they likely approached the exercise from a spirit of collaboration rather than ownership.

Another individual shared a comment that expressed similar concerns about collaboration but then offered CRM as a possible solution:

“There are times when you're sitting across the desk from somebody and you know [other offices] are [trying to interact] in chat collaboration. But there is no response from [your coworker], and you have to just say prod them into [responding]. It is one of the most delicate things we have to do. I think we talked about this in the CRM [webinar]. [We need an] attention getter. I can't remember the five steps exactly but it was like [clearly state that] we have a problem. Let's communicate about it. Let's work through that, but you kind of have to draw them into the conversation.”

In the case above, the participant noted that not everybody within an office seems interested in collaborating. This puts pressure on a shift leader to ensure their fellow forecasters remain engaged in the collaborative process. Expressing a difference of opinion diplomatically and in a non-threatening way can be accomplished using the Five-Step Assertive Statement method as outlined by CRM.

1. Opening/attention – Say the person’s name.
2. State concern/owned emotion – “I’m very uncomfortable with . . .”
3. State the problem as you see it – real or perceived.

4. Offer a solution – “I think we should . . .” (Major success key)
5. Obtain agreement – “What do you think?”

While the participants largely perceived an absence of conflict, the observers and OPG noticed differing opinions on the need for a conference call during the second day of the exercise, however, the disagreement never rose to the level of contention or serious debate. The OPG believes the potential for contentious debate was reduced by one participant, Alex Lamers, who used the above Five-Steps Assertive Statement to bring consensus. Alex noticed differing opinions during one of the days so he interjected two chat messages stating:

(At 10:24 CT) “Hi everyone, thanks for the thoughts. This is where it would be good to have a polling feature in the chat! WPC certainly doesn't want to take up too much of your time, but I think a quick call where we can all get on the same page would be good. I'm seeing a lot of conflicting opinions on the need for a call through chat and we haven't really discussed the actual changes yet from NBM that would be made for Day 3 and beyond. Just for clarity of communication, let's try for a 5-10 minute call beginning at 1535Z.”

(At 10:25 CT) “SR ROC and CR ROC, WPC can lead a meteorological discussion at the beginning with the WFOs. We can kick it to the NWC for a hydro overview (and they are welcome to discuss with RFCs as needed). And then I will kick it to you to wrap up with impacts and messaging concerns. Does that seem like a reasonable agenda? Let's try to keep it quick so you can all get back to the job at hand!”

Notice Alex started each chat by addressing the target audience (in this case “everyone” and the SR/CR ROCs). Then Alex expressed his concern, “*I'm seeing a lot of conflicting opinions... ,*” followed by his impression of the problem, “*We haven't really discussed the actual changes yet from the NBM... .*” He offered a solution, “*Let's try for a 5-10 minute call,*” and then asked for feedback, “*Does that seem like a reasonable agenda?*”

At the point in the chat flow, the responses quickly converged on Alex's well stated, clear, and effective synthesis of the conversation thus far. The conflict was resolved. The call that eventually took place was described as, “*better than the first day*” and the OPG believes this was due to the clear statements provided by Alex.

In summary, the OPG agrees with suggestions by our participants to help resolve some of the communication challenges. Most notably:

“There needs to be a clear set of guidelines and rules for these collaboration calls. And that's something the ROCs can work out with WPC. We could craft some rules for the call. Then, if everybody has these expectations going in, I think the calls would run a lot smoother. It would require some outreach on the ROCs part to the field offices and say, 'hey this is how it's gonna work, and please adhere to these set of rules and these set of guidelines.' And I think if those expectations are made clear on the front end, then these calls could run pretty smoothly.”

Day 2 Debrief

“We not only need a set of guidelines, but set goals for each call. For our call today, I didn’t know if there was a specific deliverable or outcome that needed to get resolved. So I was kinda trying to follow along to see, okay, what is coming out of this?” Day 2 Debrief

7.2 EVALUATE COMMUNICATION PRINCIPLES FROM CREW RESOURCE MANAGEMENT CONCEPT

Both the OPG and observers noted several cases in chat when participants used “ambiguous” or “vague” terminology. In other cases, participants made declarative statements without additional context. To be clear, our participants all performed exceptionally well during the exercise and the OPG understands the challenges associated with interacting with “strangers.” Regardless, the OPG believes that the chat experience during the experiment is representative of the chat experiences in operations.

The following examples of ambiguous or declarative statements were constructed by the OPG and represent the intent of chat posts but are not explicitly written in the chat logs. This was done to protect the participants.

“It depends on where the axis of heaviest rain sets up” - but the participant never expressed where they felt that axis may form, or the conditions that would favor one location over another.

“Heavier rain amounts are possible, so I’m fine with a headline if you want” - The word “possible” is a hedge term used excessively in the NWS as a replacement for specific probabilistic values. As such, it conveys no actionable information and making headline decisions based on a “possibility” is neither objective nor justified.

“This looks bad.” - If this statement is followed by “why it looks bad,” then it can effectively convey the severity of an event. However, in this case, the phrase stood alone with no further comment. Further, no other participant asked for clarification such as, “Why does this look bad?”

The social science team noted that the declarative sentences present in chat, like the ones expressed above, are still important for fostering relationships and building trust. The challenge involves achieving a balance between the declarative statements and statements of applied knowledge or seeking feedback from fellow collaborators. Too many declarative statements, or too many image shares without explanation simply muddle the chat experience. Further, too many formulaic statements designed around the Five-Step to Assertive Communication can appear robotic.

It may be a healthy exercise for offices to review chat logs from prior events. Count up the number of times participants simply presented information (“the water vapor values are high today”) versus the time participants assessed the data and conveyed their interpretation (“the water vapor values are high today and when combined with the approaching shortwave, I

anticipate robust and efficient rain producing storms - do you all agree with my assessment?"). Then simply discuss the impact various statements had on the flow or effectiveness of the collaboration session.

8. OBJECTIVE 3: PROBABILISTIC DATA IN THE FORECAST PROCESS

“There are some fundamental differences between how forecasters work with NWP model output in the short range and medium range....In the short range, meteorological knowhow is more important. In the medium range, statistical knowhow is more important”. [From the ECMWF Forecaster User Guide, Section 6: Using Deterministic and Probabilistic Forecasts.](#)

At the 2020 virtual NWA annual meeting, Andy Just from Central Region Headquarters shared results from his work on assessing the content of Area Forecast Discussions (AFDs). He noted that NWS forecasters strongly gravitate toward just three deterministic models in their forecast process. Furthermore, references to deterministic data were ten times more common than references to probabilistic data in NWS AFDs (Just, Andrew, *“Investigating Usage of Ensembles and Probabilistic Data in the NWS: An Analysis of Area Forecast Discussions”*, Sept. 2020, National Weather Association Annual Meeting).

As the OPG and PMO set out to design the CFP for QPF experiment, we discussed the types of data participants would need for their atmospheric analysis. Assessing the representativeness of a starting point (the NBM) is fundamentally different from creating a forecast from scratch. This is even more true when the starting point is a bias-corrected ensemble system using dynamic weighting based on past performance. Thus, the OPG felt it was necessary to evaluate whether or not ensemble based data improves forecasters' ability to determine the representativeness of the NBM.

8.1 DOES PROBABILISTIC DATA IMPROVE THE FORECAST PROCESS

Based on the CFP virtual experiment alone, we could not confirm the hypothesis. While we do have anecdotal evidence including survey feedback from our participants, the experiment could not explicitly prove the value of using probabilistic data in the forecast process. As such, the OPG is interested in designing future exercises that gather enough evidence to better assess the value of probabilistic data in the forecast process. Regardless, the experiment did teach us a few important points regarding the use of probabilistic data.

The social science team stated the following in their notes on the matter.

“While the participant chat mentioned probabilistic information, it was unclear how their decision making incorporated this information. Given that cognitive processing starts at an individual level, the chat does not provide enough context to describe the forecaster decision making process in detail. As such, it was difficult to tease out the role that probabilistic information was playing.

Having two forecasters for the Little Rock WFO, however, provided a small opportunity to understand the use of probabilistic information during the collaboration process. At one point,

when evaluating the NBM/QPF, one of the forecasters evaluated the different QPF percentile values. After flipping back and forth between the various thresholds, the forecasters determined that the 75% QPF was more in-line with their thinking. Therefore, during the call, they proposed an increase to more align with the 75% QPF.”

It is important to note though that chat content, based on the labels assigned as described in Section 6.1, included roughly twice the amount of probabilistic statements than deterministic (Figure 13) - a rather significant difference from the trends found in AFDs

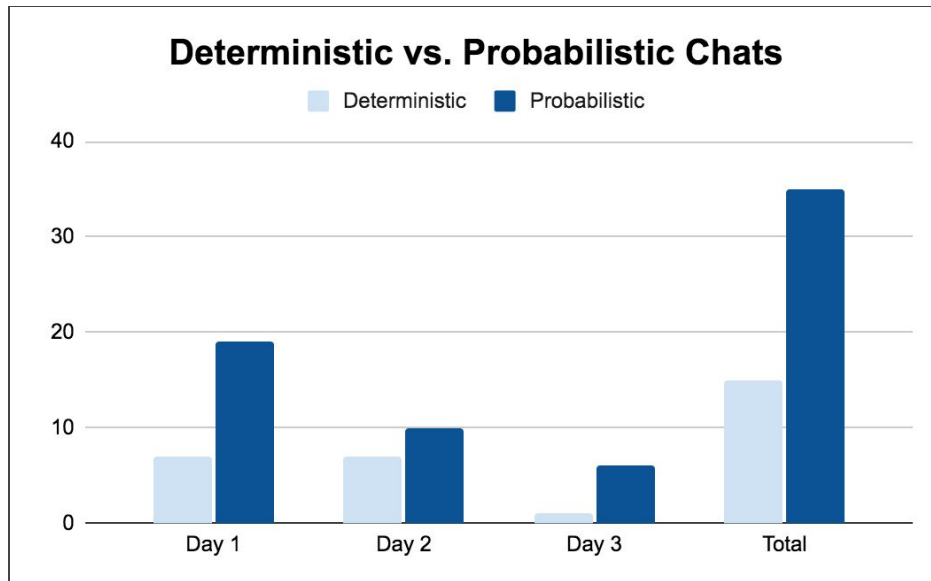


Figure 13 - Count of statements in chat logs relating to either deterministic or probabilistic guidance.

Further, participants noted the importance and value of using probabilistic data in both the daily debriefings and the final participant survey.

“I thought the probabilistic QPF from the NBM, especially looking at the 1D viewer, seeing all different ensembles, was my best tool...besides looking at just the general synoptic and mesoscale setup, was the best tool in being able to try to say with some type of skill if the NBM QPF was representative or even the WPC QPF was representative on my thinking.” From the debriefing on day 3.

“Really in convective environments probabilistic data needs a whole lot more usage.”

Participant #6

Participant #1 even linked using probabilistic data as a necessary component of the Collaborative Forecast Process stating, ***“The collaborative forecast process is using the most likely and probabilistic data, combined with the forecast funnel process, and finding common ground between WFOs, regional RFCs & national centers in the forecast.”***

Finally, two participants noted the potential future challenges if the NWS moves to probabilistic QPF grids:

"I don't know that people (NWS forecasters) understand enough statistics yet to even try to attempt to modify distributions of 10th and 90th." Final Debrief

"Yeah, I think, not necessarily editing the probabilistic grids because, again as everybody's saying, you can get into trouble very quickly, when you're trying to edit statistics." Final Debrief

8.2 DETERMINE FORECASTER PREFERENCES FOR QPF ANALYSIS

As noted above, this experiment failed to produce results that can adequately describe the forecaster preferences for QPF analysis. In our case, the ten participants certainly favored using probabilistic data, but those results may not be representative of the entire forecaster community.

9. COLLABORATIVE FORECAST PROCESS WORKFLOW EVALUATION

The CFP plan alters the flow of the forecast process in operations. The plan was designed around the timing of the NBM data rather than any other model cycle, NWS product cycle, or any other institutional timelines (for instance, producing products to meet local media needs). The PMO made this decision intentionally and through a collaborative effort from individuals representing every aspect of NWS operations (local offices, regional/national centers, etc.). This change in flow will impact when forecasters review data, what data is available for review, and when certain decisions need to be made regarding forecast changes or additional collaboration needs (like a conference call or video chat).

The OPG designed our exercise around this new timeline and gathered participant sentiments of the new process. We could not, of course, review verification from one event and make broad statements about the resulting forecast accuracy from this approach. We would need to conduct this exercise multiple times, with multiple participants, around multiple types of events in order to make such a determination. Thus, our results are focused on how the forecasters felt about the process.

The PMO suggests that using a common starting point (the NBM) and collaborating early in the forecast process (prior to making grid/forecast modifications) will lead to a more productive, informative, and valuable collaborative experience. An experience that eliminates bias based on the concept of "forecast ownership" and helps focus our participants on a common goal: identifying if and when the starting point is not representative, and discussing the necessary changes. Further, the most recent NWS Directive describing NDFD states, "Ownership of the NDFD is shared among all those involved in the collaborative process." (Directive 10-201 section 5).

9.1 THE COLLABORATION FLOW (TALK FIRST, EDIT SECOND)

The participants were asked if they preferred to edit the forecast first and then collaborate, or if they would rather start with the NBM, collaborate, and then modify the forecast. Only two of

the ten participants preferred to edit the forecast first, but one of those two individuals later stated in the short answer section that, *"...my current preferences are to load & edit grids & then collaborate. However, I do think that we should probably lean the other direction in terms of QPF & maybe a few other variables."*

So in reality, only one individual expressed concerns about collaborating prior to forecast editing. They explained their reasoning in the short answer section stating, *"I just think we are forcing a particular answer or thereabouts in the end by telling everyone that they have something that should be good enough most of the time, which will lead to very few changes to said starting point."* This participant's concerns do represent some percentage of field forecasters who are concerned that relying on the NBM may lead to reductions in forecaster expertise.

The remaining participants all felt strongly that the method we tested during the exercise was superior to prior forecast processes:

"Seems like a waste of time to edit grids, then collaborate, then edit grids again. Just load NBM, collaborate, then edit as necessary based on collaboration." Participant #8

"I believe it makes more sense for WFOs to provide feedback to one entity (WPC) and have them make high-level edits and try to integrate all the feedback in a meteorologically consistent way. I think having everyone come up with their own independent answers makes us susceptible to anchoring effects to our own initial thoughts, and seems counterproductive to the aim of getting one agency QPF." Participant #5

"From a personal perspective, I'd rather not essentially double the work in some cases where I would edit grids first and then collaborate and then potentially have to re-edit grids. Using the NBM and the collaborating from there would, in my opinion, involve less grid editing, which would seem to save time for other important elements of the forecast process. I think it is more productive and prevents the potential for time-consuming disagreements that may lead to more workload if the forecast already produced needs to be adjusted." Participant #10

"To me, it's important that the collaboration occurs before the grids are edited. This allows forecasters to draw upon the expertise of other forecasters at the local, regional, and national levels." Participant #2

During the debriefings, the participants brought up an issue that occurs during the current forecast process in operations. They essentially described a scenario as a *"shotgun blast of grids"* with the apparent intent of getting the "first say" in any subsequent collaboration process. By saving and sending grids prior to neighboring offices, forecasters essentially create an anchoring bias used in negotiation tactics. One participant stated, ***"Many people already do have the mentality that they own the forecast, they own the grids, they created it. It's a lot harder to convince someone to change when you have that mentality."***

As described in the [Harvard Law School Daily Blog from July, 2020](#):

“The negotiators who made the first offer felt more anxiety than those who did not – and, as a result, were less satisfied with their outcomes. Yet, backing up prior bargaining studies, those who made first offers did better in economic terms than those who did not. If you value only the economic outcome of your deal, make the first offer in order to anchor the negotiation in your favor. But if you value satisfaction with the negotiation process more than the outcome itself, you may want to avoid the stress and anxiety of making the first offer.”

Put in the context of a weather forecast, the first person to share their forecast with neighboring offices sets the anchor for negotiation. This will force neighboring offices to come into alignment with the anchor value. Further, the forecaster who sends their forecast first may experience increased anxiety and reduced job satisfaction.

9.2 THE VALUE OF A COMMON STARTING POINT

All participants who chose to respond to this question (one did not) indicated that there was significant value in using a common starting point to more efficiently arrive at a consistent forecast. The NBM was most commonly noted as the desired starting point. There were some who indicated a preference for having WPC make adjustments to the NBM first. However, determining the validity of using the NBM alone or leveraging WPC’s modified NBM as a starting point was not an objective of this exercise and therefore warrants further investigation.

In this exercise, the deterministic QPF from the NBM was deemed “pretty reasonable” by our participants for most of the forecasts. They had little reason to debate the accuracy of the starting point and **therefore spent more time discussing impacts, scenarios, and messaging (something the participants felt was more valuable than discussing QPF amounts)**.

“I think it was pretty effective. Even with some differences in thoughts, ideas & familiarity of the forecast funnel methodology/techniques & improving local high resolution data, the NBM seemed to capture the QPF fairly well.” Participant #1

“If we have to use a common starting point..., one designed to be as robust as possible like the NBM is the best way to increase collaboration, as it should take out much of the guess work yourself, leaving more time to focus on the most important event.” Participant #4

“Having a starting common point helps the process move quickly and everyone can collaborate on the same thing.” Participant #7

“It helped a little in that it cut down on the number of ‘well the GFS says this and the ECM says this,’ however I still saw a lot of references to other model data (Hi-res ensembles, IVT, PWAT, etc.) that made for a bunch of meteorological back and forth, and again, it made it hard to keep up. On the other hand, looking at the other data sources to back up why people felt NBM was too high or too low is probably necessary ...” Participant #8

"It was more effective this way. If we all began the CFP with the same forecast, then it shifted the emphasis on collaboration: If you identify a need to stray away from the common starting point, you have to communicate why during collaboration - or else you risk being the one responsible for creating an inconsistent forecast." Participant #9

"In most cases, having a common starting point allowed for more time for productive conversations...and reduced extreme discrepancies... and with more productive conversations, a more consistent forecast could be developed." Participant #10

There was at least one participant who saw the need for additional training to accomplish this change in the forecast process.

"First, [using the NBM as a common starting point is] a completely different approach to the forecast process than what most are used to, so there needs to be training/guidance provided to the forecasters so they can tactfully and effectively identify targets of opportunity to improve upon the common starting point." Participant #9

Finally, it is important to note that some participants expressly noted the value of having WPC in the process. No participant expressed any resentment or concern of leveraging WPC as national experts in the QPF production process.

*"I'd have to say using the NBM gave all of us the same starting point. However, **having an expert(s) at WPC already modifying the delivered NBM fields ahead of time may be quite advantageous** as they have the expertise and ability to modify that dataset as needed based on the latest favored models, trends, and known biases."* Participant #2

9.3 THE CONTENT OF COLLABORATION

When asked, "What should be the focus of the collaboration phone calls/video calls?," every participant agreed that some aspect of messaging, impacts, or hazards should be discussed on the collaboration calls/video calls. Half of the participants mentioned that some portion of the calls should focus on QPF collaboration, but four of the participants made it clear that meteorology should only be minimally discussed or NOT discussed on calls at all.

Specific responses indicated some disagreement on what the topical focus of the calls should be. One participant stated, *"If you have local Google chat rooms & Google Meet video calls, you can hammer out the meteorology. However, on national/regional collaboration calls, impacts, hazards & messaging should be the key."*

Another response similarly stated, *"For the majority of them it should all be about impacts: Hazards and messaging. Keep QPF collaboration as much to chat or something like AGOL."*

By contrast, another participant stated, *“The majority of time should be spent discussing any significant changes regarding the QPF forecast from what was initialized, and why those changes would be made.”* However, that same individual did go on to state that, *“perhaps close to half of the call could focus on expected impacts and messaging to partners.”*

In short, **the main point of agreement was that there is no one-size-fits-all call agenda.** The topics will necessarily vary depending on the situation. This will, in turn, result in variable representation on the call (see section 9.4 for more details).

9.4 The Participants on Collaboration Calls

“Throughout the experiment, the role of the ROC was not immediately clear to the participants or the observers.” From the social science feedback on the CFP Experiment.

When the OPG creates an exercise, we necessarily modify some of the reality in order to focus on specific concepts. Normally, this is not a major inhibitor to achieving quality and meaningful results. Indeed, during the virtual CFP exercise, the OPG was able to evaluate collaboration tools, timing, and processes even without perfect realism. One aspect though was very difficult to evaluate because it is highly variable during real world events. That is the aspect of “who participates on a collaboration call/video meeting?”

Typically, in real events, identifying participants for a call is ad hoc, but based on common principles. For example, during rain and flood events, both WFOs and RFCs who will be impacted by the event along with WPC and the regional ROCs are expected to participate. It is less clear when the National Water Center joins these calls or exactly where to draw the line between offices who are, or are not, impacted by the event.

As such, some offices may be left off a call when they should have been invited, and other offices participate in a call when there was no real need. A critical aspect of a CFP for any program area is the structured and deliberate selection of call participants. **Too many participants on a call can create chaos and a lack of focus. This is a challenge for the national CFP for QPF.** Rainfall and flooding events can occur across a large number of local offices when considering an event that lasts for several days, or progresses across the US over the course of a week.

Thus, one weakness of the virtual CFP experiment is that we were limited to our ten participants for the collaboration calls. Even though the participants represented a cross section of the NWS, they did not represent the amount of participants that are common for significant events. Further, it was clear, based on the design of the exercise, who was expected to be on the call.

During the debriefings, and as noted in the quote at the top of this section from our social science team, the role of certain participants was less clear. We asked our participants in the final survey, “who SHOULD be on these collaboration sessions” and the results were mixed. For some, it was situationally dependent.

"If it is just focused on QPF - WPC, WFOs, RFC. If it is on hydro impacts and messaging - WPC, WFOs, WFCs, NWC and ROCs." Participant #7

"This answer can vary considerably depending on what the weather threat is. If we are just focusing on QPF/flooding (similar to how this CFP exercise was designed), then I would imagine we had the right balance of participants actively collaborating." Participant #9

For others, they felt that all of the interested offices were welcome to be on the call for situational awareness, but that not everyone needed to speak.

"...I suppose, I don't necessarily see a problem with inviting a large cross-section of offices -- again for situational awareness. However, we need to have protocols for narrowing speaking roles and keeping them relatively concise and efficient." Participant #5

And still some others simply felt that everyone should be on the collaboration sessions.

As already alluded to, there were certain entities whose role the participants specifically questioned. Such was the case when Participant #1 said, *"I think the ROCs being on these calls, especially if we have more targeted rooms & chats & additional video calls, could get overwhelmed. I don't think the ROCs or RFCs have to be on every call & we need to target the WFOs most impacted or largest changes needed/hazards introduced."* Participant #8 added, *"To me, having NWC and ROCs on the call is nice to have, but they don't own the forecast and don't necessarily need to be part of the FORECAST process."*

While some questioned the need for RFCs to participate, such as Participant #1, others felt the RFCs should join the calls. Participant #10 stated, *"I particularly liked having the RFC around the calls. ... Would be nice to have [RFCs] on these calls to get their perspective and thoughts on things. I've always thought communication between RFCs and WFOs could improve."* With the CFP making it more likely for collaboration calls to take place focusing on QPF amounts, the role of the RFCs will need clarity.

The need for clarity on the role of the ROCs was also apparent. Participant #5 said, *"I've heard conflicting things about ROCs. Some have expressed that they would prefer the relevant national center leads the call. Others would prefer the ROCs to lead the call. Of course there are issues with 24x7 staffing."* Historically, the ROCs would participate, or lead, conference calls during major events (like landfalling hurricanes). They would not, generally speaking, engage in calls focused on moderate or weaker events (like the case we used for the exercise) even if headlines were expected. During the exercise, the ROC participants expressed their concerns stating, *"we wouldn't normally participate in a call for an event like this [the exercise event]."*

Our ROC participants crafted a set of notes for the OPG describing their experience. They echoed the sentiments of the social scientists and noted they did not feel the ROCs should participate in collaboration calls focused on the forecast (QPF values). They expressed concerns about having "too many cooks in the kitchen" when the calls focused on QPF amounts.

Finally, **it was not obvious to our participants who should lead collaboration calls.** WPC, by default, facilitated each call, but call leadership is an important role that is not expressly described in the CFP.

10. NATURE OF FORECASTING - INFLUENCERS TO THE CFP

During the course of the CFP experiment, as with many OPG experiments, the OPG and observers noticed comments or behaviors that raised questions beyond the scope of the stated objectives. In these cases, the OPG is faced with a dilemma; do we ignore these situations entirely or do we address them in the report with caveats? For example, participants hinted at the concept of “picking the most representative model” to form the basis for their forecast, or “waiting for the next model run to make a decision”, and finally, “questioning the lack of detail in the NBM QPF at medium and long ranges.”

We felt the implications of the above statements are significant enough to include in the report with the following caveats:

- We believe the concerns in 10.1 and 10.2 represent a large number of NWS meteorologists and hydrologists and therefore justify further rigorous investigation
- The OPG did not gather enough evidence to confirm or deny the validity of various forecast methods within the context of the experiment, but evidence from the greater atmospheric science community support the need for additional investigation
- The participants of our experiment performed exceptionally well and the following topics do not in any way question or negate their expertise, skill, or capabilities as forecasters or hydrologists

10.1 FUTURE CONSIDERATIONS FOR FORECAST METHODS

Participants expressed a few concerns about using the NBM run as the starting point of the CFP. Paraphrasing, their three main concerns were:

- *“The 13Z NBM is based on old data because the 13Z NBM leverages either 00Z or 06Z global model guidance.”* Shortly after the 13Z NBM arrives in AWIPS (around 14:30Z), the 12Z GFS starts to flow into AWIPS as well. As such, participants preferred to wait for the latest global deterministic model data (the 12Z runs in this case) to help decide if the 13Z NBM was representative or not.
- The NBM is a “black box” and without access to all of the inputs of the NBM and understanding of the post-processing techniques, some participants felt they could not determine whether or not the NBM was representative. Specifically, participants suggested that having access to all the NBM inputs would allow them to select the most representative solution instead of relying on an ensemble mean.
- The perceived lack of accuracy, or rather detail, in the NBM QPF especially in the medium or extended range suggests the NBM is not capable of diagnosing extreme events.

The OPG believes these sentiments to be pervasive among NWS forecasters. After all, the sentiment makes logical sense when phrased as a question, “Why should forecasters use 6 to 12 hour old data when updated information will arrive soon?” Or, “How can an ensemble possibly represent an extreme event?” However, underneath the concerns are implied capabilities or facts that are not yet objectively confirmed.

For example, the desire to wait for the “latest deterministic model” implies that said model will provide information that allows forecasters to make informed changes to the NBM starting point. Specifically, the 12Z GFS provides forecasters with evidence to support changes to the 13Z NBM (which is based on 06Z GFS data among other guidance). Of course there are two inherent flaws in the desire to wait for new data because: 1 - there is always new data on the horizon and 2 - the new data does not always improve over the prior forecast (as noted in the ECMWF Forecast User Guide). **At some point, forecasters must make a decision, put their pencils down, and publish the forecast.** The current plan for the CFP demonstration lists this time as either ~8:00Z or 20:00Z (roughly 5 to 6 hours after the NBM starting point is available).

Further, the desire to access all, or a majority of, the individual NBM inputs implies forecasters have the ability to select an input that is more representative than the post-processed solution. This concept is commonly referred to as “**model picking**” **but this ability has not yet been objectively validated** (to OPG’s knowledge).

10.2 FUTURE CONSIDERATIONS FOR NDFD

What do, or what should, NDFD grids represent? Over the course of the exercise, including the debriefings, it became clear to the OPG and observers that the operational forecast community lacks a common understanding of the nature of NDFD. The current NDFD Directive, 10-201, states in Appendix A section 2, “Element values represent conditions of meteorological fields at the resolution of the grid.” However, this definition focusing on grid resolution does not specifically address the bigger question of what should be represented in NDFD. Some forecasters believe NDFD should present the most likely deterministic forecast. Others feel that forecasters should account for uncertainty in grid production and therefore necessarily decrease precision at various time scales.

As an example, forecasters have noted concerns over the NBM appearing “too smooth” for QPF especially in the extended range forecast. These forecasters would prefer to utilize a raw deterministic solution as a starting point because those QPF grids “look realistic”. Others would argue that the NBM is justifiably “smooth” because it attempts to represent uncertainty in a single deterministic value. In either case, the OPG believes the multitude of perceptions among forecasters regarding the nature or purpose of NDFD will necessarily inhibit collaboration efforts.

What is clear is that NDFD is only capable of providing a single deterministic forecast to end users for most forecast elements (snowfall being a notable exception). Our participants frequently discussed alternate rainfall scenarios during the exercise, but would not be able to

articulate these scenarios via NDFD. Thus, they would need to rely on IDSS content to express worst case or best case scenarios.

Using the NBM as a common starting point should reduce the debate, but as noted above, some forecasters would prefer to select a deterministic model rather than use the NBM. It is not clear in the CFP plan if forecasters will be able to switch from the NBM starting point to a deterministic option at some point during the forecast production process.

The OPG agrees with some of the exercise participants and many others in the forecast community that it is worth investigating alternative options for NDFD - including the concept that QPF should not be represented as a single number, but rather represented as a probabilistic distribution.

11. TRAINING NEEDS

"Development of standard/best practices and training will be crucial to ensure a smooth transition [to a collaborative forecast process] and efficient workflow. Training on "How to collaborate" could [have a] design similar to IDSS training (be clear and concise, focus on specific goals and topics to discuss, limit the meteorology talk)" Participant #9

For the national CFP for QPF demonstration, forecasters and hydrologists need to understand the spirit and intent of the CFP. The most important training needs for the national CFP for QPF demonstration are related to the human factors of collaboration. Specifically, forecasters and hydrologists need to understand (not an exhaustive list):

- The changes to operational workflow during a CFP
- The proper use of collaboration tools
- The roles and responsibilities of all parties involved in the CFP
- The types of content that should, and should not, be discussed in collaboration sessions
- The types of roadblocks to effective collaboration and how to resolve them
- Effective techniques for improving the collaborative experience
- The importance of using the NBM as a common starting point
- The proper use of deterministic and probabilistic guidance in maximizing forecasters' ability to assess the representativeness of the NBM

It is not the role of the OPG to dictate how this training should be developed, what materials should be used, or how it should be delivered. Instead, we humbly offer our experience prior to and during the exercise as examples for consideration.

It is not our intent to suggest the CRM framework should be required for operational forecasters or hydrologists. Rather, we merely note that our participants found value in the principles taught in CRM and that the NWS should review CRM materials for possible inclusion in any training program.

Further, the OPG does not intend to denigrate, or otherwise negate the value of deterministic data in the forecast process. Rather, we offer our exercise experience, and the expertise of subject matter

experts, to suggest there are proper, and improper, methods of applying deterministic and probabilistic data in the forecast process. We simply believe forecasters and hydrologists in the NWS have different levels of understanding regarding the use of probabilistic data in the forecast process.

12. FINDINGS AND RECOMMENDATIONS

The following findings and recommendations were derived from analysis of the collective comments, suggestions, recurring opinions, observer input, anecdotal evidence, and important takeaways shared by the ten individuals who participated in the 2020 virtual CFP for QPF experiment. We have broken down the summary into three parts:

1. Findings and recommendations for the National CFP for QPF Demonstration
2. Findings and recommendations for the forecast process
3. Findings and recommendations for the future

It should be understood that “the future” findings and recommendations are not necessarily required for the CFP for QPF National Demonstration, but rather are intended to speak broadly about an eventual operational CFP for all program areas.

Regarding the CFP for QPF National Demonstration:

Finding 1: The OPG found no serious concerns during our experiment that would prevent the CFP for QPF National Demonstration from proceeding. There are a few “low hanging fruit” challenges to address, such as providing clarity on roles and responsibilities, but these are not major roadblocks to the demonstration. Addressing the human factors of collaboration (see Recommendation 2) are the most pressing needs prior to the start of the CFP demonstration. Otherwise, the NBM provides a quality starting point and is critical to the success of the CFP for QPF. Forecasters and hydrologists have access to the data they need to evaluate the representativeness of the NBM and can effectively share their thoughts with the collaboration tools at our disposal today.

Recommendation 1: Proceed with the CFP for QPF demonstration as soon as possible.

Finding 2: Existing collaboration methods and practices (what we say, how we say it, who we say it to, what we used to say it) within the NWS evolved over time based on subjective experiences or anecdotal evidence. To the OPG’s knowledge, there are no formal or vetted (by social scientists) guidelines detailing roles, responsibilities, expectations, best practices, structure, or conditions for engaging in a highly effective collaborative session. Specifically, determining who organizes a collaboration call, who joins that call, who leads the call, what should be discussed, and what should not are not well defined. As such, current collaboration experiences within the NWS were viewed with mixed feelings among our participants and likely represent the broader operational community. Further, ***the current collaboration experience in NWS operations can be better defined as “negotiation” or “coordination.”*** That is, forecasters will use tools like chat to inform their neighbors on expected changes (coordination) and may offer to adjust the forecast IF their neighbors also make adjustments (negotiation). But, ***by engaging in collaboration prior to the production of QPF grids or other forecasts, and rejecting the concept of individual forecast “ownership,” the participants found the resulting collaboration to be more productive and effective.***

Recommendation 2: Prior to the CFP for QPF National Demonstration, the NWS should provide training to operational forecasters and hydrologists on effective collaboration methods during time-sensitive, uncertain, and high stress situations. Additionally, provide guidance to operational forecasters and hydrologists describing clear roles, responsibilities, and expectations for the CFP. During the CFP for QPF National Demonstration, local offices and ROCs should conduct structured debriefings routinely with operational staff to evaluate the effectiveness of collaboration sessions and then apply lessons learned to future sessions. After the CFP for QPF National Demonstration, the NWS should work with social scientists to conduct an overarching analysis of the collaboration experiences to help inform a robust collaborative structure for future CFP initiatives.

Regarding the Forecast Process:

Finding 3: Discussions and debriefs that took place during the CFP exercise revealed an apparent misunderstanding of how to apply current science to the forecast process. In particular, the longstanding practice of attempting to select the best deterministic model solution as the basis for a forecast appears to remain common, while the value of applying calibrated, post-processed ensemble-based model information to forecast decisions is underappreciated and perhaps not well understood. Moreover, the assumption that a more recent deterministic run trumps an older ensemble remains anchored into the mindset of many forecasters. This is consistent with anecdotal evidence of current operational practices as well. For example, it is apparent from the verbiage in AFDs around the country that many forecasters treat the NBM as another deterministic model, rather than the post-processed blend of more than 100 models. Failing to incorporate effective use of probabilistic data in the forecast process threatens to hinder our commitment to continually strive for science-based service evolution in an effort to improve our value to society. Finally, there are differing opinions among the operational forecast community (and others) regarding the nature of NDFD and what the grids should represent. This difference in interpretation will pose significant challenges to the collaborative process if not resolved.

Recommendation 3: The NWS, potentially led by the Office of Science and Technology Integration (STI), should form a team, or teams, including NWS operational forecasters and subject matter experts, to objectively assess the the following concepts (among others):

- The time required to evaluate the representativeness of the NBM starting point
- The value of waiting various periods of time for new deterministic model data to inform forecaster decision making
- The ability of forecasters to select the most representative deterministic model when preparing a forecast
- The proper amount of precision to include in a 2.5 km grid at various time scales
- The variability of grid production methods and their impacts on the forecast quality
- The impact to forecast operations in a fully probabilistic NDFD environment

Results from these assessments will inform a forecasting framework that best aligns with and improves the collaborative experience.

Finding 4: In moving to the NBM as a common starting point, forecasters will be expected to evaluate the representativeness of a frequently updating, multi-model, post-processed ensemble system. Therefore, leveraging ensemble data is likely the best approach in evaluating the representativeness of the NBM. In addition, recent cognitive psychology research indicates people make more effective risk mitigation decisions when provided probabilistic information compared to those who are given strictly deterministic data. (Joslyn, S. L., & Grounds, M. A. 2015. The use of uncertainty forecasts in complex decision tasks and various weather conditions. *Journal of Experimental Psychology: Applied*, 21(4), 407–417. <https://doi.org/10.1037/xap0000064>). Probabilistic information was used heavily during the CFP for QPF experiment by the ten participants, but it is yet unclear if the broader operational forecasting community in the NWS leverages the data within the forecast process. To the OPG’s knowledge, the NWS currently lacks specific, detailed, and objective guidelines for determining when to use deterministic and probabilistic data in the forecast process. For example, the European Centers for Medium Range Weather Forecasting (ECMWF) produced a robust web based user guide for forecasters that clearly defines the objective reasoning, and use cases for both deterministic and ensemble based data in the forecast process.

Recommendation 4: The NWS should provide operational forecasters with training on the application of ensemble based information in the forecast process, specifically addressing methods for assessing the representativeness of the NBM. Operational forecasters and hydrologists should review the [ECMWF Forecaster User Guide](#) (the OPG humbly suggests conducting group discussions) and apply lessons learned to the forecast process.

Regarding the Future Needs for Operational CFPs:

Finding 5: The collaborative tools (chat, video, geospatial) available to forecasters today are effective enough to support the CFP for QPF Demonstration. Forecasters currently leverage several different tools for the collaborative process including, but not limited to: AWIPS collaboration chat, NWSChat, Google Chat chat rooms, Google video meetings, hurricane hotline video systems, GFE, and of course the telephone. Unfortunately, these collaboration tools exist on different systems that are not necessarily interconnected, that require separate login credentials, and have their own unique user experiences. The OPG, through discussions with our exercise participants, personal experience, and anecdotal evidence, is concerned that the current piecemeal collaboration framework will hinder future collaborative efforts. The OPG is unclear if a singular system that meets all user collaboration needs is required, or who should develop such a system, but we do recognize our collaborative environment today is largely unstructured and inefficient.

Recommendation 5: The NWS should conduct a brainstorming meeting with operational forecasters and hydrologists, AWIPS developers, and potentially include third party software vendors, to craft idealized collaborative environments to meet our future CFP needs.

13. NOTE OF APPRECIATION

The OPG would like to thank our incredible participants who trusted our methods and performed exceptionally well during the exercise. All of the observers and the OPG were incredibly impressed with your efforts during the exercise.

We would also like to thank Dr. Trevor Alcott, Matthew Jeglum, and Tim Oram for providing the participants with excellent information prior to the exercise.

We would like to thank our social science team who provided expertise, analysis, and incredible insights into our exercise.

We would like to thank the group of observers who provided extra eyes and analysis during this complex unique exercise.

Finally, I would like to thank the team at the OPG. Their incredible ability to innovate and adjust to a virtual exercise was simply extraordinary.

This report represents a massive team effort built upon the expertise and hard work of many. We hope this report inspires readers to engage in thoughtful discussions about the forecast process, the CFP, and our future.

Sincerely,

John J. Brost

Director, Operations Proving Ground

October, 22nd 2020

APPENDIX 1 - CHAT ANALYSIS CATEGORY NAMES AND DESCRIPTIONS

Categories

- **General Conversation:** Welcome, good mornings, introductions
- **Confirmation:** Noting that a message was received, “Got it - thanks!” or expressing thanks for something, “Thanks for the data.” Confirmation is usually directed to another person rather than the content of their post.
- **Deterministic:** Discussing signals/data from deterministic models)
- **Probabilistic:** Discussing signals/data from a probabilistic guidance)
- **Diagnostic/Reasoning:** Discussing the meteorology or hydrology)
- **Request for Feedback:** A question that drove a response, “What are your thoughts in increasing rain amounts?”
- **Feedback Given:** Response to a feedback request that represents a decision or triggers and action, “We agree with increasing QPF.” Feedback is usually directed toward the content of a post, rather than the individual.
- **Filler:** Comments that are related to the conversation but lack substance, “That’s impressive.”
- **Noise:** Distracting, unrelated comments
- **Impact / Hazard Info:** Discussions focused on impacts or flooding - not including specific discussions of headline coordination.
- **Headlines:** Explicit discussions of headlines
- **Forecast:** Explicit statement regarding their forecast values
- **Uncertainty:** A statement where the forecaster expresses their lack of confidence, or certainty, in either the information they are reviewing, or the forecast they are making
- **Expression of Emotion / Gut Feeling:** A referencing some kind of internal feeling regardless if it includes data as support or not. “I’ve got a bad feeling about this”. Or, “I’m growing concerned about day 5...”)

APPENDIX 2 - EXAMPLE ANALYSIS OF CHAT STATEMENTS

The following statement was provided by an individual on the first day of the experiment. The content is unedited.

“The SA table indicates nearly 97-99th percentile IVTs for even the 1st system to prime my eastern areas. The ridge is anomalous like nearly maxed out in the NAEFS table climatology. So we won't get out of this pattern of continuous rounds of convection. Thank you Paula for that information. Its great to hear that streamflow is low. Needs to to stay that way 😊 but I know it won't stay that way for long. GEFS is maxed out on PW climatology for last system, which goes back to 1985 for the Day 7. Agreed, even though that seems to be a target, if anything, any change to NBM could be in shorter term, for now.”

Statement Made in Chat	Applied Category Label
The SA table indicates nearly 97-99th percentile IVTs for even the 1st system to prime my eastern areas.	Probabilistic because the main focus is the percentile values of the IVT.
The ridge is anomalous like nearly maxed out in the NAEFS table climatology.	Probabilistic because of the reference to “anomalous” based on NAEFS data.
So we won't get out of this pattern of continuous rounds of convection.	Diagnostic because they provided expectations based on the prior two statements.
Thank you for that information.	Confirmation because they were acknowledging information provided by another participant.
Its great to hear that streamflow is low.	Feedback because they are commenting on the content of a post (not just acknowledging it).
Needs to to stay that way 😊 but I know it won't stay that way for long	Filler because the comment is relevant to the discussion, but is mostly opinion.
GEFS is maxed out on PW climatology for last system, which goes back to 1985 for the Day 7.	Probabilistic because the focus of the statement is the GEFS.
Agreed, even though that seems to be a target, if anything, any change to NBM could be in shorter term, for now.	Feedback, Uncertainty, and Forecast because the first statement is providing feedback to a prior post but uses the term “seems” and “if anything” which suggest uncertainty, and the then specifically addresses the forecast values.

APPENDIX 3 - CFP EXERCISE DEVELOPERS

Name	Position	Office
John (JJ) Brost	Director	NWS OPG and PMO Initiative Lead
Matt Foster	SOO	NWS OPG
Dr. Katie Crandall Vigil	Research Scientist	OU-CIMMS/NWS OPG
Ryan Difani	Research Associate	OU-CIMMS/NWS OPG
Jack Richardson	Sys Admin / ITO	NWS OPG
Dr. Chad Gravelle	Tech Dev Met	NWS SRH STSD
Kim Runk	Senior Planning Adviser	NWS OSTI
James Nelson	Chief - Development and Training Branch	NWS WPC

APPENDIX 4 - CFP EXPERIMENT AGENDA

<u>TUESDAY (All Times CDT)</u>	
9:00 AM - 9:15 AM	Exercise - Receive Weather Briefing
9:15 AM - 10:00 AM	Exercise - Analyze Provided Forecast Data
10:00 AM - 10:30 AM	Exercise - Collaborate as Needed Using Google Chat (No Graphic Sharing)
10:30 AM - 11:00 AM	Exercise - Opportunity for a Collaboration Call via a "Conference Call" (Occurs as Needed)
11:00 AM	Exercise - WPC Provides Updated QPF
11:00 AM - 11:30 AM	Exercise - Produce Deliverables focused on Partner Communication
11:30 AM - 1:00 PM	Lunch
1:00 PM - 2:00 PM	Debrief
<u>WEDNESDAY</u>	
9:00 AM - 9:15 AM	Exercise - Receive Weather Briefing
9:15 AM - 10:00 AM	Exercise - Analyze Provided Forecast Data
10:00 AM - 11:00 AM	Exercise - Collaborate as Needed Using Google Chat (Full Functionality) and Google Meet
11:00 AM	Exercise - WPC Provides Updated QPF
11:00 AM - 11:30 AM	Exercise - Produce Deliverables focused on Partner Communication
11:30 AM - 1:00 PM	Lunch
1:00 AM - 2:00 PM	Debrief
<u>THURSDAY</u>	
9:00 AM - 9:15 AM	Exercise - Receive Weather Briefing
9:15 AM - 10:00 AM	Exercise - Analyze Provided Forecast Data
10:00 AM - 11:00 AM	Exercise - Collaborate as Needed Using Google Chat (Full Functionality), Google Meet, and ArcGIS Online
11:00 AM	Exercise - WPC Provides Updated QPF
11:00 AM - 11:30 AM	Exercise - Produce Deliverables focused on Partner Communication
11:30 AM - 1:00 PM	Lunch
1:00 AM - 2:00 PM	Debrief
<u>FRIDAY</u>	
8:00 AM - 11:00 AM	End-of-Week Survey
11:00 AM - 1:00 PM	Final End-of-Week Debrief

APPENDIX 5 - END-OF-WEEK ANONYMOUS PARTICIPANT SURVEY QUESTIONS

1. In your own words please define the collaborative forecast process.
2. How effective was QPF collaboration using phone conferencing? *(Answer choices: Not effective, Slightly effective, Moderately effective, Extremely effective)*
3. How effective was QPF collaboration using Google Chat? *(Answer choices: Not effective, Slightly effective, Moderately effective, Extremely effective)*
4. How effective was QPF collaboration using Google Video Chat? *(Answer choices: Not effective, Slightly effective, Moderately effective, Extremely effective)*
5. Rank your preferred collaboration medium. With 1 being your first choice, 2 being your second choice, and 3 being your third choice. *(Answer choices: Phone, Chat, Video)*
6. Now that you have used chat, phone calls, and video chat for collaboration, please describe what your ideal collaboration experience would be.
7. How valuable was it to be able to see the other people while you collaborated? *(Answer choices: Not valuable, Slightly valuable, Moderately valuable, Extremely valuable)*
8. How valuable was using screen sharing to collaborate during Google Meet? *(Answer choices: Not valuable, Slightly valuable, Moderately valuable, Extremely valuable, Not applicable)*
9. How valuable was highlighting an area of concern using ArcGIS Online? *(Answer choices: Not valuable, Slightly valuable, Moderately valuable, Extremely valuable, Not applicable)*
10. Did you experience conflict during the collaboration portion of the experiment? *(Answer choices: Yes, No)*
11. If you experienced conflict during collaboration how was it resolved?
12. Please describe your thoughts on how collaborating prior to forecast production impacted the collaborative experience.
13. How did using the NBM as a common starting point (instead of....) impact the effectiveness of the collaboration session?"
14. Do you prefer editing grids first and then collaborating instead of using NBM as a common starting point? *(Answer choices: Yes, No)*
15. Please explain your answer to question 14.
16. In each collaboration session we had WPC, WFOs, RFCs, ROCs, and NWC participating. Who do you think SHOULD be on these collaboration sessions during real-life operations?

17. What should be the focus of the collaboration phone calls/video calls?
18. If you were given a choice between producing the QPF forecast/grid the way we have done so on in the past, or utilizing the methods we tested during the exercise, which would you prefer and why?
19. Please describe how you feel about the collaborative process on your operational workflow.
20. Please describe your thoughts on how the collaboration process may change based on weather event types.
21. How important is it to train NWS employees on collaboration methods and techniques?

Additional questions about home setup, pre-requisite training, data presentation, and how to improve future evaluations by the OPG were answered by the participants. Those questions are available upon request.

APPENDIX 6 - END-OF-WEEK ANONYMOUS OBSERVER SURVEY QUESTIONS

1. Now that you have been given a chance to observe different collaboration methods (phone, chat, video with screen sharing), please tell us how you would choose to collaborate.
2. Please describe your thoughts on how collaborating prior to forecast production impacted the collaborative experience for participants.
3. How important is it to train NWS employees on collaboration methods and techniques? *Answer choices: Not important, Slightly important, Moderately important, Extremely important*

Additional questions about experiment success and how to improve future evaluations by the OPG were answered by the observers. Those questions are available upon request.

APPENDIX 7 - END-OF-WEEK FINAL DEBRIEF SESSION QUESTIONS

1. What does collaboration mean to you?
2. How would you resolve conflict in collaboration discussions?
3. How did you feel about collaborating before you edited any grids?
4. How do you feel about the NBM as your starting point coming into this?
5. If there are disagreements with the NBM starting point do you have any documented, or unwritten guidelines in your office of when you should adjust the starting point? Basically, how off does it have to be before you make changes or edits?
6. Do you think there needs to be a standard answer (related to the previous question) across all offices, or do you think it has to vary depending on what's going on?
7. If you were to be presented with information that was more probabilistic in nature than deterministic, would that be helpful or hurtful in your process? Would you be interested in manipulating probabilistic information to achieve the kind of messaging you're working towards? Do you want to edit 90th percentiles and make it more extreme for instance?
8. Thinking back through your experience this week: What technological features or tools do you think you need to collaborate effectively going forward?
9. Was there something from this exercise this week that you might miss when you go back, quote unquote to work? We present you with anything that you may not necessarily have in ops or don't take full advantage of that you're going to miss when you get back?

APPENDIX 8 - WORKSHOP PARTICIPANT'S HOME OFFICES

