

Community Gridpoint Statistical Interpolation (GSI) System and its role in bridging research and operational data assimilation community

Xiang-Yu Huang, Ming Hu, Hui Shao, and Don Stark

Developmental Testbed Center

Abstract:

The Developmental Testbed Center (DTC) serves as a bridge between research and operational data assimilation community by transiting the operational data assimilation system to a community resource and committing the contributions from research community to the operational repository.

Over the last three years, the DTC has made a significant investment to transit the National Centers for Environmental Prediction (NCEP) operational data assimilation system, Gridpoint Statistical Interpolation (GSI), to a community resource. The DTC has released three community versions of the GSI system with a User's Guide. The DTC also hosted the 1st Community GSI Tutorial in Boulder CO, and built a community GSI User's Page (<http://www.dtcenter.org/com-GSI/users/index.php>) to provide documentation and on-line tutorials to the research community. The DTC staff has been providing support to GSI users through gsi_help@ucar.edu since the release of version 1.0 of the community code.

During the same time, the DTC also plays a key role in bringing the contributions from the research community to the operational GSI repository. For this purpose, the DTC has created and maintained a community GSI repository and an extended regression test suite, and organized the GSI review committee. DTC has been supporting several community GSI development groups to use the trunk version of GSI and has worked with them to commit their contributions back to the repository since the initial setup of the operational and community GSI repositories.

1. Introduction

The Gridpoint Statistical Interpolation (GSI) system is a unified variational data assimilation (DA) system for both global and regional applications. It was initially developed by the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction (NCEP) as a next generation analysis system based on the then operational Spectral Statistical Interpolation (SSI) analysis system. Instead of being constructed in a spectral space like the SSI, the GSI is constructed in a physical

space and is designed to be a flexible, state-of-art system that is efficient on available parallel computing platforms.

After initial development, the GSI analysis system became operational as the core of the North American Data Assimilation System (NDAS) in June 2006 and the Global Data Assimilation System (GDAS) in May 2007 at NOAA. Since then, the GSI system has been adopted in various operational systems, including the National Aeronautics and Space Administration (NASA) global atmospheric analysis system, the NCEP Real-Time Mesoscale Analysis (RTMA) system, the Hurricane WRF (HWRF), the Rapid Refresh (RR) system, and the Air Force Weather Agency (AFWA) operational system. The number of groups involved in operational GSI development has also been expanded to include more development groups, including NASA Goddard Global Modeling and Assimilation Office (GMAO), NOAA Earth System Research Laboratory (ESRL) Global Systems Division (GSD), and National Center for Atmospheric Research (NCAR) Earth System Laboratory's Mesoscale and Microscale Meteorology Division (MMM).

Starting from 2007, the Developmental Testbed Center (DTC) has been collaborating with major GSI development groups to transit the operational GSI system into a community system, supporting distributed development. The DTC has complemented the development groups in providing GSI documentation, porting GSI to multiple platforms, and testing GSI in an independent and objective environment, while still functionally equivalent to operational centers. Working with NCEP Environmental Modeling Center (EMC), the DTC is maintaining a community GSI repository that is equivalent to the operational developmental repository and facilitates community users to develop GSI. Based on the repository, the DTC releases GSI code annually with intermediate bug fixes. The first community version of the GSI system was released by the DTC in 2009. Since then, the DTC has provided community support through the GSI helpdesk, community GSI webpage, and annual community GSI tutorial/workshop.

In this extended abstract, we will first introduce the efforts and achievements to bring the operation GSI to the research community and the available resources to the research community to start a GSI application. The section 3 describes the DTC efforts to provide an environment for research community to develop and test their code following the operational development and then to commit their contributions back to the GSI trunk. The summary and the future plan are in section 4.

2. From Operation to Research (O2R)

Since 2007, the DTC has been investing significant resources to make the operational GSI system to a community system, which includes a set of functions to serve and facilitate the research community using GSI as their research system:

- Built and maintaining a user support interface through GSI Users Webpage and Helpdesk

As a community system, the system itself and related information must be available in a free and easy way to the whole research community. The DTC has built and is maintaining a webpage of the community GSI system to serve the research community since the very beginning of the community GSI project. As the other community systems supported by the DTC, the community GSI webpage is called “the GSI Users Page”, which is hosted as a part of the DTC web server:

<http://www.dtcenter.org/com-GSI/users/index.php>

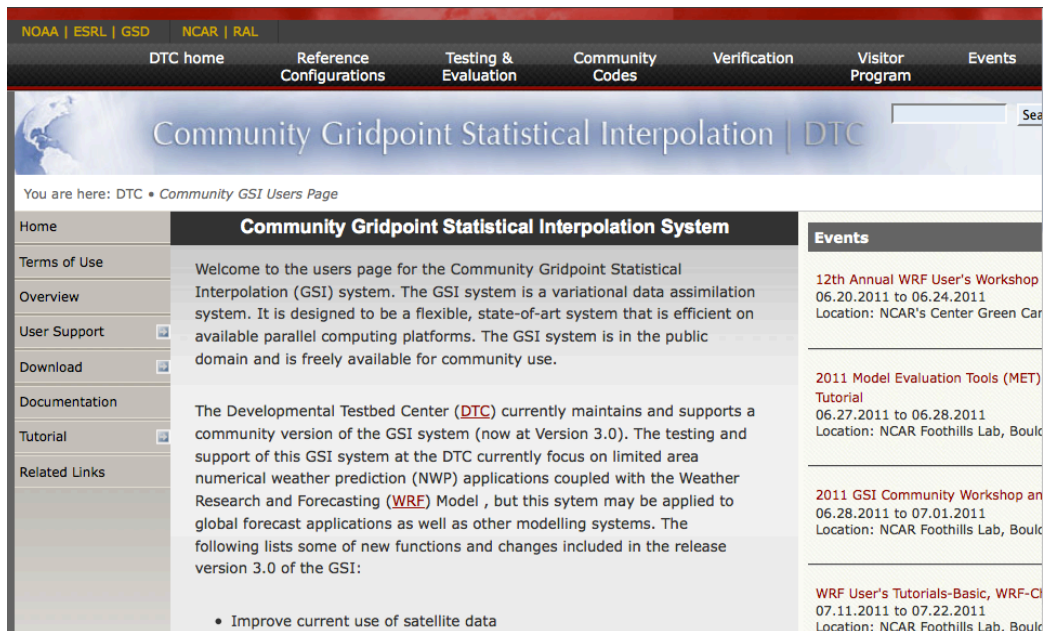


Fig 1. The main page of the GSI Users Page

From this website, GSI users can easily obtain the community GSI software package and useful documentation and information on the GSI such as the community GSI User’s Guide, the lectures from the GSI Summer Tutorial, the sample data sets for each release version, the announcement of GSI workshop and tutorial, and how to get support if you have questions or comments on the community GSI application.

As a part of the support interface of the GSI community system, the GSI Helpdesk provide a unified way for users to feedback questions and comments back to the DTC and operational community during their application of GSI. The GSI helpdesk is an email address:

gsi_help@ucar.edu

This Helpdesk is supported by DTC staffs with different expertise of the GSI system. If the users’ question is beyond the expertise of the DTC staff, the DTC will work with EMC and other system developers to solve the question. Based on the question, DTC staff usually answers them in the same day of the receiving but some questions need

many email iterations and several days to figure out a right answer for the users. DTC is committed to try the best to help users apply GSI in their research smoothly.

- Test GSI code on multiple platform and release GSI package annually

As a community code, the community GSI needs to be able to easily install and run on multiple computer platforms. GSI is an operational system that was mainly developed on IBM supercomputer. The DTC has made enormous efforts to test and port the IBM GSI system to multiple platforms, which include Linux cluster and Linux workstation with both Ifort and PGI compilers and Mac system with PGI compiler. The DTC has been working with NCEP and other GSI development groups closely to fix the portability issues founded in the multiple platform tests. From these collaborations, the portability of the GSI code has been improved significantly. Now, the GSI trunk head can be easily installed and run on multiple platforms.

The key to an easy installation of a community system is a well-tested user-friendly compiling system on top of the code. The DTC has added and maintained such a compiling system on top of the GSI source code. Also, the DTC added all the external libraries needed by GSI except for the WRF system as a part of the community GSI package for the convenience of the users. After some simple computer environmental configuration steps, users can compile the community GSI with a single command. This command intrigues the compiling system to compile both external libraries and main source code of GSI to make a final executable. The available computer configurations include IBM supercomputer, Linux Ifort and PGI, Mac PGI, and SGI Ifort. The GSI installation is explained in details in the Chapter 2 of the GSI User's Guide.

To help users to run GSI smoothly, DTC created a sample GSI run script. For most of simple cases, users only need to set several computer and study case options at the head of the sample run script to the cases. The run script will create a run directory, copy or link all needed data into the run directory, create GSI namelist, and run GSI. This sample run script is provided as a part of the community GSI package in each release version. The GSI run scripts is explained in details in Chapter 3 of the GSI User's Guide.

The community GSI package also includes several GSI tools added by the DTC to help users processing BUFR/PrepBUFR data and diagnosing the GSI analysis results in their research. The DTC has created and enhanced these tools mostly based on the users' questions and requirements. For example, many of the users need to process BUFR/PrepBUFR files in their GSI application. The DTC therefore created a set of simple BUFR/PrepBUFR processing examples for users to learn and start such kind of work. These BUFR tools in the new release version 3 have been significantly enhanced based on the questions asked by the users in the applications of these tools since the release 2. The DTC also added a new Chapter in the User's Guide to explain how to apply these BUFR/PrepBUFR tools in detail. For other community tools, please refer the GSI User's Guide Appendix A for more details.

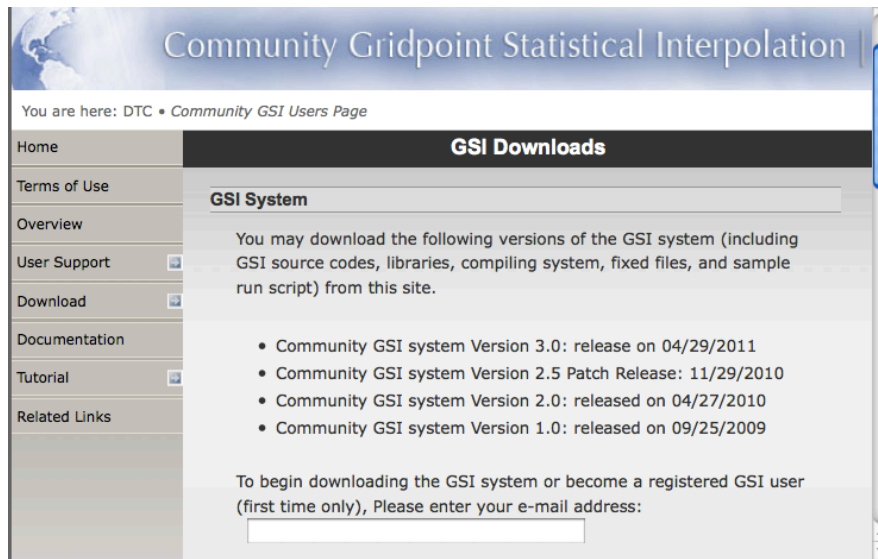


Fig.2 Web Page for GSI package download

Since September 2009, the DTC has officially released three version of the community GSI package with the corresponding User’s Guide. The release history can be seen from the GSI Users Page under Download/GSI System page (Fig.2). As we introduced before, each release version includes well-tested and documented GSI source code, libraries, compiling system, GSI run scripts, and community tools. As you can see from the download webpage, we do need users to register before download the package. Right now, there are about 360 registered users across the world.

- Write GSI User’s Guide and provide other documentation

As a key part of the community system, the GSI User’s Guide is written based on the release community GSI package to provide important guidance and information for users to apply GSI in their researches. The DTC has worked with NCEP and other GSI development groups closely since the very beginning of the community GSI project to write and update this GSI User’s Guide. The User’s Guide has been released with the each official release version of the community GSI package and the latest version is for release 3 in April, 2011. The GSI User’s Guide version 3 includes 8 chapters and 2 appendixes to cover the history and current status, GSI installation, running, diagnostics, theory and code structure, applications, observation and background error, BUFR/PrepBUFR, tools, and complete namelist options. Users can find all three versions of the GSI User’s Guide from GSI users Page under “Documentation”.

In the same webpage, users can also find the slides from the lectures in the 2010 Summer GSI Tutorial. These lectures were given by the GSI experts from NCEP, GMAO, MMM, and DTC to cover both basic GSI application skill and advance scientific topics related to GSI and data assimilation. User can also find other documentation like the slides from the presentations on the DTC GSI support given by DTC staff in several conferences, the GSI code Browsers, and reference list of important published GSI papers.

- Provide training through online and on-site tutorial

Before start to apply GSI in researches, we suggest users to go through an online tutorial in the GSI Users Page to test and get familiar with the GSI system. The DTC has created several data analysis cases with details instructions as the on-line tutorial for each release version to help users start. The on-site tutorial hands-on practical cases were also posted as a part of the on-line tutorial after the tutorial to give users more cases to practice.

The first GSI on-site Summer Tutorial was hold from June 28-30, 2010. This tutorial was a joint effort of NCEP, MMM, GMAO, GSD, and DTC. There were 13 lectures covering not only the basic skills of GSI application but also many advanced scientific topics such as radiance data assimilation, background error covariance, GPS data assimilation, radar data assimilation, and GSI based 4DVar. The DTC also designed and prepared 6 hands-on practical cases to give students some first-hand experiences of using GSI in several kinds of applications. We are now preparing the second GSI Summer Tutorial, which will be hold from June 29 to July 1, 2011. Both lectures and practical cases have been improved based on the first tutorial. Users are encouraged to attend the on-site GSI tutorial to interact with the experienced GSI experts.

3. From Research to Operation (R2O)

As a bridge between the operational and research data assimilation community, DTC also make enormous efforts to coordinate the GSI development, built and maintains the community GSI repository to facilitate the GSI development in research community, and work with research community developers to test and commit the new code back to the GSI repository.

- GSI review committee

In recent several years, several new developers groups, such as NOAA/ESRL, NCAR/MMM, AFWA, joined in the GSI development community. With the community GSI code available to the research community, many researchers showed interests to study their new data assimilation ideas based on the GSI system. To steer new developments and avoid conflicts in the GSI development, a GSI review committee was formed to coordinate distributed development of the GSI system in late 2010. The committee is composed of primary GSI research and operational centers, including NCEP/EMC, NASA/GMAO, NOAA/ESRL, NCAR/MMM, AFWA, and DTC. As one of the committee members, the DTC represents the research community with GSI users from all over the world.

The GSI Review Committee primarily steers distributed GSI development and community code management and support. The responsibilities of the Review Committee are divided into two major aspects: coordination and code review. The purpose and guiding principles of the Review Committee are as follows:

Coordination and Advisory

- Propose and shepherd new development
- Coordinate on-going and new development
- Process management
- Community support recommendation

GSI Code Review

- Establish and manage a unified GSI coding standard followed by all GSI developers.
- Establish and manage a process for proposal and commitment of new developments to the GSI repository.
- Review proposed modifications to the code trunk.
- Make a decision on whether code change proposals are accepted or denied for inclusion in the repository and manage the repository.
- Oversee the timely testing and inclusion of code into the repository.

The review committee is committed to facilitating transition from research to operations (R2O). Prospective contributors of code should contact the DTC through the GSI helpdesk (gsi_help@ucar.edu). The DTC will help the prospective contributors go through a scientific review from the GSI Review Committee to avoid any potential conflict in code development. Upon the approval of the GSI review committee, the DTC will work with the prospective contributors in the preparation and integration of their code into the GSI repository (Fig. 3).

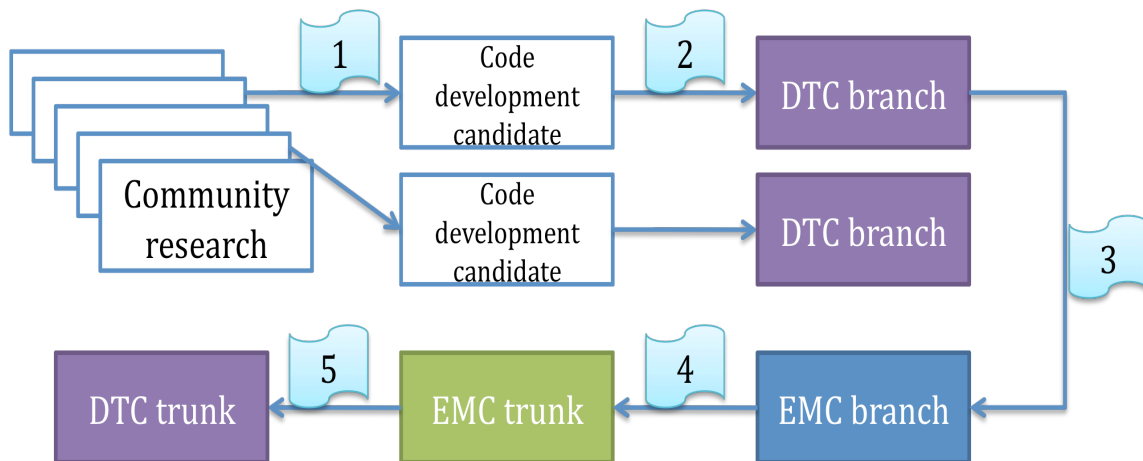


Fig.3. Illustration of DTC's community code management structure for the GSI data assimilation system to share code between the NOAA Environmental Modeling Center (EMC) and the community. (1) GSI Review Committee provides initial scientific review; (2) Developer code merging and testing; (3) DTC code ticketing and testing; (4) GSI Review Committee code review and DTC->EMC commitment; (5) EMC->DTC GSI repository syncing.

- Community GSI repository

Since 2009, EMC have put all its operational systems under subversion code manage system. The operational GSI repository in EMC has been actively used by NCEP and GMAO to track and share the GSI code development since then. At the same time, the DTC built a community GSI repository to support the community GSI developers using the latest GSI code in the research. In figure 4, we illustrated the purpose and relation of these two GSI repositories: the EMC GSI repository is mainly used by GSI developers in EMC and GMAO for the operational development, while the community GSI repository is mainly used by community GSI developers such as NOAA/GSD and NCAR/MMM for the research or operation development in a computer environment other than EMC computers. The community GSI repository also provides the code manage system to hold the official released GSI packages.

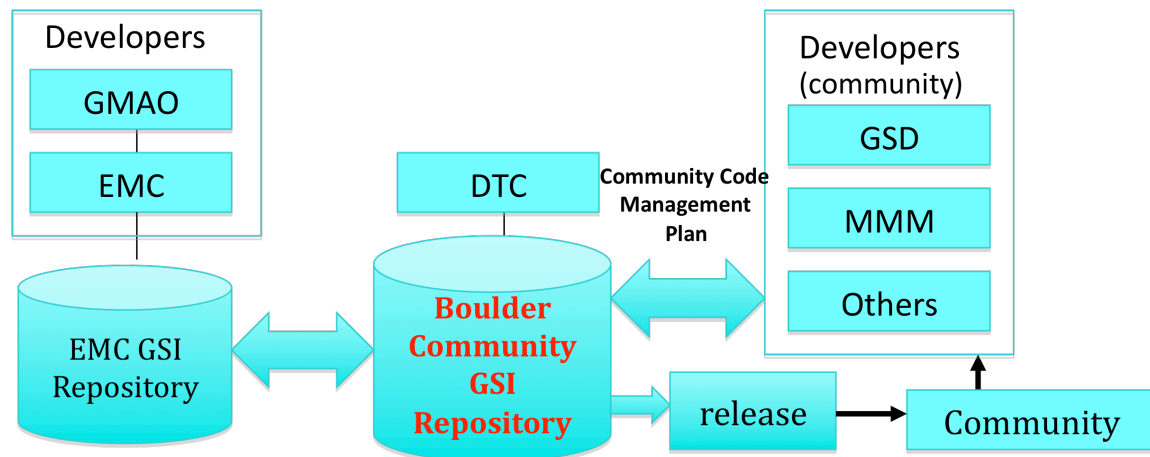


Fig. 4. The function of the EMC GSI repository and community GSI repository

The figure 5 illustrates the structure and contents of the two GSI repositories. Here, we can see another major difference between the two repositories: the community repository includes several components that are not in the EMC repository. As we introduced before in section 2 on the release of community GSI package, these components such as external libraries, compile system, and utilities are part of the released community GSI package. DTC created and maintain these components to make the community GSI system can be easily installed on multiple platforms other than the EMC computer system.

Although the two GSI repositories are maintained by the EMC and DTC separately to serve the operational and research community, the core part of the repository, the GSI source code, is always identical in the two repositories. As indicated in the figure 3 and figure 5, the DTC keeps this identical GSI source code by syncing the GSI source code from the EMC repository (./src directory) to the community repository (./src/main directory) weekly. When bug fixes or new codes from the DTC and research community are ready to commit to the GSI trunk, the DTC always work with the EMC GSI group to

commit the changes back to the EMC GSI trunk first. Then, the changes from the DTC and research community will be synced back to the community GSI trunk after they are in the EMC trunk. In this way, both operational and research community are actually doing their GSI development based on the same code.

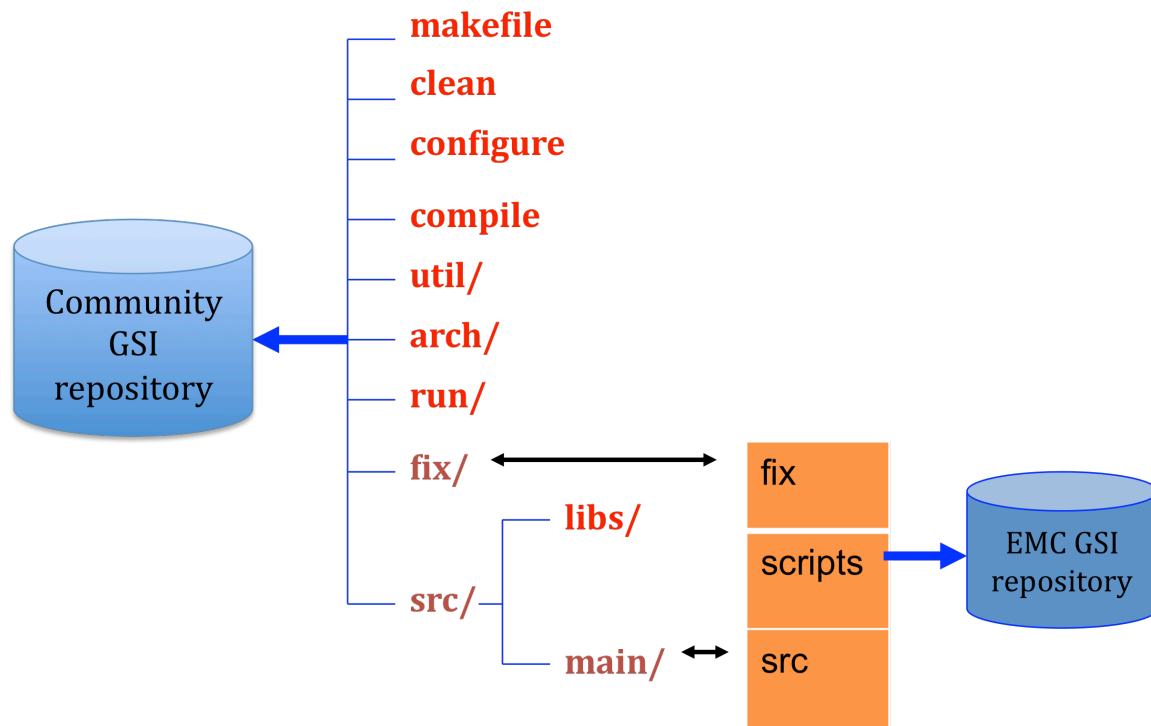


Fig.5 The structure of the EMC GSI repository and community GSI repository

As described in the GSI code management (figure 3), all the new GSI development to the GSI trunk need to be reviewed by the review committee. After the committee approves the new development, DTC staff will work with the community GSI developers to tests the new changes on multiple platforms with simple cases to catch the portability and other issues before send code to EMC for a code review by the committee. So, the research community can easily conduct the research with the community GSI trunk on a computer platform different from the operation ones but still get new development back to the GSI trunk smoothly.

- GSI workshop and monthly meeting

As a part of coordinate efforts, the DTC is organizing the first GSI workshop at NCAR on June 28 2011 to provide a pathway between operational centers and research community to communicate and share experience on:

1. GSI development, implementation and future plans;
2. New techniques and skills in data assimilation.

The workshop will include a few invited talks from major operational and research centers as well as a general section for community GSI and other data assimilation system users.

DTC also hosts a monthly community GSI telecom meeting to provide a chance for all GSI developers and users to discuss their GSI related works. Any users are welcome to join the meeting through phone. Please email GSI helpdesk for the instructions of how to join the meeting.

4. Summary and Plan

In the past several years, the DTC have been making significant efforts not only to bring the operational GSI system to the research community but also to support the research community to participate the GSI development and commit their contributions back to the operation GSI trunk. The interested GSI users are welcome to contact the GSI help desk for further information on this bridge role of the DTC.

Please note the DTC currently focuses on testing and evaluation of GSI for limited area Numerical Weather Prediction (NWP) applications, however, the long-term plan includes transitioning to global forecast applications. Currently, the GSI v3.0 is still a three-dimensional variational (3D-Var) system with some modules developed for the upcoming four-dimensional variational (4D-Var) and observation sensitivity capabilities. Combined with an ensemble system, this GSI can also be used in an ensemble-variational hybrid data assimilation system with appropriate configuration. The DTC will follow up the operational development and application to support the community users on these new functions in the future based on the processing of the development and available support resources.