Summary Report of the 2021 Science Review of the NOAA Global Systems Laboratory

May 10 - 13, 2021

Review Panel:
Shaima Nasiri (Department of Energy), Chair
Chris Bretherton (Vulcan, Inc., and University of Washington)
Gregory Carmichael (University of Iowa)
Anke Kamrath (National Center for Atmospheric Research)
Humphrey Lean (UK Met Office)
Vernon Morris (Arizona State University)
Gabriele Pfister (National Center for Atmospheric Research)
Danny Sims (Federal Aviation Administration)
John Williams (The Weather Company, an IBM business)

1 Overview

Laboratory science reviews are conducted every five years to evaluate the quality, relevance, and performance of research conducted in the National Oceanic and Atmospheric Administration (NOAA) Office of Oceanic and Atmospheric Research (OAR) laboratories. The reviews are used internally by NOAA/OAR for planning, programming, and budgeting, and external interests, and by the Laboratory in strategic planning of future research directions. These reviews are also intended to ensure that OAR laboratory research is aligned with NOAA's research mission and priorities and other relevant strategic plans, is of high quality as judged by preeminence criteria, and is carried out with a high level of performance.

This report covers the research of the Global Systems Laboratory (GSL) over a period of approximately five and a half years (November 2015 through May 2021). The members of the review panel participated in a virtual review May 10-13, 2021. Prior to the review, the reviewers participated in two pre-review videoconferences with OAR and GSL leadership and support staff that helped clarify the charge to reviewers, explain the review process, and make minor changes to the review agenda. The pre-review videoconferences provided the reviewers the opportunity to comment on and make suggestions regarding the specific questions to stakeholders and request that the stakeholder pool be broadened. Throughout the pre-review and review process, OAR and GSL leadership were responsive to reviewer concerns and requests.

The review panel was charged with evaluating the Quality, Relevance, and Performance of GSL in four areas: Organizational Excellence, Advanced Technologies, Earth System Prediction, and Decision Support. Each reviewer was asked to independently prepare his or her written evaluations of one or more research areas. The Chair, a Federal employee, was tasked with creating a report summarizing the individual evaluations. For this report, the Chair also drew upon the summary documents prepared by the panelists in preparation for panel report-out/recommendations provided at the end of the virtual review on May 13 to OAR and GSL leadership. In accordance with Federal Advisory Committee Act (FACA) rules, the review panel did not seek consensus in its evaluations. However, there was substantial agreement on the general findings as well as in the specific findings in each of the four areas presented below. Each of the reviewers was given the opportunity to review and comment on the final report before it was submitted to NOAA.

2 Summary of Laboratory-Wide Findings and Recommendations

GSL is a recognized leader in Advanced Technologies, Earth System Prediction, and Decision Support. GSL designs models and software tools that operational weather forecasters, the research community, and the broader public find useful. GSL's activities clearly benefit society and are aligned with NOAA's mission. GSL has dedicated and forward-looking leadership as well as engaged, qualified, and productive staff. As an OAR laboratory, GSL *Exceeds Expectations* and is outstanding in many areas.

The likelihood of continued and future success is high. However, there are things that GSL, OAR, and NOAA can do to increase the probability of GSL's future success. While GSL's leadership should be commended for their success in weaving together diverse funding mechanisms, many from outside NOAA, the current high concentration of funding on short-term projects (1 to 2-year) has far-reaching limitations and it will be difficult for GSL to achieve its stated goal of increasing the focus on research without having longer term funding. The benefits of a better balance between longer-term projects and short-term projects would be felt across many facets of the organization including publication metrics and active external collaborations and would better enable GSL to engage in higher risk, yet potentially revolutionary, R&D.

Discussions with and feedback from stakeholders demonstrated that GSL's role within OAR is not always clear and there are perceptions of overlap and duplication of effort with other OAR labs. The scales of GSL's new Grand Challenge, NOAA's plans for the Unified Forecasting System, as well as the societal importance of NOAA's mission will all require close coordination and collaboration across OAR labs and the National Weather Service (NWS). It should not fall on GSL alone to define its roles and responsibilities, to coordinate collaboration, or to avoid duplication across NOAA; NOAA leadership and direction will be necessary here. GSL would benefit from better branding and promotion of its intellectual leadership and strengths, role in NOAA's mission, and services and outreach to the general public.

	Highest Performance	Exceeds Expectations	Satisfactory
Organizational Excellence	1 Quality	**2 Overall** 2 Relevance 1 Performance	1 Quality 1 Performance
Advanced Technologies	**1 Overall** 1 Quality 1 Relevance	**2 Overall** 2 Quality 1 Relevance 3 Performance	1 Relevance
Earth System Modeling	2 Quality 1 Relevance 1 Performance	**5 Overall** 3 Quality 4 Relevance 3 Performance	1 Performance
Decision Support	**1 Overall** 1 Relevance 1 Performance	**1 Overall** 2 Quality 1 Relevance	1 Performance

The table summarizes the overall ratings and quality, relevance, and performance ratings from the individual panelist evaluation forms for each of the four themes. The lowest tier rating of "Needs Improvement" was not utilized by any of the panelists hence the rating is not included in the table. Note that individual thematic evaluation forms often included comments that spanned multiple themes or the laboratory as a whole.

3 Findings and Recommendations by Research Area

3.1 Organizational Excellence

GSL has a dedicated, competent, and creative leadership team. The blended federal, Cooperative Institute, and contractor workforce is engaged in the day-to-day operations as well as strategic planning of the lab. The GSL leadership team has had the foresight and flexibility to find solutions that have enabled GSL to continue to be productive and collaborative during the disruptions of the 2020 and 2021 global pandemic. The high satisfaction levels of GSL's staff are strong indicators of GSL's culture of inclusion. Overall, GSL *exceeds expectations* in the thematic area of Organizational Excellence.

3.1.1 Quality

GSL conducts preeminent research, and its scientific products and technological products are meritorious and significant contributions to the scientific community. However, publication numbers (average of 42 peer-reviewed publications/journal articles per year and 54 in 2020 for an organization with 180 staff, 50 of whom are classified as scientists) seem low, even compared with peer organizations focused on research to operations. Several stakeholders mentioned low publication numbers, as well as difficulty finding out what GSL researchers are working on without downloading conference PowerPoint slides. This can be a perennial feature in organizations where delivery of models and products is often viewed as a higher priority than writing papers. However, since GSL is aspiring to move further towards the research end of the "funnel" it would be good to try to encourage staff to produce more publications and to give them the time and resources to do so. Additionally, if GSL wants to prioritize transitions, it should develop a methodology for tracking, documenting, and publicizing transitions; crediting them to creators; and highlighting them at the same level, and in a similar way, as journal articles. Elevating the profile of transition and tracking them will make a case to GSL's peer institutions (e.g., other NOAA as well as other agency labs) that GSL's scientists and software developers are at their same level. Awards numbers and service to organizations, journals, and societies seem satisfactory although it appears that a subset of staff members perform the bulk of professional and journal service. There was no mention of service on proposal review panels during the review or in the accompanying documentation; if lab staff are not regularly serving as proposal reviewers, GSL should consider encouraging them to do so.

GSL exceeds expectations in terms of collaboration with groups within and outside of NOAA. GSL leadership maintains alignment with NOAA Science & Technology strategic goals and leverages partnerships with the cooperative institutes to catalyze efforts to advance high quality work and expand leadership and career pathways for new talent. The collaborations with NWS, Federal Aviation Administration (FAA), Joint Center for Satellite Data Assimilation (JCSDA), and Developmental Testbed Center (DTC) are especially noteworthy and are examples where GSL is contributing intellectual leadership as well as support. For the most part, the stakeholder questionnaires and meetings reinforce this evaluation. However, GSL is encouraged to expand academic and NOAA Center partnerships beyond the immediate locale, raise its international profile, and increase its number of international collaborations. GSL clearly excels at research to operations transitions, however, these are more difficult to keep track of and understand. E.g., it was necessary to expand the "Research to Operations Transitions" document from the web site to full screen on a large monitor to see it in its entirety. As

mentioned above, GSL is encouraged to think creatively about how to make these transitions available, findable, and understandable.

3.1.2 Relevance

It is clear from the review documents, presentations, and stakeholder meetings that GSL designs models and software tools that address societally relevant needs. In general, stakeholders had positive comments about working with GSL. Several stakeholders had suggestions for additional things that GSL could do. This can be seen as a good sign because it shows that the stakeholders want to continue working with GSL and feel that GSL is capable and does good work. Here are a few takeaways from the stakeholder engagement surveys and interviews:

- Stakeholders had very positive comments about GSL leadership.
- The strength and value of partnerships with NWS really came through in the stakeholder responses.
- Stakeholders had very positive comments about the quality of GSL's work, as well as rapid response to stakeholder needs, collaborative relationships, continual improvement of working relationships, fluid and effective communications, and cost-effective measures to produce results.
- Several stakeholders felt that there were areas of duplication between GSL and other NOAA labs including the Geophysical Fluid Dynamics Laboratory (GFDL) and the National Severe Storm Laboratory (NSSL) and that there was not a clear path for collaboration in some of the model development areas.
- Depending on the stakeholder, engagement and communication were viewed as satisfactory in some areas and exceeding expectations in others.

The review presentations and documents demonstrated that public engagement is important to many GSL staff members and that they take the opportunity to participate in the public sphere through outreach activities. In addition, GSL's public engagement through Science on a Sphere is very good. However, one reviewer commented that while they had known about Science on a Sphere for ages, they were unaware of GSL's role in it. GSL could raise its public profile through more effective branding and by showcasing its visualizations and model development on their webpage in a form that is readily consumable by the general public.

GSL could do more for workforce development by increasing the number of postdoctoral research associates that it supports and by extending training and mentoring at all career stages more uniformly through GSL. For example, GSL could consider extending the successful peer-to-peer mentoring program at the Cooperative Institute for Research in Environmental Sciences (CIRES) to include Cooperative Institute for Research in the Atmosphere (CIRA) and other federal employees. Because of the strong national need to grow the workforce of software developers who understand and are interested in numerical weather prediction and earth system models, GSL might consider starting a post-masters software development internship program. To be successful, such an internship program would need to focus on helping recent graduates in atmospheric sciences and related fields develop new skills and thrive in a software development and research-to-operations environment. Together with a recruitment strategy that extends beyond a handful of institutions, an internal focus on workforce development and training may help GSL meet its diversity goals.

3.1.3 Performance

GSL's leadership team is very strong and appears to have had the flexibility to take advantage of new opportunities and shift into new directions (e.g., fire weather and smoke forecasting). The teamwork demonstrated by GSL's management and the shared respect between GSL management and employees are impressive. GSL's new draft strategic plan and grand challenge appear to have been adopted and internalized by lab staff. Overall, GSL is organized and managed effectively for R&D. Annual strategic retreats are open to all staff; weekly all-lab meetings during the pandemic have been valuable (and at least some staff hope they will continue); awards for team member of the month are appreciated; and staff report good work-life balance. GSL's leadership is seen by staff and stakeholders as being responsive, collaborative, and willing to listen, and as a result is deeply respected. Another important point is that GSL has been willing to take on high-risk initiatives even though the initiatives have not always been successful or had an immediate path to operations.

Based on GSL's Final Report on Actions available from the GSL Science review web page, GSL appears to have done its due diligence with respect to implementation of recommendations from the previous laboratory review. However, this was not discussed very much within the review itself and there would have been benefit in hearing more about the challenges of implementation. In addition, there have been significant changes to GSL and to forecasting systems and technology since the review of GSD in 2015, so it is difficult to comment fully on the extent of the implementation since the recommendations are not all equally relevant now.

While NOAA's definition for diversity is broad, "Diversity is the mixture of the unique attributes that shape an individual's identity which they bring into the workplace to help NOAA accomplish its goals. Diversity refers to demographic diversity (e.g., race, gender, sexual orientation), experiential diversity (e.g., affinities, hobbies, and abilities), and cognitive diversity (e.g., sensory processing and problem solving)", GSL's presentation and materials focused only on demographic diversity. GSL leadership appears to be committed to improving workforce diversity; however, the progress towards increasing laboratory demographic diversity since the last review was limited. The review panel recognizes that low demographic diversity within the atmospheric sciences is a systemic challenge and one that cannot be solved at the level of an individual laboratory. Equity, diversity, and inclusion need to become embedded ethics within all STEM communities. Accomplishing this will require strategies for implementation at the NOAA level, as well as within each of the OAR laboratories. Establishing a fully inclusive workforce requires a comprehensive plan to make progress and NOAA and OAR need to provide resources and work with the labs to develop, implement, and evaluate the success of these plans. Common elements may include broader accountability measures in the organization assessment; support for affinity groups; flexibility for employees to do remote work; enhanced visitor programs, and internship programs (e.g., internship "for credit" programs with minority serving institutions), however no single element will be sufficient. Instead of trying to create a program by re-inventing the wheel or mimicking the square wheels of other organizations that are struggling with the challenges of diversity, equity, and inclusion, labs should partner with individuals or organizations to develop a sustainable strategy beyond summer programs and one-off internships. Better utilizing the NOAA Cooperative Science Centers may help GSL and OAR develop a robust pipeline of diverse employees.

GSL demonstrated that it considers NOAA and OAR priorities when developing its strategic plan and vision; however, the overall coordination between and among GSL, other OAR laboratories, and OAR leadership, as well as the process for prioritizing new projects was not always clear. Additionally, there were few mentions by GSL staff, other than lab leadership, of connections to OAR's Weather Portfolio which, along with stakeholder comments about duplication between NOAA laboratories, raised questions about the effectiveness of OAR's portfolio approach. It does appear that GSL makes effective use of NOAA's Research and Development Database and program management table.

One of the greatest challenges to GSL's ability to do long-term planning, make commitments, and engage in exploratory, but potentially transformative, research and development is lack of dedicated funding and the limitations of short funding cycles. GSL has managed to take advantage of base funds to anticipate needs and directions and to develop visionary proofs-of-concepts. While some have gone on to be supported by stakeholders, others have been more difficult to sustain long-term. GSL leadership needs to carefully balance the impacts of short-term, non-core projects on GSL's portfolio against the limited amount of base funding and longer-term, strategic investments. While the lab would clearly benefit from more base funding and consistent direction, GSL's diverse efforts nevertheless seem to be woven into a coherent story, with external funding enhancing the lab's ability to serve its core mission.

The GSL leadership is encouraged to develop a strategy for branding the lab and its excellence. Whether the fact that stakeholders are not always clear which Lab within OAR is doing what is the fault of the stakeholder or the fault of the Lab, the impact is the same: less credit for successes. As a brand, "Global Systems Laboratory" is not very descriptive. While a name change may be prohibitively difficult, some skillful marketing to enhance GSL's name recognition may be worthwhile. A branding strategy could include ensuring that your websites reflect the high quality of GSL's work and the great value of GSL's products as well as prioritizing sharing of research results, software, documentation, and products in more referenceable/findable outlets than conference proceedings/presentations.

3.2 Advanced Technologies

GSL is a recognized leader within NOAA in Advanced Technologies including data systems, high performance computing, machine learning, cloud computing and visualization. GSL's R&D in Advanced Technologies supports the end-to-end capabilities needed to improve weather predictions. GSL's contributions have been recognized both nationally and internationally. In addition, the GSL Information Technology (IT) team operates a reliable and quality environment that is generally meeting the needs of the staff. In the area of Advanced Technologies, GSL *exceeds expectations* and is outstanding in many areas.

3.2.1 Quality

GSL conducts high-quality research and development of Advanced Technologies to advance NOAA and the US weather enterprise that is outstanding in many areas. Lab scientists and engineers are widely recognized as pathfinders within NOAA for IT, computing, and modeling; are involved in multiple external committees and boards; present their work at conferences and workshops; and publish their

research results. However, as in other areas, the number of publications appears to lag some peer institutions of similar size.

GSL has made significant and impressive progress in developing, enhancing, and delivering high quality products. One example is the Meteorological Assimilation Data Ingest System (MADIS) which has had four major releases during the past four years. Following the recommendations of the previous review panel, GSL has added new data systems and data sets and has established a MADIS data and metadata standard. Another example is Science on a Sphere (SOS) which continues to provide an inspiring science communication and support tool that has a broad installation base including 77 exhibit installations across 23 countries. The SOS Explorer (SOSx) has seen over 25,000 downloads and provides an important option beyond the SOS exhibit installations. The uptake and downloads of these and other GSL products, including the Advanced Quantitative Precipitation Information (AQPI) system, are impressive and speak to their relevance and quality.

The breadth of GSL's work in the thematic area of Advanced Technologies should be noted. GSL is leading NOAA in exploring the use of cloud technologies and services to augment traditional supercomputing for "overflow," intermittent or collaborative workloads. GSL is also deeply involved in shaping NOAA's strategy for adopting artificial intelligence and machine learning, organizing NOAA-sponsored workshops and collaborating in an NSF artificial intelligence institute, and is exploring ways to exploit artificial intelligence and machine learning (ML/AI) throughout the lab's portfolio. The current GSL computing infrastructure appears to be very well run, with a good balance between security and useability. In addition, GSL is also responsible for the proper facilitation, support, coordination and system reporting for a substantial fraction of NOAA's computing infrastructure at other locations (e. g., Orion in Mississippi).

3.2.2 Relevance

GSL's R&D in Advanced Technologies is deeply relevant to NOAA's mission and uniquely aids innovation throughout OAR, the NWS, and beyond. The technologies that GSL applies and develops are essential to making more accurate forecasts, extracting information, and providing better products to forecasters and the public. GSL's Advanced Technology thematic area has a very strong and relevant portfolio and is held in high regard by GSL's stakeholders. The focus on research to operations and delivering transitional products and solutions to GSL's customers is clear. In general, panelists felt that GSL's relevance is currently outstanding in many areas, but they did raise a few areas where planning and forward thinking will be necessary for GSL to maintain a high level of relevance into the future. A few of those areas are discussed below.

GSL continues to be forward-looking and at the forefront of high-performance computing (HPC)-related R&D. However, more funding for HPC and a broader technology portfolio would be beneficial. The Graphics Processing Unit (GPU) strategy around performance-portability is appropriate given the rapid technology shifts that are on the horizon. Regarding the future of conventional HPC and GPUs, one panelist questioned whether GSL's concerns regarding the Finite Volume Cubed Sphere model (FV3) and decision to focus efforts on the more GPU-friendly dycore was premature. The panelist noted that openACC, a standard for directive-based parallelization of codes, is not the only approach that could be

applied to speed up the FV3 dycore on GPUs and that the NOAA community is heavily invested in FV3 for the next decade.

GSL is working on novel cloud-computing and GPU applications that complement its traditional computing infrastructure, and its work in this area is relevant and well-considered. Currently, GSL's cloud computing effort is relatively small and focused on applications that complement GSL's own HPC. Panelists did raise questions about GSL's cloud strategy for the future while noting it will be important to manage the pressures to do more and more on the cloud and balance them with the (often) higher costs of operating in the cloud. The panel heard from stakeholders outside of NOAA who noted that it can be difficult to get approvals to work within the NOAA HPC firewall. For example, this can make it difficult for academic research groups that include foreign nationals to collaborate with GSL scientists. Many potential community users/developers of NOAA models such as the UFS are stymied by the challenges of the NOAA HPC firewall. While there may not be clear cost benefits of cloud computing to GSL's scientists now, GSL is encouraged to consider the potential benefit to the Unified Forecast System (UFS) community. The type of technology development that would be necessary to enable UFS to be run on cloud computing is something at which GSL excels. GSL should look for opportunities to co-design modeling initiatives with the National Centers for Environmental Prediction (NCEP), including Earth Prediction Innovation Center (EPIC), to fully realize the potential of new advanced technologies including GPUs and ML/AI.

Additional areas of importance and relevance in Advanced Technologies include preparing models for exascale, incorporating ML/AI in the many aspects of GSL's workflow, handling larger and larger data challenges, incorporating both on and off-premises cloud solutions, and tackling the end-to-end workflow challenges that come with all these technologies. GSL is currently successfully balancing the competing priorities in these areas, however increased funding will be needed to tackle more of these growing areas to maintain GSL's relevance to NOAA/OAR and to support the next level of problems in the sub-seasonal to seasonal forecasting domain. For example, a potential missing area to consider is big data end-to-end workflows. GSL has many workflow elements within their organization, but the integration between these and optimizing across them was not evident. The challenges around Big Data workflows will require a holistic view to tackle the challenges ahead in GSL's customer and collaboration portfolio.

3.2.3 Performance

In the area of Advanced Technologies, GSL performs at a consistently high level with a surprisingly modest internal computing budget. However, as with the other areas, shorter-term, non-core funding appears to have a strong influence on the portfolio of GSL advanced technology efforts. It is especially important for Advanced Technologies to have longer-term, higher-risk elements in its portfolio, and to make investments from a strategic organizational perspective. Sometimes efforts will fail, as with the initial GPU port of the FV3 dycore, but it is imperative that the organization have the flexibility to take risks. How much more science and research to operations could GSL do with more access to computational resources and longer-term funding and how would those projects be prioritized? There are a lot of options for future GSL investment, but also a need for them to be strategic. One clear area of opportunity is in developing strategic partnerships. One suggestion that came up during the stakeholder

discussions was that a regular (e.g., quarterly) portfolio/opportunity review would provide an opportunity to share collaborative opportunities and help make connections.

It should be noted that GSL's advanced technology successes, such as GSL's development and transition to operations of the High-Resolution Rapid Refresh model (HRRR) and the documentation and public release of the Gridpoint Statistical Interpolation (GSI) data assimilation system, far outnumber the failures. In general, GSL's long-term vision in Advanced Technologies is appropriate. In addition, the continuity of vision across the GSL leadership changes since the previous review is good to see and seems to be well supported by the leadership team and laboratory staff. The Advanced Technologies area appears to have done a good job of addressing the recommendations in the 2015 report. Appropriately, some of the 2015 recommendations continue to be an ongoing focus.

3.3 Earth System Prediction

GSL is one of the leading players in the world in developing rapidly updating regional models and transitioning them to operations. Notable achievements in Earth System Prediction include development of an hourly-updating convection-permitting model, the HRRR; extension of HRRR capabilities to include application to smoke and renewable energy; and important contributions to the premier US community atmospheric chemical transport model, the Weather Research and Forecasting model with inline chemistry (WRF-Chem). Overall, GSL's work in the area of Earth System Prediction *exceeds expectations* and is outstanding in many areas.

3.3.1 Quality

GSL has many talented staff members working on parameterization development, software engineering supporting community modeling (e.g., the Common Community Physics Package or CCPP effort), and visualization and interpretation tools that help stakeholders understand simulation results. The work of individual scientists, as presented during the review, was of very high quality and covered significant breadth. It should be noted that the HRRR capability developed by GSL is state-of-the-art, and GSL has demonstrated its commitment to extending its capabilities, including the development of the HRRR ensemble, in part in response to the previous review. The forecast improvements over time for different HRRR implementations are notable. In addition, in their written input and discussions with the review panelists, stakeholders were very positive about the quality of GSL's R&D efforts in Earth System Prediction. Additionally, there is considerable planetary boundary layer expertise within GSL and strong contributions to this component of Earth system modeling are being made.

GSL plays an important role within NOAA in advancing the atmospheric composition component of the Earth system modeling framework. WRF-Chem enables high quality science and the success of GSL's contributions to it is notable, even with the limited funding commitment from NOAA in this area. GSL's development of model parameterizations at the nexus between aerosols and physics is another area where GSL scientists have excelled and received recognition. The Global Ensemble Forecast System (GEFS Aerosols model upgrade is a valuable contribution to an enhanced global model capability. It is likely that the need for a predictive capability, as well as applications and services, for atmospheric composition will continue to grow, especially related to forecasting of smoke and fire, air quality, ecosystems, and human

impacts. However, there does not seem to be a clear vision across NOAA regarding atmospheric composition. While NOAA plans to establish geostationary satellite capabilities for atmospheric composition as a follow-on to the National Aeronautics and Space Administration's (NASA's) Tropospheric Emissions: Monitoring Pollution (TEMPO) mission, the review panel was not aware of plans to incorporate the capabilities of HRRR-Smoke or atmospheric chemistry into the UFS. In addition, the relationship of GSL's atmospheric composition work to NOAA's operational air quality forecasting system was not clear. Atmospheric composition and chemistry are areas where strategic partnerships within NOAA, as well as outside with outside groups including National Center for Atmospheric Research (NCAR), NASA, the Department of Energy (DOE), and academic institutions, will likely be necessary. With strategic planning and long-term, dedicated funding, GSL could play an important role in this area in the future.

While GSL has several scientists with strong publication records, outstanding numbers of citations, and both US and international recognition, publication numbers appear to be on the low side, even in the area of Earth System Prediction. Much of the model development, especially focused on global modeling and associated parameterization development, is presented in talks rather than journal articles. There are clearly strong collaborations across NOAA, e.g., with the Hazardous Weather Testbed. GSL should consider expanding those collaborations with other testbeds, including the Aviation Weather Testbed. However, compared with similar international organizations, the number of international collaborations seems low. Increasing the number of international collaborations could bring GSL the added benefit of seeing how the models perform and are perceived in different areas of the world and could result in increased model enhancements.

3.3.2 Relevance

The contributions of GSL to weather and air quality/smoke forecasting are clearly relevant to society, are aligned with NOAA's mission and OAR and NWS objectives, and are also highly valuable for the research community. There is strong evidence of GSL working closely with both the operational and research organizations that make use of GSL models. Stakeholders were generally of the view that GSL works effectively with them and is responsive and agile in addressing their needs. All research appears to be driven by the needs of NOAA or those of other stakeholders, including FAA and DOE. While it does not appear that any of GSL's research is not relevant, at times it appears that the NOAA-driven research is left to the discretion of GSL. For example, management guidance from higher levels in NOAA might help the lab identify a clear path for integration of GSL's global-scale seasonal-forecasting R&D into UFS and the NWS Environmental Modeling Center. There has been considerable development in response to direct needs and requirements of FAA and DOE stakeholders. Both the FAA and DOE appear to be engaged to ensure needs are met.

For GSL to maintain its high degree of relevance, it will need to anticipate future opportunities and stakeholder needs. GSL's involvement in the NOAA Boundary Layer and Fire Weather Cross-Lab initiatives will be helpful for ensuring that GSL remains a major player in smoke and high resolution predictions. The collaborative focus on the planetary boundary level is welcome both in terms of improving overall forecasts and improving the representation of phenomena at low levels (e.g., low level winds for wind power). There are also opportunities for expanding on this to focus on the representation of the urban surface and boundary layer since urban applications are becoming more relevant with higher

resolution models that can start to resolve cities. The increasing importance of urban issues was recently highlighted by the World Meteorological Organization (WMO) with their initiative on "Integrated Urban Hydrometeorological, Climate and Environmental Services" and some international modeling centers are viewing 100-m scale regional models as one of the next resolution challenges alongside global km scale models. GSL may want to consider taking an active part in the WMO 2024 Paris Olympics RDP if the timing does not conflict with other priority activities in the next few years.

3.3.3 Performance

In general, GSL's performance in the thematic area of Earth System Prediction *exceeds expectations* and is outstanding in many areas. While acknowledging the high level of performance, panelists focused on the challenges GSL may face in maintaining that level into the future.

The fact that the GSL was able to work collectively to identify a grand challenge is a strong indicator of effective laboratory leadership and planning. Throughout the review, GSL staff demonstrated their belief in and dedication to this challenge. The challenge of global rapidly updating storm-scale modeling is relevant both scientifically and from the point of view of applications. However, the challenge is very large- "global", in fact. Panelists had questions regarding whether this is a realistic challenge for GSL alone, or whether it will require substantial collaboration with other NOAA labs. If so, how would that collaboration be enabled? Panelists had questions about the steps that GSL plans to take, and the timeframe, to make substantive progress. One panelist raised concerns about whether a global focus would spread GSL's resources too thin and questioned whether it would be better to maintain superiority in regional storm-scale model development and contribute to global development as opportunities arise rather than attempt to achieve the same level of success in the global domain. Panelists noted that much of the planning for GSL's model development appears to happen in a bottom-up fashion and that individual scientists and groups have the freedom to get involved in projects that they feel will be useful. Undoubtedly this freedom contributes to GSL's high level of workplace satisfaction, but it does require care to ensure that the overall effort is well-coordinated and avoids holes and duplication. At the scale of GSL's grand challenge, it may be necessary for that coordination to occur at higher levels within NOAA with clearer guidance regarding the roles and responsibilities of the different labs and more formalized collaborations.

Another Earth System Prediction challenge that will require effective management is transitioning GSL's modeling success stories into the UFS. The transition of the HRRR/RAP to the RRFS in the UFS framework will be a key change for GSL and it is important to ensure that it is well managed and that there is awareness from a management point of view of how it will change the balance of activities within GSL. Ensuring that the additional capabilities of HRRR, such as HRRR-Smoke do not get left behind in the move to the UFS is also a challenge. Successful transitions will require sufficient resources, as well as clear project plans including timelines and milestones. The move to the FV3 dycore is likely to be a significant challenge, especially since GSL staff do not appear to be enthusiastic about FV3, but the move will be necessary for GSL to continue to have impact within NOAA on storm-scale to global-scale modeling. The HRRR and WRF-Chem physical parameterizations are tuned to a different dycore, and there will likely be initial discouraging performance and data assimilation degradation to overcome. Sufficient resources, frequent motivation, clear milestones, and oversight will be important to manage the transition.

3.4 Decision Support

GSL's program in impact-based decision support services, verification, and evaluation is of very high quality, is clearly relevant to NOAA's mission, and is outstanding in many areas. Stakeholders rely on GSL to be thought leaders in verification and decision support services and GSL demonstrated the lab's many impressive successes in this area. The deep involvement of social scientists in GSL's Decision Support research and development is impressive and commendable. GSL's R&D in Decision Support *exceeds expectations* and is outstanding in nearly all areas.

3.4.1 Quality

GSL conducts preeminent Decision Support research that plays a key role within NOAA and the US weather enterprise. This is enabled by the fact that GSL decision support scientists are leaders in their fields. GSL continues to build on its strong reputation. The addition of social science researchers since the last review has increased the quality of decision support developments by looking at how users actually use weather information in their decision-making processes. GSL leadership has found a way to weave together diverse sources of funding which has helped enable GSL to build strong collaborations outside of the lab. The fact that GSL is called on to lend its expertise to a wide array of projects, many competitively awarded, together with positive statements of support and stakeholder feedback, is a testament to the quality of GSL's contributions. For example, continued year-to-year funding from the FAA's Aviation Weather Program to support verification of FAA-funded weather products by GSL indicates a high level of satisfaction and a healthy long-term relationship.

As has been mentioned in other sections, some indicators of quality, like refereed publications, appear to lag peer institutions of similar size. However, GSL has a very successful track record in transitioning research to operations and providing high-quality international data to the weather enterprise, e.g., via MADIS. While GSL has been successful in developing decision support tools and placing them into operations for NWS forecasters, there is untapped opportunity for GSL to connect GSL's tools and methods to other users and decision makers, e.g., industry (including airlines and other forecast providers), additional international users, and emergency managers.

3.4.2 Relevance

GSL's research and development in the area of Decision Support is clearly relevant and contributes both directly and indirectly to NOAA's mission. GSL's work is particularly relevant to the OAR goals to "Make forecasts better: improve accuracy, precision and efficiency of forecasts and predictions to save lives and property and support a vibrant economy" and "Drive innovative science: cultivate and deliver mission-relevant research to lead the environmental science community." Surveys from and conversations with NOAA stakeholders demonstrated that GSL has led the way in recognizing needs or opportunities. In some cases, GSL has prototyped solutions before a stakeholder requirement was even articulated, as with Impact-based Decision Support Services (IDSS). In the area of Decision Support, GSL's innovations are essential in evaluating the suitability and value of forecasting technologies to a wide range of users and in providing information in a way that simplifies forecaster workflows and communicates the most

important information to end-users. For example, the Hazard Services Program builds on GSL's Advanced Weather Interactive Processing System (AWIPS) legacy and will play an essential role in unifying and advancing forecaster tools, while Weather Archive and Visualization Environment (WAVE) is paving the way to better communication of weather forecasts to users. Stakeholders provided very positive assessments of the value of GSL's verification assessments and decision support system prototyping and research-to-operations efforts.

While GSL work benefits a wide range of weather applications and customers, Decision Support is primarily NOAA-oriented with a focus on providing tools to NWS forecasters. Expanding the user base beyond NWS and NOAA is a real area of opportunity for GSL, but one that will require thoughtful resource planning and engagement with external stakeholders. It may also require new avenues for external funding.

For GSL to continue to demonstrate relevance at the highest levels, it will be necessary to anticipate what will be relevant in the future. Ensemble forecasts are here, and there is a need to ensure that users understand how to use the wealth of information available and not fall back to using only a small part of the output (e.g., the mean). GSL could lead the way in helping users interpret and extract value from ensemble output. Another area of opportunity is in verification. GSL is moving towards a process-oriented verification that will examine the physics contributions to overall model error; this will be very valuable for model development work. But while verification is part of many GSL projects, it is not clear how embedded verification is with the overall GSL end-to-end process presented by leadership. Integrating verification into earlier stages of development is a potential area of opportunity, but it seems that currently much of GSL's verification is motivated by a response to external stakeholder needs and funding (e.g., the FAA support for aviation weather products). Verification has mostly focused on meteorological accuracy with some work attempting to examine and measure impacts. GSL could consider strategies to extend verification activities to other areas of the lab, including smoke and chemistry forecasting. During the review, GSL expressed a desire to measure economic benefits of weather information; this is another area of opportunity that could expand relevance beyond the current set of stakeholders.

3.4.3 Performance

GSL is well-organized and managed to conduct R&D. GSL is appropriately willing to take on high-risk initiatives that may not prove to be successful or may not have an immediate path to operations. GSL has a reputation for completing objectives and milestones and being effective and efficient in delivering high-quality products. For instance, GSL's independent impact-based assessments of new weather products for the FAA provide a basis for their approval for use or identify the need for additional development; the skill scores provided by the Development Testbed Center and via the model evaluation tools are essential for comparing NWP model alternatives and assessing progress; and the use of verification methods pervades GSL's own weather product development activities. However, there is a tendency for verification to be conducted at the project level rather than from a strategic, lab-wide approach and efficiencies could be found by adopting the latter approach. In general, stakeholders are engaged and complimentary about the applications and information delivered; however, stakeholder feedback indicated opportunities for improved communication. And while GSL appears to be interested in

partnering with organizations and agencies on user-driven decision support, the mechanisms for reaching out to other organizations and agencies and developing new partnerships are not always clear.

4 Lab Review Process

Overall, the review process went well. Review panelists appreciated the thoughtfulness, time, effort, and planning that went into the review from both GSL and NOAA OAR. While there were a few drawbacks to holding the review entirely virtually, these were minimized by GSL and OAR's organization of the process and attention to detail. Panel members appreciated the opportunity to participate and enjoyed learning more about GSL. Several noted that the experience would be valuable for future reviews of their own organizations.

Overall, reviewers felt that the videos that were provided before the reviews were informative and gave panel members the ability to watch as time permitted before and after the review. The availability of subject matter experts for discussion during the review was beneficial because it enabled reviewers to hear more details about the work being conducted, ask detailed questions, and engage in discussion. It also provided reviewers the opportunity to see how members of thematic areas worked together as teams. All of the presentation materials were high quality and panelists appreciated the slides that explicitly highlighted a topic's connection to Quality, Relevance and Performance. The lunchtime sessions with federal, cooperative institute, and contractor employees provided insight into the GSL work environment. However, the sample size was small, and the amount of time was short. In addition, receiving the stakeholder questionnaires before the stakeholder meetings helped focus discussion during the brief meetings. If future reviews are conducted virtually, GSL and OAR may consider mixing up the panel member groups between different sessions. A key takeaway is that all of the information provided was useful to the panel members.

There were a few drawbacks of the virtual review and suggestions for future reviews. The format of the review meant that there were few opportunities for informal interactions either with GSL or between the panelists. If future reviews take place virtually, a chat function available just to the reviewers would be useful. Panelists noted that a leadership-level presentation to summarize actions taken to address the main issues raised at the last review and any challenges in addressing them would have been useful, as well as a top-down presentation that clearly identifies NOAA priorities flowing to OAR priorities to NWS priorities (the last being GSL's primary customer). While some of this information was available, it required panel members to review multiple NOAA documents and strategic plans to find the information. Panelists also felt that more written documentation with overarching information about the different Lab programs and activities would have been useful for reference, especially when writing the individual evaluations. Other suggestions include providing the review panelists:

- a printed executive summary;
- a separate document that includes a more clearly articulated path towards improving diversity, equity, and inclusion at GSL that indicates the rationale, goals, expectations/challenges, and the means for assessing and evaluating progress;
- and a decision tree that reflects the process for project prioritization or a similar type of element in the presentation to make it more clear to future review teams.

Panelists had additional feedback for OAR regarding the evaluation forms. Panelists noted that much of the guidance for rating Quality, Relevance and Performance seemed related more to overall lab leadership, organization, and activities than just an assigned thematic area. This could lead to repetitive responses from reviewers evaluating more than one thematic area. It might make more sense to provide a separate section for overarching feedback rather than embedding it within the specific thematic areas. Additionally, there were a lot of levels to the "Performance" category and not all seemed relevant to each of the thematic areas.

5 Recommendations

5.1 Organizational Excellence Recommendations

1. GSL should work with NOAA OAR leadership to develop and implement an all-OAR laboratory plan and timeline for increasing diversity.
2. To the extent possible, coordinate mentoring programs for early career and postdocs across the embedded stakeholders (e.g., the cooperative institutes). Giving special attention to the postdoctoral fellows within GSL may be low-hanging fruit. Such a program should include mentor and coach training.
3. Develop stronger relationships with the NOAA Cooperative Science Centers. These NOAA-funded entities routinely recruit, train, and support hundreds of students from diverse ethnic identities in NOAA mission sciences.
4. Develop a formalized relationship to the Cooperative Science Centers and other Minority-Serving Institution partners that engages their faculty expertise with roles or membership on governance, advisory, and strategic planning committees.
5. GSL should work to increase the number of journal articles from staff members who are primarily working on science/research. Consider surveying similar organizations when developing the goal to figure out reasonable targets and learn about possible difficulties and solutions. (Options include incentivizing publications in appropriate journals and lowering any existing organizational barriers to publishing.)
6. GSL should consider elevating the status of research to operations transitions and developing a way to present them as a parallel performance metric to journal articles.
7. Related to the previous two recommendations, GSL should strive to prioritize its documentation of research results and software development so that current information is easy to find and cite, and authors are easy to contact.
8. If GSL wants to move forward with its plans to focus on lower readiness-level research, GSL should assess the risks and identify/document current barriers to research and development and share with OAR leadership before putting together a plan with milestones.
9. Develop a consistent outreach and branding strategy that makes clear to stakeholders, peer institutions, and potential funding sources who GSL is, what GSL does, and why GSL is unique. Such a strategy should engage GSL staff at all levels.
10. GSL should consider adding to its website more high-quality graphics and model output visualizations targeting the general public.

5.2 Advanced Technologies Recommendations

11. GSL should play a leadership role within NOAA in developing and showcasing best practices to improve computing efficiency, save money, and reduce environmental harm. To start, GSL should identify the carbon footprint of its lab, computing resources, and ultimately the models/tools it develops. This footprint could be used as an additional criterion in decision making at many levels (e.g., procurement, model development, etc.).
12. Identify one or more driving needs around exascale computing and develop and implement an end-to-end plan that includes model readiness (e.g., GPUs, machine learning, etc.), data and workflow requirements, codesign, leadership, and stakeholder buy-in around this need.
13. Strategically develop and demonstrate applications of AI/ML to simplify, accelerate and improve the quality of NOAA products, including improving forecaster workflow by automating routine tasks, guiding forecaster attention to where it can add the most value and providing decision recommendations to end-users. Ensure that GSL has a strategic plan for AI that is in line with NOAA's strategic plan.
14. Continue to help shape and engage with the NOAA Center for AI and maintain collaboration with the NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography.
15. Look for longer-term funding opportunities to broaden the Advanced Technologies area to reach critical mass or to broaden your portfolio in highly relevant advanced technologies.
16. Scale, replicate and document the AQPI project for additional regions and use cases, and as a paradigm for building end-user-focused weather systems.
17. Draw on GSL's expertise in environmental observations, modeling, cloud, AI and verification to create and make available curated datasets and compute environments to support workforce education and experimentation with AI training and evaluation techniques.
18. Adopt coding standards and encourage uniform staff use of version control (e.g., GitHub) to facilitate code sharing, building a shared reservoir of capability, onboarding new staff, and ensuring sustainability.
19. As part of longer-term NOAA strategy, GSL could explore how quantum computing may be used to dramatically accelerate NWP, AI and other computing relevant to NOAA's mission.

5.3 Earth System Prediction Recommendations

20. NOAA has labs with strongly overlapping interests and research programs in earth system model development, e. g. GSL, NCEP/EMC, and GFDL. At a higher level, it is important for NOAA leadership to clearly define the distinct roles of these labs so that they are rewarded for collaborating and do not duplicate effort more than is necessary.
21. GSL should work with other NOAA labs and OAR leadership to establish clear targets and responsibilities regarding UFS development.
22. A detailed, resourced plan should be produced to enable understanding of the implications for GSL of the transition of the RAP/HRRR to RRFS in the UFS framework. In particular which types of activities will GSL be doing more of and which less and what are the implications for your deployment of resources? This ties in with GSL's goal of doing more research. As part of this there needs to be a realistic and resourced plan for incorporating the FV3 dycore.

23. Develop a long term (5-10 year) plan to carry out your stated "Grand Scientific Challenge" to develop global, rapidly updating storm-scale models. This project may need to start relatively slowly due to the resource limitations of the UFS transition discussed above. The plan needs to be developed in collaboration with other labs and organizations and will also require significant additional HPC resources.
24. Consider a long term (5-10 year) project, in collaboration with other centers or groups, to develop the capability to run 100-m/urban scale regional models. This will need to include plans to improve the model/parameterizations for 100m scale configurations, improve the representation of the urban surface and work with potential stakeholders in these models. A plan to obtain sufficient HPC resources will also be key here.
25. Increase the development and use of satellite information for data assimilation.
26. Build up the collaboration with other testbeds such as the Aviation Weather Testbed to the same level as with the Hazardous Weather Testbed
27. Develop collaborations with groups using the relatively new convection permitting climate models (which are used for example to understand how heavy rain might change in a future climate).
28. Try to seek more strategic international collaborations. For example, understanding how your models perform and are perceived in other parts of the world would probably be very informative for your model development.

5.4 Decision Support Recommendations

29. Continue with the inclusion of social science researchers as part of the GSL Decision Support development team.
30. Further extend the target of Decision Support and user experience R&D beyond NWS forecasters and traditional stakeholders to end-user consumers of the forecast information and underserved countries (recognizing that external funding may be required).
31. Expand efforts in Decision Support for ensemble weather products to include ensemble uncertainty quantification and how users, both meteorologists and non-meteorologists, should interpret ensemble output.
32. Develop the capability within GSL to measure economic benefits of weather information.
33. Prioritize verification at the lab level and make more use of verification in the earlier stages of development by approaching data assimilation, verification, and forecasting in a more holistic approach to fully utilize GSL's strengths and resources.
34. Re-establish site visits for Decision Support feedback and needs as soon as is safe and feasible.
35. Consider expanded use of dissemination venues such as public GitHub repositories, blog posts, user manuals, software documentation, social media, and tracking of related key performance indicators such as views or downloads to supplement journal and conference publications in highlighting successes, sharing R&D results and growing the set of GSL stakeholders.
36. Consider expanding connections with private industry as collaborators, recipients, and implementers of GSL decision support R&D outputs. For example, recommended routes provided by Integrated Support for Impacted Air-Traffic Environments (INSITE) might be valuable to airline dispatchers. Create open-source, cloud-ready containerized implementations, and track GSL's expanded impact using performance indicators like GitHub downloads.
37. Offer objective verification services to industry, e.g., in comparing performance of aviation turbulence forecast skill. This could leverage the techniques developed for FAA evaluations,

attract additional funding, and benefit society by fostering competition to improve the quality of commercial forecast products.