SF 424

Table of Contents

Table of Contents 1
SF 424 Application for Federal Assistance 2
SF-424a Budget Information - Non-Construction Programs 5
CD511 8
SF-424b Assurances - Non-Construction Programs 9
Project Narrative Attachments 11
NOAA_Proposal_20241037457671 11
Budget Narrative Attachments 51
NOAA_Budget_Justification1037457665 51
GTRC_FY26_27_FA_Rate_Agrmt1037457693 53

Application for Federal Assistance SF-424

* 1. Type of Submission:	* 2. Type of Application:	* If Revision, select appropriate letter(s):
❍ Preapplication	● New	
● Application	❍ Continuation	* Other (Specify)
❍ Changed/Corrected Application	❍ Revision	

* 3. Date Received:	4. Applicant Identifier:
12/02/2024	XLIU976

5a. Federal Entity Identifier:	5b. Federal Award Identifier:

State Use Only:

6. Date Received by State:	7. State Application Identifier:

8. APPLICANT INFORMATION:

* a. Legal Name: Georgia Tech Research Corporation

* b. Employer/Taxpayer Identification Number (EIN/TIN):	* c. UEI:
1-580603146-11	EMW9FC8J3HN4

d. Address:

* Street1:	926 Dalney Street NW
Street2:	
* City:	Atlanta
County/Parish:	Fulton
* State:	GA: Georgia
Province:	
* Country:	USA: UNITED STATES
* Zip / Postal Code:	303320415

e. Organizational Unit:

Department Name:	Division Name:
VP Research Administration	EVP Research

f. Name and contact information of person to be contacted on matters involving this application:

Prefix:		* First Name:	Douglas
Middle Name:	E		
* Last Name:	Heaton		
Suffix:			

Title: Research Associate I

Organizational Affiliation:

Georgia Tech Research Corporation

* Telephone Number:	404-710-9101	Fax Number:	

* Email: douglas.heaton@osp.gatech.edu

Application for Federal Assistance SF-424

*** 9. Type of Applicant 1: Select Applicant Type:**

H: Public/State Controlled Institution of Higher Education

Type of Applicant 2: Select Applicant Type:

Type of Applicant 3: Select Applicant Type:

* Other (specify):

*** 10. Name of Federal Agency:**

DOC NOAA - ERA Production

11. Catalog of Federal Domestic Assistance Number:

11.459

CFDA Title:

Weather and Air Quality Research

*** 12. Funding Opportunity Number:**

NOAA-OAR-WPO-2025-28603

* Title:

FY2025 Weather Program Office Research Programs Announcement - Air Quality Research and Forecasting (AQRF)

13. Competition Identification Number:

Title:

14. Areas Affected by Project (Cities, Counties, States, etc.):

File Name:

*** 15. Descriptive Title of Applicant's Project:**

A Physics-Informed Machine Learning System for Real-Time Air Quality Forecasting during Wildfire Events

Attach supporting documents as specified in agency instructions.

File Name:

Application for Federal Assistance SF-424

16. Congressional Districts Of:

* a. Applicant: GA-005

* b. Program/Project: GA-005

Attach an additional list of Program/Project Congressional Districts if needed.

17. Proposed Project:

* a. Start Date: 08/01/2025

* b. End Date: 07/31/2028

18. Estimated Funding ($):

* a. Federal	1,050,000.00
* b. Applicant	0.00
* c. State	0.00
* d. Local	0.00
* e. Other	0.00
* f. Program Income	0.00
* g. TOTAL	1,050,000.00

*** 19. Is Application Subject to Review By State Under Executive Order 12372 Process?**

❍ a. This application was made available to the State under the Executive Order 12372 Process for review on ______.

❍ b. Program is subject to E.O. 12372 but has not been selected by the State for review.

● c. Program is not covered by E.O. 12372.

*** 20. Is the Applicant Delinquent On Any Federal Debt? (If "Yes", provide explanation in attachment.)**

❍ Yes ● No

21. *By signing this application, I certify (1) to the statements contained in the list of certifications and (2) that the statements herein are true, complete and accurate to the best of my knowledge. I also provide the required assurances** and agree to comply with any resulting terms if I accept an award. I am aware that any false, fictitious, or fraudulent statements or claims may subject me to criminal, civil, or administrative penalties. (U.S. Code, Title 218, Section 1001)**

☑ ** I AGREE

** The list of certifications and assurances, or an internet site where you may obtain this list, is contained in the announcement or agency specific instructions.

Authorized Representative:

Prefix:

* First Name: Douglas

Middle Name: E

* Last Name: Heaton

Suffix:

* Title: Research Associate I

* Telephone Number: 404-710-9101

Fax Number:

* Email: douglas.heaton@osp.gatech.edu

* Signature of Authorized Representative: Douglas E Heaton

* Date Signed: 12/02/2024

BUDGET INFORMATION -
Non-Construction Programs

OMB Approval No. 4040-0006
Expiration Date 02/28/2025

SECTION A - BUDGET SUMMARY						
Grant Program Function or Activity (a)	Catalog of Federal Domestic Assistance Number (b)	Estimated Unobligated Funds		New or Revised Budget		
		Federal (c)	Non-Federal (d)	Federal (e)	Non-Federal (f)	Total (g)
1 . FY2025 Weather Program Office Research Programs Announcement	11.459	$0.00	$0.00	$350,000.00	$0.00	$350,000.00
2 . FY2025 Weather Program Office Research Programs Announcement	11.459	$0.00	$0.00	$0.00	$0.00	$0.00
3 . FY2025 Weather Program Office Research Programs Announcement	11.459	$0.00	$0.00	$0.00	$0.00	$0.00
4 .						$0.00
5. Totals		$0.00	$0.00	$350,000.00	$0.00	$350,000.00

SECTION B - BUDGET CATEGORIES					
6. Object Class Categories	GRANT PROGRAM, FUNCTION OR ACTIVITY				**Total (5)**
	(1) FY2025 Weather Program Office Research Programs Announcement	(2) FY2025 Weather Program Office Research Programs Announcement	(3) FY2025 Weather Program Office Research Programs Announcement	(4)	
a. Personnel	$160,929.00	$165,757.00	$164,558.00		$491,244.00
b. Fringe Benefits	$24,712.00	$25,453.00	$25,729.00		$75,894.00
c. Travel	$6,958.00	$495.00	$2,160.00		$9,613.00
d. Equipment					
e. Supplies					
f. Contractual					
g. Construction					
h. Other	$46,850.00	$48,256.00	$47,088.00		$142,194.00
i. Total Direct Charges (*sum of 6a-6h*)	$239,449.00	$239,961.00	$239,535.00		$718,945.00
j. Indirect Charges	$110,551.00	$110,039.00	$110,465.00		$331,055.00
k. TOTALS (*sum of 6i and 6j*)	$350,000.00	$350,000.00	$350,000.00		$1,050,000.00
7. Program Income					

Standard From 424A (Rev. 7-97)

SECTION C - NON-FEDERAL RESOURCES					
(a) Grant Program		(b) Applicant	(c) State	(d) Other Sources	(e) TOTALS
12. TOTAL (sum of lines 8-11)					

SECTION D - FORECASTED CASH NEEDS					
	Total for 1st Year	**1st Quarter**	**2nd Quarter**	**3rd Quarter**	**4th Quarter**
13. Federal	$350,000.00	$87,500.00	$87,500.00	$87,500.00	$87,500.00
14. Non-Federal	$0.00	$0.00	$0.00	$0.00	$0.00
15. TOTAL (*sum of lines 13 and 14*)	$350,000.00	$87,500.00	$87,500.00	$87,500.00	$87,500.00

SECTION E - BUDGET ESTIMATES OF FEDERAL FUNDS NEEDED FOR BALANCE OF THE PROJECT				
(a) Grant Program	FUTURE FUNDING PERIODS (Years)			
	(b) First	(c) Second	(d) Third	(e) Fourth
20. TOTAL (*sum of lines 16-19*)				

SECTION F - OTHER BUDGET INFORMATION	
21. Direct Charges:	22. Indirect Charges: MTDC 57.4%
23. Remarks:	

Standard Form 424A (rev. 7-97) Page3

FORM **CD511** U.S. DEPARTMENT OF COMMERCE
(REV 1-05) CERTIFICATION REGARDING LOBBYING

Applicants should also review the instructions for certification included in the regulations before completing this form. Signature on this form provides for compliance with certification requirements under 15 CFR Part 28, 'New Restrictions on Lobbying.' The certifications shall be treated as a material representation of fact upon which reliance will be placed when the Department of Commerce determines to award the covered transaction, grant, or cooperative agreement.

LOBBYING

As required by Section 1352, Title 31 of the U.S. Code, and implemented at 15 CFR Part 28, for persons entering into a grant, cooperative agreement or contract over $100,000 or a loan or loan guarantee over $150,000 as defined at 15 CFR Part 28, Sections 28.105 and 28.110, the applicant certifies that to the best of his or her knowledge and belief, that:

(1) No Federal appropriated funds have been paid or will be paid, by or on behalf of the undersigned, to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress in connection with the awarding of any Federal contract, the making of any Federal grant, the making of any Federal loan, the entering into of any cooperative agreement, and the extension, continuation, renewal, amendment, or modification of any Federal contract, grant, loan, or cooperative agreement.

(2) If any funds other than Federal appropriated funds have been paid or will be paid to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a member of Congress in connection with this Federal contract, grant, loan, or cooperative agreement, the undersigned shall complete and submit Standard Form-LLL, 'Disclosure Form to Report Lobbying.' in accordance with its instructions.

(3) The undersigned shall require that the language of this certification be included in the award documents for all subawards at all tiers (including subcontracts, subgrants, and contracts under grants, loans, and cooperative agreements) and that all subrecipients shall certify and disclose accordingly.

This certification is a material representation of fact upon which reliance was placed when this transaction was made or entered into. Submission of this certification is a prerequisite for making or entering into this transaction imposed by section 1352, title 31, U.S. Code. Any person who fails to file the required certification shall be subject to a civil penalty of not less than $10,000 and not more than $100,000 for each such failure occurring on or before October 23, 1996, and of not less than $11,000 and not more than $110,000 for each such failure occurring after October 23, 1996.

Statement for Loan Guarantees and Loan Insurance

The undersigned states, to the best of his or her knowledge and belief, that:

In any funds have been paid or will be paid to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with this commitment providing for the United States to insure or guarantee a loan, the undersigned shall complete and submit Standard Form-LLL, 'Disclosure Form to Report Lobbying,' in accordance with its instructions.

Submission of this statement is a prerequisite for making or entering into this transaction imposed by section 1352, title 31, U.S. Code. Any person who fails to file the required statement shall be subject to a civil penalty of not less than $10,000 and not more than $100,000 for each such failure occurring on or before October 23, 1996, and of not less than $11,000 and not more than $110,000 for each such failure occurring after October 23, 1996.

As the duly authorized representative of the applicant, I hereby certify that the applicant will comply with the above applicable certification.

* NAME OF APPLICANT
Georgia Tech Research Corporation

* AWARD NUMBER

*PROJECT NAME
A Physics-Informed Machine Learning System for Real-Time Air

Prefix: * First Name: Douglas Middle Name: E
* Last Name: Heaton Suffix:
* Title: Research Associate I

*** SIGNATURE:**
Douglas E Heaton

*** DATE:**
2024-12-02

ASSURANCES - NON-CONSTRUCTION PROGRAMS

OMB Approval No. 4040-0007
Expiration Date 02/28/2025

Public reporting burden for this collection of information is estimated to average 15 minutes per response, including time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding the burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to the Office of Management and Budget, Paperwork Reduction Project (0348-0040), Washington, DC 20503.

PLEASE DO NOT RETURN YOUR COMPLETED FORM TO THE OFFICE OF MANAGEMENT AND BUDGET. SEND IT TO THE ADDRESS PROVIDED BY THE SPONSORING AGENCY.

NOTE: Certain of these assurances may not be applicable to your project or program. If you have questions, please contact the awarding agency. Further, certain Federal awarding agencies may require applicants to certify to additional assurances. If such is the case, you will be notified.

As the duly authorized representative of the applicant, I certify that the applicant:

1. Has the legal authority to apply for Federal assistance and the institutional, managerial and financial capability (including funds sufficient to pay the non-Federal share of project cost) to ensure proper planning, management and completion of the project described in this application.
2. Will give the awarding agency, the Comptroller General of the United States and, if appropriate, the State, through any authorized representative, access to and the right to examine all records, books, papers, or documents related to the award; and will establish a proper accounting system in accordance with generally accepted accounting standards or agency directives.
3. Will establish safeguards to prohibit employees from using their positions for a purpose that constitutes or presents the appearance of personal or organizational conflict of interest, or personal gain.
4. Will initiate and complete the work within the applicable time frame after receipt of approval of the awarding agency.
5. Will comply with the Intergovernmental Personnel Act of 1970 (42 U.S.C. §§4728-4763) relating to prescribed standards for merit systems for programs funded under one of the 19 statutes or regulations specified in Appendix A of OPM's Standards for a Merit System of Personnel Administration (5 C.F.R. 900, Subpart F).
6. Will comply with all Federal statutes relating to nondiscrimination. These include but are not limited to: (a) Title VI of the Civil Rights Act of 1964 (P.L. 88-352) which prohibits discrimination on the basis of race, color or national origin; (b) Title IX of the Education Amendments of 1972, as amended (20 U.S.C. §§1681-1683, and 1685-1686), which prohibits discrimination on the basis of sex; (c) Section 504 of the Rehabilitation Act of 1973, as amended (29 U.S.C. §794), which prohibits discrimination on the basis of handicaps; (d) the Age Discrimination Act of 1975, as amended (42 U.S.C. §§6101-6107), which prohibits discrimination on the basis of age; (e) the Drug Abuse Office and Treatment Act of 1972 (P.L. 92-255), as amended, relating to nondiscrimination on the basis of drug abuse; (f) the Comprehensive Alcohol Abuse and Alcoholism Prevention, Treatment and Rehabilitation Act of 1970 (P.L. 91-616), as amended, relating to nondiscrimination on the basis of alcohol abuse or alcoholism; (g) §§523 and 527 of the Public Health Service Act of 1912 (42 U.S.C. §§290 dd-3 and 290 ee- 3), as amended, relating to confidentiality of alcohol and drug abuse patient records; (h) Title VIII of the Civil Rights Act of 1968 (42 U.S.C. §§3601 et seq.), as amended, relating to nondiscrimination in the sale, rental or financing of housing; (i) any other nondiscrimination provisions in the specific statute(s) under which application for Federal assistance is being made; and, (j) the requirements of any other nondiscrimination statute(s) which may apply to the application.
7. Will comply, or has already complied, with the requirements of Titles II and III of the Uniform Relocation Assistance and Real Property Acquisition Policies Act of 1970 (P.L. 91-646) which provide for fair and equitable treatment of persons displaced or whose property is acquired as a result of Federal or federally-assisted programs. These requirements apply to all interests in real property acquired for project purposes regardless of Federal participation in purchases.
8. Will comply, as applicable, with provisions of the Hatch Act (5 U.S.C. §§1501-1508 and 7324-7328) which limit the political activities of employees whose principal employment activities are funded in whole or in part with Federal funds.

9. Will comply, as applicable, with the provisions of the Davis- Bacon Act (40 U.S.C. §§276a to 276a-7), the Copeland Act (40 U.S.C. §276c and 18 U.S.C. §874), and the Contract Work Hours and Safety Standards Act (40 U.S.C. §§327- 333), regarding labor standards for federally-assisted construction subagreements.

10. Will comply, if applicable, with flood insurance purchase requirements of Section 102(a) of the Flood Disaster Protection Act of 1973 (P.L. 93-234) which requires recipients in a special flood hazard area to participate in the program and to purchase flood insurance if the total cost of insurable construction and acquisition is $10,000 or more.

11. Will comply with environmental standards which may be prescribed pursuant to the following: (a) institution of environmental quality control measures under the National Environmental Policy Act of 1969 (P.L. 91-190) and Executive Order (EO) 11514; (b) notification of violating facilities pursuant to EO 11738; (c) protection of wetlands pursuant to EO 11990; (d) evaluation of flood hazards in floodplains in accordance with EO 11988; (e) assurance of project consistency with the approved State management program developed under the Coastal Zone Management Act of 1972 (16 U.S.C. §§1451 et seq.); (f) conformity of Federal actions to State (Clean Air) Implementation Plans under Section 176(c) of the Clean Air Act of 1955, as amended (42 U.S.C. §§7401 et seq.); (g) protection of underground sources of drinking water under the Safe Drinking Water Act of 1974, as amended (P.L. 93-523); and, (h) protection of endangered species under the Endangered Species Act of 1973, as amended (P.L. 93- 205).

12. Will comply with the Wild and Scenic Rivers Act of 1968 (16 U.S.C. §§1271 et seq.) related to protecting components or potential components of the national wild and scenic rivers system.

13. Will assist the awarding agency in assuring compliance with Section 106 of the National Historic Preservation Act of 1966, as amended (16 U.S.C. §470), EO 11593 (identification and protection of historic properties), and the Archaeological and Historic Preservation Act of 1974 (16 U.S.C. §§469a-1 et seq.).

14. Will comply with P.L. 93-348 regarding the protection of human subjects involved in research, development, and related activities supported by this award of assistance.

15. Will comply with the Laboratory Animal Welfare Act of 1966 (P.L. 89-544, as amended, 7 U.S.C. §§2131 et seq.) pertaining to the care, handling, and treatment of warm blooded animals held for research, teaching, or other activities supported by this award of assistance.

16. Will comply with the Lead-Based Paint Poisoning Prevention Act (42 U.S.C. §§4801 et seq.) which prohibits the use of lead-based paint in construction or rehabilitation of residence structures.

17. Will cause to be performed the required financial and compliance audits in accordance with the Single Audit Act Amendments of 1996 and OMB Circular No. A-133, "Audits of States, Local Governments, and Non-Profit Organizations."

18. Will comply with all applicable requirements of all other Federal laws, executive orders, regulations, and policies governing this program.

19. Will comply with the requirements of Section 106(g) of the Trafficking Victims Protection Act (TVPA) of 2000, as amended (22 U.S.C. 7104) which prohibits grant award recipients or a sub-recipient from (1) Engaging in severe forms of trafficking in persons during the period of time that the award is in effect (2) Procuring a commercial sex act during the period of time that the award is in effect or (3) Using forced labor in the performance of the award or subawards under the award.

* SIGNATURE OF AUTHORIZED CERTIFYING OFFICIAL	* TITLE
Douglas E Heaton	Research Associate I
* APPLICANT ORGANIZATION	* DATE SUBMITTED
Georgia Tech Research Corporation	12-02-2024

Standard Form 424B (Rev. 7-97) Back

A Physics-Informed Machine Learning System for Real-Time Air Quality Forecasting during Wildfire Events

Notice of Funding Opportunity: NOAA-OAR-WPO-2025-28603
Project Period: August 1, 2025 – July 31, 2028

AQRF Program Priorities: **AQRF-3**, **AQRF-5**, and **AQRF-6**

PI: Xiao Liu, David M. McKenney Family Associate Professor,
H. Milton Stewart School of Industrial and Systems Engineering,
Georgia Institute of Technology
Address: 755 Ferst Drive, NW Atlanta, GA 30332-0205
Phone: (914)-826-6347, email: xiao.liu@isye.gatech.edu

Co-PI 1: Yao Xie, Coca-Cola Foundation Chair and Professor,
H. Milton Stewart School of Industrial and Systems Engineering,
Georgia Institute of Technology
Phone: (404) 385-1687, email: yao.xie@isye.gatech.edu

Co-PI 2: Yuhang Wang, Professor,
School of Earth and Atmospheric Sciences, College of Sciences
Georgia Institute of Technology
Phone: (404)-894-3995, email: yuhang.wang@eas.gatech.edu

Authorized Institutional Representative, Georgia Institute of Technology
Douglas Heaton, Office of Sponsored Programs, 926 Dalney Street, NW Atlanta, GA
Phone: (404)-710-9101, email: douglas.heaton@osp.gatech.edu

Budget Summary

Georgia Institute of Technology	Year 1	Year 2	Year 3	Total
Direct Cost	$239,449	$239,961	$239,535	$718,945
Indirect Cost (57.4%)	$110,551	$110,039	$110,465	$331,055
Total	$350,000	$350,000	$350,000	$1,050,000

Additional Funding Notes
Subawards: None

Starting Readiness Level: RL 3
Ending Readiness Level: RL 7

Abstract

A Physics-Informed Machine Learning System for Real-Time Air Quality Forecasting during Wildfire Events

- **Project Goal:** This project will develop and validate *computationally-fast*, *physics-informed*, and *multi-source data-driven* Machine Learning (ML) algorithms and software tools for generating low-latency hourly air quality predictions (updated every 10 minutes) for the next 24 hours during hazardous wildfire episodes. The ML system will consist of a complete suite of three components: (i) a fire spread and source estimation module, (ii) a smoke propagation module, and (iii) an air quality prediction module. The project addresses Air Quality Research and Forecasting (AQRF) Priorities #3, #5 and #6.

- **Problem/Opportunity Statement:** Physics/Chemistry-based and computationally-intensive numerical models—such as CMAQ which aims to fully resolve governing equations—form the backbones of the current National Air Quality Forecast Capability (NAQFC). On the other hand, to inform the public—in real time—about hazardous air pollution during wildfire episodes, complementary tools can be added to the existing NAQFC toolbox by harnessing the latest advances of physics-informed ML that efficiently utilizes multi-source streaming observational data (addressing Goals 2&4 of WPO Strategic Plans, as well as the Priorities for Weather Research related to forecasting capabilities).

- **Methodology/Activities to be Performed:** (i) For Component I (i.e., the "Fire Spread and Source Estimation" module), the project will develop a physics-informed spatio-temporal Deep Hierarchical Dynamic model for fast prediction of fire spread, by integrating multi-source data including NOAA's GOES-R remote sensing fire products; (ii) For Component II (i.e., the "Smoke Propagation" module), this project will develop a dynamic data fusion and prediction model for aerosol propagation regularized by advection-diffusion processes, leveraging GOES-R multi-satellite aerosol data streams; and (iii) For Component III (i.e., the "Air Quality" module), an infinite-dimensional dynamic model will be utilized to make air quality predictions based on the outputs from Components I & II.

- **Primary Project Products/Outputs:** This project will generate: (i) ML algorithms for fire spread, smoke propagation and air quality predictions; (ii) a software package that implements the developed ML system to facilitate NOAA's potential R2O transitions; (iii) technical reports/papers/conference presentations for dissemination; and (iv) training materials (user manuals) for the ML algorithms/system/software.

- **Expected Results, Outcomes, and Benefits:** The physics-informed ML system will generate complementary capabilities, which can be added to the current NAQFC toolbox, by providing accurate and real-time air quality predictions during wildfire events. The system will enhance forecasting, situational awareness and emergency response capabilities for Federal and local agencies, and timely inform the public to take preventive measures during hazardous air pollution episodes. The project activities will include a diverse body of researchers/students, and promote the integration of ML into fire weather research and environmental awareness, helping to achieve NOAA's long-term strategic goals.

1 Problem/Opportunity Statement

Physics-based and computationally-intensive numerical models—such as AQM (Air Quality Modeling system[33]) and CMAQ (Community Multiscale Air Quality Modeling System[37, 38])—form the backbones of the current National Air Quality Forecast Capability (NAQFC). On the other hand, to inform the public—in real time—about hazardous air pollution during wildfire episodes, complementary capabilities can be added to the existing NAQFC toolbox by developing *computationally-fast*, *physics-informed*, and *multi-source data-driven* Machine Learning (ML) models that fully utilize streaming observational data.

Leveraging the latest advances in ML, this project will develop a **physics-informed multi-source data-driven ML system** that generates low-latency, probabilistic and hourly air quality predictions—updated every 10 minutes for the next 24 hours over a continuous spatial domain—during hazardous wildfire episodes. The system integrates 3 inter-connected modules:

1. Component I: the "Fire Spread and Source Estimation" module based on a physic-informed deep hierarchical dynamic fire spread model.
2. Component II: the "Smoke Propagation" module for real-time aerosol propagation prediction utilizing multi-satellite remote-sensing data.
3. Component III: the "Air Quality" module for real-time air quality prediction by integrating fire spread, smoke propagation and air quality monitoring data.

The project will also create **software tools** that facilitate potential Research to Operations (R2O) transitions and other NOAA programs such as the Testbed Program.

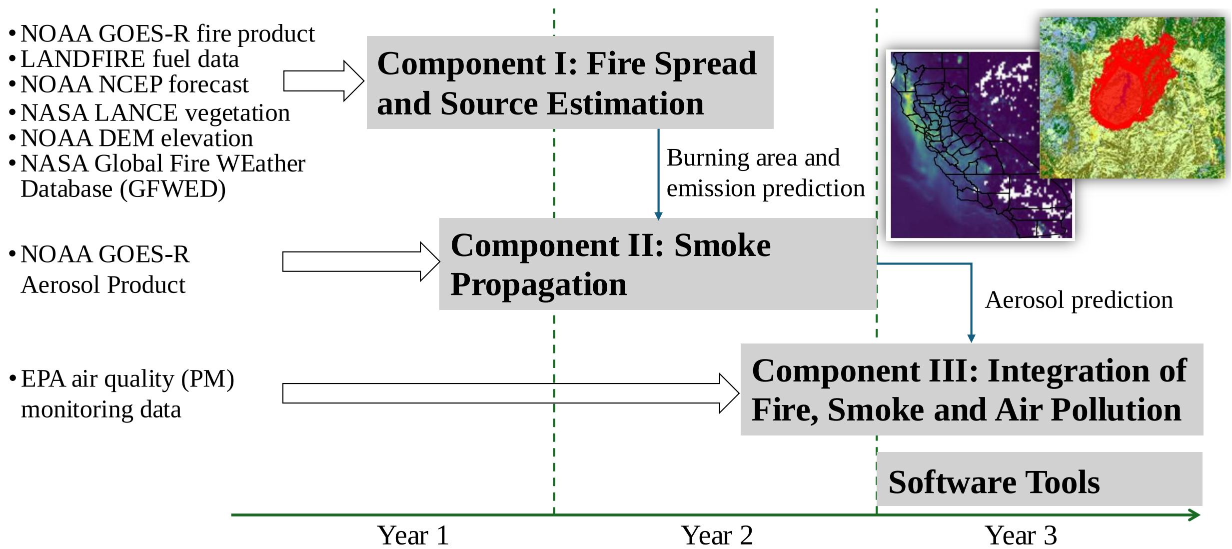

Figure 1: An overview to the physics-informed multi-source data-driven ML system

Fig.1 provides an overview to the physics-informed multi-source data-driven ML system that integrates 3 inter-connected modules: fire, smoke and air pollution:

(i) Component I: the "Fire Spread and Source Estimation" module. Integrating NOAA GOES-R remote sensing fire products (binary burned area), landscape information, vegetation, meteorological variables and existing fire spread models, the team will develop **a physics-informed spatio-temporal Deep Hierarchical Dynamic model for fast prediction of fire spread** (which provides information needed for pollution source estimation). Unlike black-box ML, this approach leverages the key structure of existing fire spread models (to make the output *interpretable*). Unlike deterministic fire spread models, the proposed ML approach will naturally provide uncertainty quantification (to inform better *decisions under uncertainty*) and enable computationally efficient real-time forecasting [*Component-I addresses the program priority AQRF-3 to provide real-time spatial and temporal source estimates].

(ii) Component II: the "Smoke Propagation" module. Leveraging the output of Component-I and multi-satellite NOAA GOES-R remote-sensing aerosol data streams, **a dynamic data fusion and prediction model regularized by advection-diffusion processes** will be developed for aerosol/smoke propagation. This approach will enable the estimation of underlying true physical smoke propagation using heterogeneous multi-satellite data streams with different data missing rate, bias, measurement errors, etc. [*Component-II addresses AQRF-5 to provide improved fusion of multi-source remote sensing data].

(iii) Component III: the "Air Quality" module. An **infinite-dimensional dynamic model** will be developed to explain how air pollution processes are driven by fire spread and smoke propagation processes, and enable accurate predictions of hourly PM2.5 concentrations [*Component-III addresses AQRF-5 to provide improved air quality prediction].

Finally, the three modules above will be integrated into a software package and tested against real data, addressing AQRF-6 to develop software packages for verification.

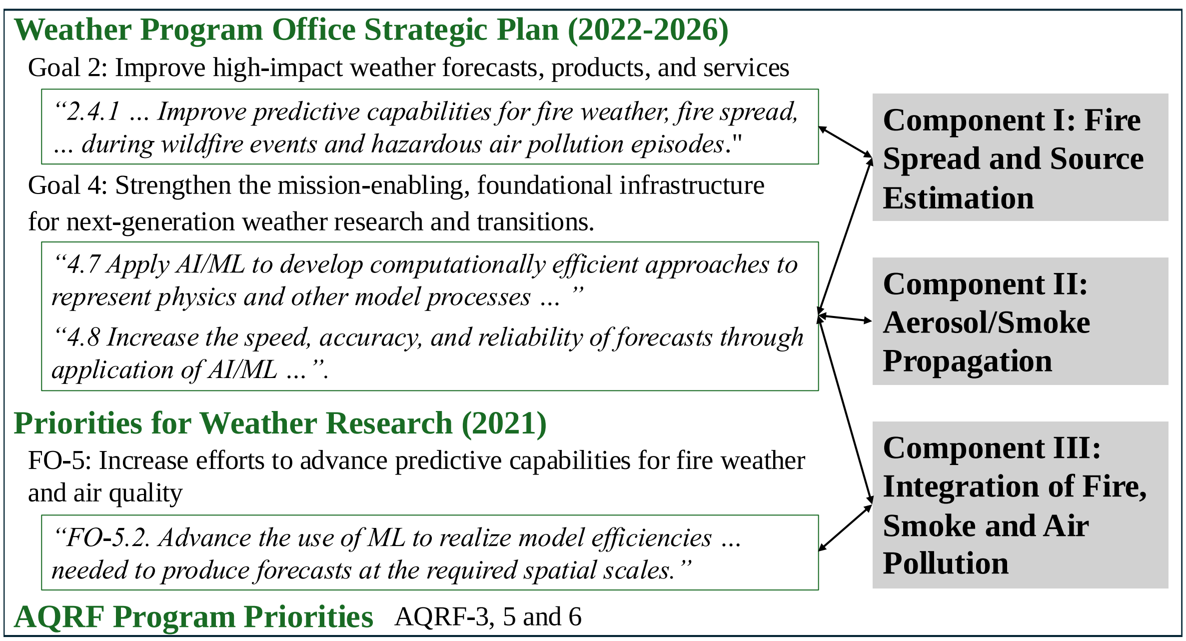

Figure 2: Mapping of the project components to NOAA's strategic priorities

1.1 Relevance to NOAA's Mission and Data Products

Fig 2 provides the mapping between this project and NOAA's missions. Over the last few centuries, the natural fire regime of regions and ecosystems across the U.S. has been disrupted by many factors, including fire exclusion, land management, and socio-economic issues. The effects of climate change (extended droughts, heat waves, increases in surface and air temperatures) have created more favorable conditions for ignition and fire spread, even in areas that would not normally burn[15, 17, 28]. Hence, fast and accurate air quality forecasting capabilities will create high societal and economic value.

NOAA's Data Products. This project will heavily utilize a large set of heterogeneous data from multiple sources, especially NOAA's data products. These data sources typically have different spatio-temporal resolutions and characteristics, and the proposed solution will enable the data fusion capabilities that integrate multi-source data into a unified ML framework. Some major datasets to be utilized are summarized as follows:

(i) **Component-I**: (a) NOAA GOES-R Fire Detection and Characterization Algorithm (FDCA) data product, including fire metadata mask, fire size, temperature and radiative power, obtained from the ABI Channels 7 (3.9μm band) and 14 (11.2μm band), (b) NOAA NCEP Global Forecast System (GFS) data, (c) NASA Global Fire WEather Database (GFWED), (d) Fuel data (e.g., Landscape Fire and Resource Management Planning Tools Project (LANDFIRE)); (e) Vegetation data (e.g., the Normalized Difference Vegetation Index (NDVI) data from NASA LANCE); and (f) Topography data (e.g., Digital Elevation Model (DEM) from NOAA); (ii) **Component-II**: NOAA GOES-R Advanced Baseline Imager (ABI) dimensionless Aerosol Optical Depth (AOD) data at a wavelength of 550nm over land and ocean; and (iii) **Component-III**: EPA's air quality monitoring data.

1.2 The Complementary Strengths of Physics-Informed Machine Learning and An Interdisciplinary Team

This section provides further rationales on why the ML system to be developed by this project can bring complementary strengths to the existing NAQFC.

• **Physics-based numerical air quality models.** Air pollution and wildfires are complex physical processes governed by the interaction of environmental processes across multiple scales[4, 18]. The fundamental idea of building a physics-based numerical air quality model aims to couple and resolve (i) a meteorological model (e.g., WRF[37] and MPAS[38]), (ii) an emission model (e.g., RAP- and HRRR-Smoke[21]), and (iii) a 3D chemistry transport model (e.g., CMAQ developed by the U.S. EPA). Additional components (e.g., GEFS-Aerosols[20], Fengsha Dust Module[1], data assimilation algorithms, etc.) can further enhance the NAQFC for specific scenarios.

[**advantages**] Physic-based numerical simulations provide a powerful way to understand the interactions between air quality and wildfires under various atmospheric conditions. The coupled governing equations are highly interpretable (with solid scientific basis) and provide actionable insights on long-term forecasting, planning, and "what-if" analysis.

[**opportunities**] (***i***) For real-time operations and decisions, computational time becomes a major obstacle that prevents us from using numerical simulation models[2] (e.g., the HRRR-Smoke is available every 6 hours); (***ii***) Because of the high computational cost, it remains a challenging task to incorporate heterogeneous multi-source streaming data that contain valuable real-time information (e.g., geostationary remote-sensing, aerial/ground observations, etc.) even with data assimilation techniques; (***iii***) Physics-based simulation models require accurate inputs for a large number of parameters (e.g., meteorological information, fuel roughness height/depth, etc.) that are rarely precisely known for real-time operations; and (***iv***) It is challenging to perform uncertainty quantification (probabilistic forecasting) that requires repeated simulations considering the computational cost.

• **Pure data-driven black-box models.** ML approaches, such as Deep Learning[10] and expert systems[32], provide complementary strengths for air quality predictions. For example, [9] and [26] employed the Convolutional Neural Networks (CNN) to generate the burn map for the 2016 Beaver Creek fire in Colorado, [19] modeled the wildfire spread process as a Markov Decision Process based on deep Reinforcement Learning (RL), and Google recently used a Convolutional Autoencoder for one-day-ahead fire prediction[10].

[**advantages**] Although the off-line training of ML models can still be computationally expensive, the on-line (probabilistic) prediction is fast enough for supporting real-time decisions. The ML approaches have the capabilities to quickly extract useful information from streaming observational data, which have become increasingly abundant, for tasks including fire/smoke spread prediction, fuel characterization, emission estimation (where physics models are either unavailable or computationally infeasible for real-time operations).

[**opportunities**] One limitation, which has received much attention in recent years, is the lack of explainability of black-box ML. We note that (***i***) air quality management involves high-stake decisions for which domain knowledge imposes critical constraints on how data should be modeled and how models can be interpreted; and (***ii***) stakeholders who fight against air pollution in the front line may not be data scientists by training. The lack of explainability and actionable insights has become a barrier that impedes the implementation (in operational environments) of the latest advances in ML for air quality and fire weather management (an increasingly *data-rich* and *domain-knowledge-intensive* domain).

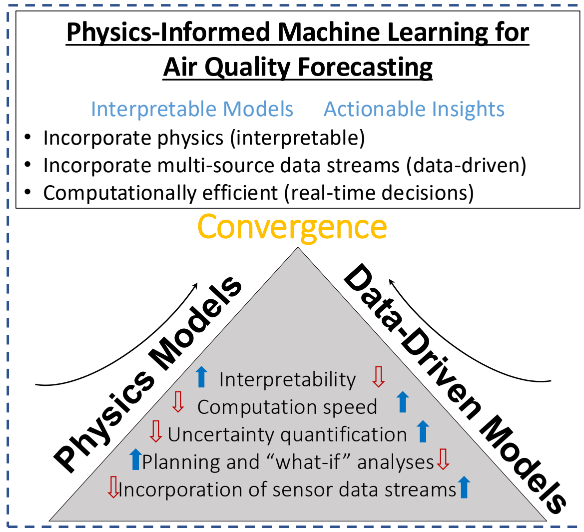

Figure 3: Harnessing physics-informed ML for real-time air quality predictions

• **The call for convergence: physics-informed ML.** Following the discussions above, the proposed solutions live at the intersection of physics-based models and data-driven approaches (see Fig.3), generating salient modeling and operational advantages: (***i***) real-time predictions capabilities, (***ii***) capabilities of fusing multi-source data streams, (***iii***) improved model interpretability by explicitly modeling the fire/smoke spread processes, and (***iv***) statistical uncertainty quantification.

Before presenting the technical details in Sec.2, it is also critical to highlight that:

> ML models, even if physics-informed or motivated by governing equations, are still essentially data-driven models. Such models will not replace existing physics-based models in understanding the science behind air pollution processes. On the other hand, in terms of generating accurate and probabilistic predictions for real-time decisions, ML has proven to be an important asset that provides the advantages listed above. By integrating governing equations and data, this project will create interpretable and accurate air quality predictions that can be quickly updated using streaming data, providing complimentary capabilities to the NAQFC toolbox.

• **An Interdisciplinary Project Team** has been formed with the required expertise in Statistics, Machine Learning, and Atmospheric Science: (***i***) PI Liu has focused on integrating governing equations into statistical models for environmental processes (relevant prior work:[12, 13, 14, 34, 35, 36, 40]), and also has a 5-year R&D experience with IBM Research (he served as the technical lead of IBM's first environment-related project with the National Environment Agency of Singapore[24]); (***ii***) co-PI Xie has expertise in high-dimensional data analysis, spatio-temporal modeling, anomaly and change-point detection, and is the Associate Director of Georgia Tech's Machine Learning Center. She will provide strong theoretical guidance on the ML models to be developed (relevant prior work:[6, 7, 8, 39, 42]); and (***iii***) co-PI Wang will bring in the essential domain knowledge on atmospheric chemistry, aerosols, and remote-sensing (relevant prior work:[15, 16, 17, 28, 43]), constituting an indispensable part for building physics-informed ML systems for air quality forecasting.

2 Methods and Activities

This section provides the technical details for constructing the 3 modules of the ML system.

2.1 A Deep Hierarchical Dynamic Model for "Fire Spread and Source Estimation" (Component I).

For Component I, this project will develop a wildfire spread model, known as the **Deep Hierarchical Dynamic Model (Deep-HDM)**, which consists of 3 layers: (i) the Observation Layer, (ii) the Fire Spread Layer, and (iii) the Parameter Layer.

• The **Observation Layer** is where the fire burning area observation enters the model. Note that, for short-term predictions, real-time physical states of an active fire can rarely be continuously and accurately monitored (for example, fire radiative power, in megawatts, is an important state that captures the amount of energy given off by a forest fire, but it is difficult to obtain real-time observations of such parameters typically affected by many other factors including weather, fire size, moisture content, plant chemistry, etc.).

Hence, this project considers a realistic scenario where only binary observations of burning areas are available from NOAA's GOES-R Series (see e.g., Fig.4). Then, to predict fire spread, a link needs to be established from the binary observations (i.e., image mask) to the underlying fire dynamics. Let $\psi(t, \boldsymbol{s})$ be the unobserved space-time fire dynamics, we consider a link function, $\text{logit}(p(t, \boldsymbol{s})) = \gamma_0 + \gamma_1 \psi(t, \boldsymbol{s})$, where $p(t, \boldsymbol{s})$ is the probability that fire spreads to location $\boldsymbol{s}$ at time t, and the *logit* function is given by $\log[p(t, \boldsymbol{s})/(1-p(t, \boldsymbol{s}))]$.

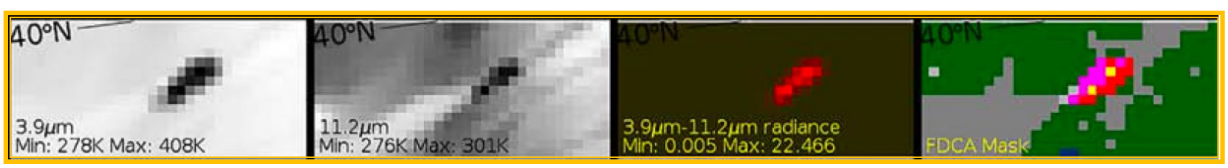

Figure 4: An illustrative example of the binary burning area data[30]. Left to right: GOES-R ABI at 3.9 and 11.2μm bands, radiance difference between the 3.9 and 11.2μm bands, and the Fire Detection and Characterization Algorithm (FDCA) data mask for Camp Fire, 2018.

• The **Fire Spread Layer** captures the spatio-temporal fire dynamics $\psi(t, \boldsymbol{s})$. The proposed Deep-HDM will treat this layer as "hidden" that needs to be estimated and continuously updated from the binary observations of fire burning area, as well as other factors including weather, fuel, topography, etc. In particular, we invoke the level set equation (a stochastic Partial Differential Equation, sPDE) which allows a submesh representation of the burning region and a flexible implementation of various kinds of ignition[23]:

$$\dot{\psi}(\boldsymbol{s}, t) = S||\nabla\psi(\boldsymbol{s}, t)|| + e(\boldsymbol{s}, t), \tag{1}$$

where $\psi(\boldsymbol{s}, t)$ is a level-set function (i.e., the burning region at time t is represented as the set of all spatial points $\boldsymbol{s}$ such that $\psi(\boldsymbol{s}, t) \leq 0$), S is the fire spread rate which depends on environmental factors and needs to be estimated from data (e.g., the modified Rothermel formula[27]), $e(\boldsymbol{s}, t)$ captures the uncertainty associated with the process, and ∇ is the gradient operator. The level set equation (1) has been adopted in the WRF-SFIRE model[18], and a recent statistical fire spread model[41].

Unlike WRF-SFIRE, this project will *not* pursue the direct execution of the sPDE (1). Instead, the Deep-HDM will estimate the hidden dynamics $\psi(t, \boldsymbol{s})$ from the fire burning area information. Leveraging the spectrum decomposition (eigenfunction expansion), we represent the PDE (1) by a statistical dynamic model[14]:

$$\psi(t, \boldsymbol{s}) \approx \tilde{\psi}(t, \boldsymbol{s}) = \boldsymbol{\Phi}^T(\boldsymbol{s})\boldsymbol{\alpha}(t), \quad \dot{\boldsymbol{\alpha}}(t) = \mathcal{G}(t, \boldsymbol{\alpha}(t), \boldsymbol{\theta}_P, \Omega) \tag{2}$$

where $\boldsymbol{\alpha}(t) = (\alpha_1(t), ..., \alpha_K(t))^T$ is a **multivariate stochastic process** in time whose dynamics is described by a system of Stochastic Differential Equations (SDE) characterized by $\mathcal{G}$ (to be found by this project), $\boldsymbol{\theta}_P$ is the parameter, Ω captures the randomness, $\phi_j(\boldsymbol{s})$ is the spatial basis function, and $\boldsymbol{\Phi}(\boldsymbol{s}) = (\phi_1(\boldsymbol{s}), ..., \phi_K(\boldsymbol{s}))^T$. The spatial basis function $\phi_j(\boldsymbol{s})$ can be either directly specified (e.g., Fourier basis[5, 14, 31]) or learned from simulation/historical data (e.g., Proper Orthogonal Decomposition (POD) basis[25]).

• The **Parameter Layer** sits on top of the Hidden Fire Spread and Observation layers, and contains the unknown model parameters to be estimated, including (i) a collection of parameters $\boldsymbol{\theta}_P$ in (2) that define the mapping between environmental factors (e.g., wind, fuel, temperature, topography) and fire spread rate; and (ii) the parameter, $\boldsymbol{\theta}_L = (\lambda_0, \lambda_1)$, in the logit link function. This layer also allows us to impose necessary regularizations and useful prior knowledge on these model parameters through a Bayesian framework.

• **Deep Hierarchical Dynamic Model (Deep-HDM).** Integrating the three layers above leads to the **Deep-HDM** to be developed by this project (see Fig.5):

$$\begin{aligned}
&\text{Observation Layer: } [\ \underbrace{\boldsymbol{Y}(t)}_{\text{binary fire mask}}\ ;\ \underbrace{\boldsymbol{\xi}(t)}_{\text{hidden state}}\ ,\ \underbrace{\boldsymbol{\theta}_D}_{\text{parameters}}\ ] \\
&\text{(Hidden) Fire Spread Layer: } [\ \underbrace{\boldsymbol{\alpha}(t)}_{\text{coefficient}}\ ; \boldsymbol{\alpha}(t-\Delta),\ \underbrace{\boldsymbol{\beta}(t)}_{\text{bias correction}}\ ,\ \underbrace{\boldsymbol{\theta}_P}_{\text{parameters}}\ ] \\
&\text{Parameter Layer: } [\boldsymbol{\theta}_D, \boldsymbol{\theta}_P, \boldsymbol{\theta}_L]
\end{aligned} \tag{3}$$

where $[\cdot]$ represents probability distribution, and $\boldsymbol{Y}(t)$ is a vectorized binary observations on burning areas time t from all spatial locations. The proposed Deep-HDM provides a list of salient **modeling advantages**: (i) integrate heterogeneous multi-channel GOES-R observations, (ii) incorporate prior knowledge and sparse learning through the parameter layer, (iii) quantify uncertainty by updating the posterior distributions of model parameters and predictions, (iv) handle missing data through the hierarchical structure, and (v) enable fast computation using Monte-Carlo type of algorithms and support real-time operations.

To realize the potential advantages of the Deep-HDM framework described above, this project will develop both the **off-line** procedure for model training using historical data, and the **on-line** updates of model parameters using real-time data streams.

• **[Off-line estimation]**. The off-line procedure will leverage the following state-space representation of the Deep-HDM modeling framework shown in (3):

$$\text{logit}(p(t, \boldsymbol{s})) = \boldsymbol{F}\boldsymbol{x}(\boldsymbol{t}), \qquad \boldsymbol{x}(\boldsymbol{t}) = \boldsymbol{G}_{\boldsymbol{\Omega}(\boldsymbol{t})}\boldsymbol{x}(t-\Delta) + \boldsymbol{\epsilon}(t) \tag{4}$$

Here, we omit the technical details of how (4) can be derived from (3) due to the page limit.

In summary, $p(t, \boldsymbol{s})$ is the probability that fire spread to location $\boldsymbol{s}$ at time t, $\boldsymbol{F}$ is the augmented spatial basis matrix obtained from $\boldsymbol{\Phi}$ in (2), $\boldsymbol{x}(t)$ is the stochastically-evolving augmented hidden state of wildfire (including both fire propagation and fire source) obtained from $\boldsymbol{\alpha}(t)$ and $\boldsymbol{\beta}(t)$ in (3), $\boldsymbol{G}_{\boldsymbol{\Omega}(t)}$, which is unknown and dependent on environmental factors $\boldsymbol{\Omega}(t)$ (e.g., wind, topography, fuel, etc.), determines the dynamics of the augmented state vector $\boldsymbol{x}(t)$, $\boldsymbol{\epsilon}(t) \sim N(\boldsymbol{0}, \boldsymbol{W}(t))$ is a random vector that captures the uncertainty associated with the state transition, and Δ is the computational time interval.

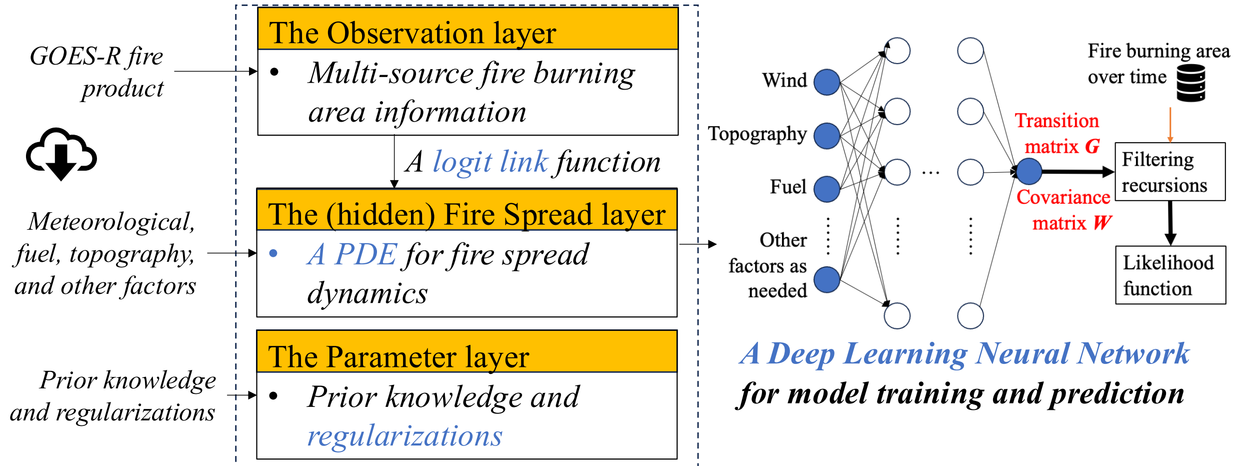

Figure 5: A high-level illustration of the proposed Deep-HDM

Hence, the main task is to learn the mapping from the environmental factors $\boldsymbol{\Omega}(t)$ to the transition matrix $\boldsymbol{G}_{\boldsymbol{\Omega}(t)}$, and estimate the covariance matrix $\boldsymbol{W}(t)$ given observed fire burning areas over time. In particular, let $\boldsymbol{y}(t)$ be a vector, with binary elements, that represents the fire burning area at time t, the joint likelihood of $\boldsymbol{G}$ and $\boldsymbol{W}$ is given by $L(\boldsymbol{G}, \boldsymbol{W}; \boldsymbol{y}(1), \boldsymbol{y}(2), \cdots, \boldsymbol{y}(T)) = \prod_{t=1}^{T} \pi(\boldsymbol{y}(t)|\boldsymbol{y}(t-1); \boldsymbol{G}, \boldsymbol{W})$. To obtain the conditional distribution $\pi(\boldsymbol{y}(t)|\boldsymbol{y}(t-1); \boldsymbol{G}, \boldsymbol{W})$, we will derive the full **Filtering Recursions** for the state-space model (4): (i) the one-step-ahead predictive density for $\boldsymbol{x}(t)$, i.e., $\pi(\boldsymbol{x}(t)|\boldsymbol{y}(t-1)) = \int \pi(\boldsymbol{x}(t)|\boldsymbol{x}(t-1))\pi(\boldsymbol{x}(t-1)|\boldsymbol{y}(t-1))d\boldsymbol{x}(t-1)$; (ii) the one-step-ahead predictive density for $\boldsymbol{y}(t)$, i.e., $\int \pi(\boldsymbol{y}(t)|\boldsymbol{x}(t))\pi(\boldsymbol{x}(t)|\boldsymbol{y}(t-1))d\boldsymbol{x}(t)$ (Laplace approximation is needed); and (iii) the filtering distribution for $\boldsymbol{x}(t)$, i.e., $\pi(\boldsymbol{x}(t)|\boldsymbol{y}(t)) \propto \pi(\boldsymbol{y}(t)|\boldsymbol{x}(t))\pi(\boldsymbol{x}(t)|\boldsymbol{y}(t-1))$ (approximations are also needed for computational efficiency). Finally, the multi-layer Artificial Neural Networks (ANNs) (see Fig.5) is used to estimate the state transition matrix $\boldsymbol{G}_{\boldsymbol{\Omega}(t)}$ (which determines both fire spread and source) from environmental factors $\boldsymbol{\Omega}(t)$, and therefore, allow us to use the state-space model (4) as a fast forward predictive model for fire propagation based on real-time environmental data streams.

• [**On-line update**]. To provide more accurate real-time predictions, the Deep-HDM model needs to be adapted to particular wildfire events by allowing on-line update of the model utilizing real-time streaming data. Given fire burning area data up to time T, i.e., $\boldsymbol{y}_{1:T} = (\boldsymbol{y}(1), \boldsymbol{y}(2), \cdots, \boldsymbol{y}(T))$, we will develop a fast online algorithm that (***i***) updates the estimated model parameters, $\boldsymbol{G}$ and $\boldsymbol{W}$, and (***ii***) re-constructs all state variables $\boldsymbol{x}_{0:T} = (\boldsymbol{x}(0), \boldsymbol{y}(1), \cdots, \boldsymbol{x}(T))$ up to time T:

$$\pi(\boldsymbol{x}_{0:T}, \boldsymbol{G}, \boldsymbol{W}|\boldsymbol{y}_{1:T}) = \pi(\boldsymbol{x}_{0:T}|\boldsymbol{G}, \boldsymbol{W}, \boldsymbol{y}_{1:T})\pi(\boldsymbol{G}, \boldsymbol{W}|\boldsymbol{y}_{1:T}) \quad (5)$$

This project will develop **sequential Monte Carlo on-line algorithms**, tailored for the proposed Deep-HDM (3), to obtain the joint posterior density (5) of $\boldsymbol{G}$ and $\boldsymbol{W}$

and the state variables $\boldsymbol{x}_{0:T}$. When a new observation $\boldsymbol{y}(T+1)$ arrives, the algorithm only adjusts the target distribution $\pi(\boldsymbol{x}_{0:T+1}, \boldsymbol{G}, \boldsymbol{W}|\boldsymbol{y}_{1:T+1})$ from $\pi(\boldsymbol{x}_{0:T}, \boldsymbol{G}, \boldsymbol{W}|\boldsymbol{y}_{1:T})$ by the estimation-error correction given by the filtering recursions, without doing all the computations again. This is a great advantage provided by the hierarchical structure of the proposed Deep-HDM (3), which allows efficient on-line simulation-based sequential updating of the posterior distribution of wildfire burning area.

2.2 Smoke Propagation (Component II)

Leveraging the (i) output generated from the fire spread model in Component-I, and (ii) NOAA GOES-R Aerosol Optical Depth (AOD) product, Component-II will develop a dynamic model for predicting the aerosol propagation using multi-source data streams.

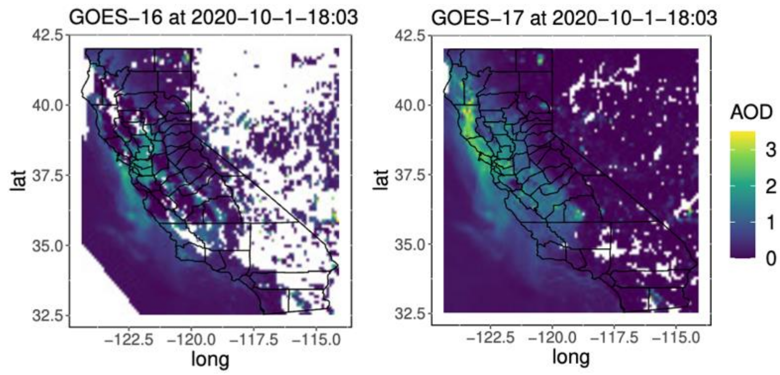

Figure 6: An illustrative example of AOD (at the same time over the same spatial area) from GOES-16 and GOES-17.

The Advanced Baseline Imager (ABI) on board the GOES-R satellites generates remote-sensing AOD data for the continental U.S. (CONUS) every 5 minutes, enabling the real-time modeling and prediction of aerosol processes. For illustrative purposes only, Fig.6 shows the AOD from GOES-16 and -17 over the West Coast (*GOES-18 will be used in this project instead of GOES-17). It is noted that, although the two images are taken at the same time over the same spatial area, they clearly present heterogeneous characteristics (including data missing rates, AOD values, bias, etc.).

Although the difference is mainly because the two satellites operate in different regions (East v.s. West), both data streams provide useful information about the underlying aerosol process (even if one source could be more reliable than the other). In fact, heterogeneity is a common challenge in modeling multi-source remote-sensing data streams, which provide imperfect observations of the same underlying physical process (e.g., aerosol propagation). There exist both practical needs and theoretical interests to simultaneously utilize (or fuse) multi-source remote-sensing data streams, so that the underlying aerosol propagation process can be accurately predicted. This challenge motivates us to develop a spatial-temporal dynamic model for wildfire aerosols propagation process, by utilizing (i) multi-source remote-sensing data streams (for the same underlying physical process), (ii) fire spread predictions from Component-I, and (iii) the advection-diffusion equation that governs the fundamental dynamics of smoke propagation processes.

• **The Modeling Details.** The ML model will consider a (stochastic) scalar transport equation for the dispersion of air-borne pollutants:

$$\dot{\xi}(\boldsymbol{s},t) = Q(\boldsymbol{s},t) - \nabla \cdot [\boldsymbol{v}(\boldsymbol{s},t)\xi(\boldsymbol{s},t)] + \nabla \cdot [\boldsymbol{K}(\boldsymbol{s},t) \cdot \nabla\xi(\boldsymbol{s},t)] - \tau^{-1}\xi(\boldsymbol{s},t) + \epsilon(\boldsymbol{s},t) \quad (6)$$

where ξ is the AOD at location $\boldsymbol{s}$ and time t, Q is the emission rate (to be linked to fire spread model in Component-I), $\boldsymbol{v}$ is the velocity vector, $\boldsymbol{K}$ is a second-order tensor of the eddy diffusivity, τ is the relaxation timescale of pollution removal, and $\epsilon(\boldsymbol{s},t)$ is a space-time random noise. In the transport equation (6), advection describes the transport mechanism of a pollutant due to the flow of air, the diffusion process describes the mixing of the pollutant by local turbulent flow, and $\tau^{-1}\xi(\boldsymbol{s}, t)$ models the physical removal process (e.g., deposition) of the air-borne pollution.

In our prior work [13, 14] (details not included due to the page limit), the transport equation (6) can be modeled by a spatio-temporal random field $\{Y(\boldsymbol{s}, t); \boldsymbol{s} \in \mathbb{R}^d, t \in \mathbb{R}\}$ which is continuous in both space and time:

$$Y(\boldsymbol{s}, t) = \int_0^\infty \exp(-\lambda u)\phi(\boldsymbol{x}; \bar{\boldsymbol{\mu}}_u, \bar{\boldsymbol{\Sigma}}_u) * \{Q(\boldsymbol{s}, t-u) + dB(\boldsymbol{x}, t-u)\}du \tag{7}$$

where $\bar{\boldsymbol{\mu}}_u = \int_0^u \boldsymbol{R}(\boldsymbol{s}, t-x)\boldsymbol{v}(\boldsymbol{s}, t-x)dx$, $\bar{\boldsymbol{\Sigma}}_u = \int_0^u \boldsymbol{R}(\boldsymbol{s}, t-x)\boldsymbol{\Sigma}(\boldsymbol{s}, t-x)dx$, and $B(\boldsymbol{s}, t)$ is a spatially correlated Brownian motion. The right-hand-side of (7) explicitly captures how the emission $Q(\boldsymbol{s}, t)$ over space and time affects the aerosol process, while the uncertainty associated with the process is captured by $dB(\boldsymbol{x}, t-u)$. Hence, the emission term $Q(\boldsymbol{s}, t)$ provides **the link** needed between the random field (7) and the fire spread model developed in Component-I. In other words, if fire has spread to location $\boldsymbol{s}$ at time t, $Q(\boldsymbol{s}, t)$ is a function of the fuel information which can be learned from data.

Following our latest work [34, 36] and applying the Fourier Transform to (6), we approximate the random field $Y(\boldsymbol{s}, t)$ by $Y(\boldsymbol{s}, t) \approx \sum \alpha_{\boldsymbol{k}}(t) f_{\boldsymbol{k}}(\boldsymbol{s})$, where $f_{\boldsymbol{k}}(\boldsymbol{s})$ is the eigenfunction (e.g., the Fourier bases) and $\alpha_{\boldsymbol{k}}(t)$ determines the random weighting of each mode. Then, a dynamic model can be constructed for heterogeneous multi-satellite AOD observations:

$$\begin{aligned} &\text{Multi-satellite data streams: } [\{\boldsymbol{Y}_j(t)\}_{j=1}^J; \xi(t), \boldsymbol{\theta}_D] \\ &\text{Dynamics of the advection-diffusion process: } [\boldsymbol{\alpha}(t); \boldsymbol{\alpha}(t-\Delta), \boldsymbol{\beta}(t), \boldsymbol{\theta}_P] \end{aligned} \tag{8}$$

where $[\cdot]$ represents probability distribution, $\boldsymbol{Y}_j(t)$ be the observations at time t from satellite j, and $\boldsymbol{\alpha}(t) = (\alpha_1(t), \cdots, \alpha_K(t))$ are the Fourier coefficients that determine the dynamics of the random field (7). It is noted that the multi-satellite data streams $\{\boldsymbol{Y}_j(t)\}_{j=1}^J$ in (8) are often correlated in both space and time. Because the conventional stationary covariance kernel functions are known to be inadequate in scenarios where heterogeneous correlations dominate the input space in cases of longitudinal and spatial data, this project will follow and extend our recent work [6], and model the multi-satellite data streams $\{\boldsymbol{Y}_j(t)\}_{j=1}^J$ by a multi-source spatio-temporal Gaussian process with a specially designed covariance kernel function, where the time, space, and data source are mutually independent and each factor describes the AOD characteristics within distinct domains (this approach has been proven effective in our prior work for wind power generation prediction[6]).

The dynamic model (8), coupled with the kernel design approach above, provides the critical capability that naturally utilizes multi-source observations, with possibly multi-level spatio-temporal resolutions, to estimate the underlying AOD process (i.e., to obtain the posterior distribution, $[\boldsymbol{\alpha}(t), \boldsymbol{\beta}(t), \boldsymbol{\theta}; \{\boldsymbol{Y}_j(t)\}_{j=1}^J])$) using the Gibbs sampler with the Forward Filtering and Backward Sampling (FFBS) to be developed by this project.

2.3 "Air quality" prediction (Component III)

The completion of Components-I and II will naturally lead to a straightforward but effective approach for predicting air quality (PM2.5) by integrating ground monitoring data. In particular, an **infinite-dimensional dynamic model** will be utilized to link the aerosol propagation (generated by Components-I and II) and pollutant concentrations:

$$\dot{\xi}(t, \boldsymbol{s}) = \mathcal{A}\xi(t, \boldsymbol{s}) + \epsilon(t, \boldsymbol{s}), \quad \boldsymbol{c}_k = \mathcal{H}_k \xi(t, \boldsymbol{s}) + \boldsymbol{e}_k \tag{9}$$

where $\xi(t, \boldsymbol{s})$ still represents the aerosol process (whose prior distribution is available from Component-II), $\boldsymbol{c}_k$ is a vector that contains the observed pollutant concentrations at time step k from EPA's monitoring stations, and $\mathcal{H}_k$ is an operator (to be estimated) that links the aerosol process to pollutant concentrations. Hence, the well-established infinite-dimensional Kalman filter, smoother and predictor in [22, 29] can be employed to predict the future state of the system (9) and pollutant concentrations. Finally, Fig.7 presents a high-level summary of how the three Components-I, II and III are coupled to create the ML system for air quality predictions during wildfire episodes.

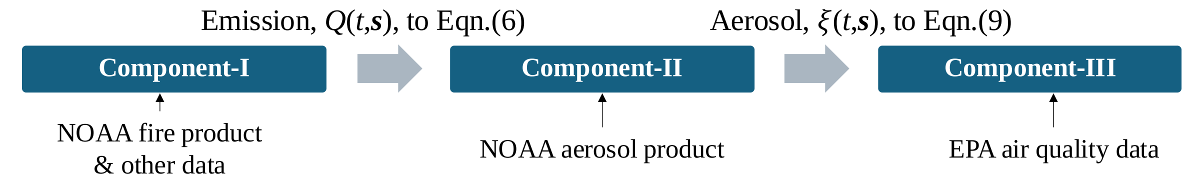

Figure 7: Coupling of Components-I, II and III for air quality prediction

2.4 Validation

The team will validate the performance (both accuracy and computational time) of the proposed ML system for at least **10** major air pollution episodes selected from the 2027 and 2028 California Wildfire Seasons, peaking between late Spring and Fall.

The validation will test all three Components of the ML system, including the predicted (i) fire burning area, (ii) aerosol propagation, and (iii) pollutant (PM2.5) concentrations. To evaluate the accuracy of the predicted (binary) fire burning area, we will report precision, recall, and AUC (area under the curve) as the performance metrics (also used by a recent study performed by a Google team[10]). To comprehensively evaluate the accuracy of the predicted AOD and air quality, we will report a set of 9 performance metrics adopted by the "openair" package of R—an open-source package for analyzing air pollution data[3]. These metrics include the Mean Gross Error, Normalized Mean Bias, Index of Agreement, Fraction of Predictions within a Factor or Two, etc. All performance metrics will be reported for 1-hour-ahead prediction up to 24-hour-ahead prediction, and compared to the performance of the predicted values (at locations where ground observations are available) available on the Air Quality Forecast Guidance Viewer supported by NOAA and National Weather Service. The comparison will enable us to quantify the improvement in air quality prediction brought by the developed ML system. Finally, computational time will also be reported to demonstrate the real-time predictive capabilities of the ML system.

3 Products/Outputs

A software package will be developed, in Python environment, to facilitate potential implementation and NOAA's R2O transitions. The package will adopt the basic structure of IBM's Environmental Intelligence Suite[11] and consist of 3 modules: (i) Module-I (**Data APIs**) provides data querying, extraction and preparation for user-specified spatial domains and time windows; (ii) Module-II (**ML System**)

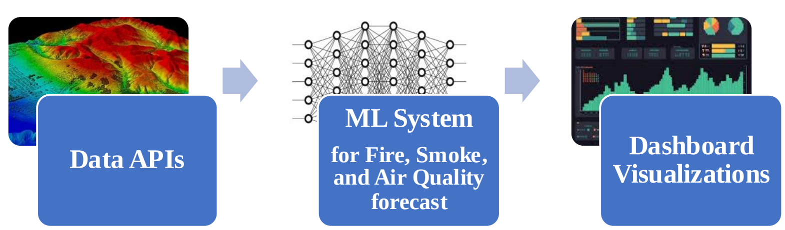

Figure 8: A software package for potential users

is the analytics engine that runs the ML algorithms developed by this project, and (iii) Module-III (**Dashboard Visualizations**) presents the forecasting results on an interactive map so that users can easily access the results generated by the ML system. The software is able to automatically update the hourly air quality predictions every 10 minutes.

Other products include: (i) ML **algorithms** for fire spread, smoke propagation and low-latency hourly regional air quality predictions; (ii) annual and final **technical reports** that document the technical details and validation results; (iii) **training materials** (user manuals) for the ML algorithms/system/software; and (iv) five **journal publications** and six **conference presentations** to disseminate the outcomes with the scientific communities and general public (Readiness Levels are provided in Sec.5).

4 Impacts, Benefits, Outcomes, & Recipients

Climate change has caused an increase in the frequency and severity of extreme wildfires. This project responds to the Weather Program Office Strategic Goals 2&4, FO-5 of the Priorities for Weather Research, and AQRF-3, 5 and 6 (as summarized in Sec.1.1).

The physics-informed ML system will generate complementary capabilities for enhancing the current NAQFC by providing accurate and real-time air quality predictions during wildfire events. Note that, our goal is *not* to replace the current physics/chemistry models by ML methods, but to add an additional useful tool to the existing NAQFC toolbox for interpretable, accurate, and fast air quality predictions that can be frequently updated using multi-source streaming data. This capability is essential for enhancing the situational awareness and emergency response capabilities of Federal and local agencies in making real-time decisions. As a result, the ML system will facilitate the delivery of services and values—by NOAA and other agencies such as EPA—to the public so as to limit the harmful health effects of poor air quality during wildfire episodes.

Once the software package has been developed and the potential value of the proposed ML system has been fully validated, the team will actively support NOAA's Research-to-Operation (R2O) transitions, and pursue opportunities to participate in other NOAA programs such as the Testbed Programs. Finally, by involving a diverse body of researchers and students, we expect to promote the integration of ML into fire weather research and environmental awareness, helping to achieve NOAA's long-term strategic goals.

5 Schedule with Key Milestones

This section presents the timelines, key milestones, and how the Readiness Level (RL) matures over milestones; see Fig.9.

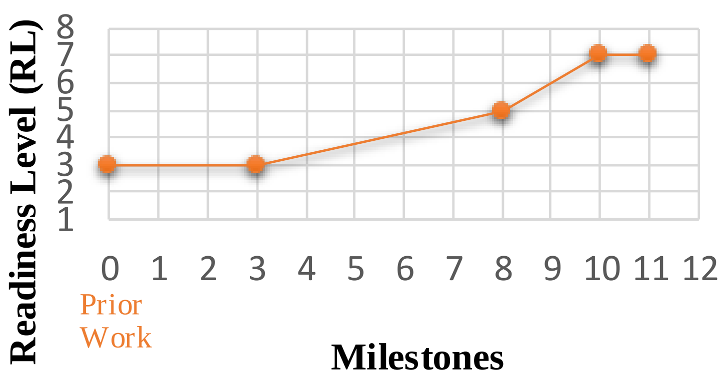

Figure 9: Readiness Level (RL) growth over milestones

• **Phase-I**: Data Preparation and GRA Recruitment [08/2025-01/2026]

*Milestone 1: all training (historical) data will be prepared, and the data preparation procedure will be automated before 01/2026.

*Milestone 2: All GRAs will be recruited before 01/2026.

• **Phase-II**: Model Development Phase [08/2025-07/2027]

*Milestone 3: Component-I of the ML system will be developed by 07/2026.

*Milestone 4: Year 1 Annual Report due in 07/2026.

*Milestone 5: Component-II of the ML system will be developed by 01/2027.

*Milestone 6: Component-III of the ML system will be developed by 07/2027.

*Milestone 7: Year 2 Annual Report due in 07/2027.

• **Phase-III**: Validation, Software and Testing [08/2027-07/2028]

*Milestone 8: validation of the ML system will be completed and the testing results will be reported by 04/2028.

*Milestone 9: software development will be completed by 04/2028.

*Milestone 10: live demonstration of the ML system and software tools by 07/2028.

*Milestone 11: Final Project Report by 07/2028.

6 Outreach and Education

• **Education**. (i) Project outcomes will be integrated into the Georgia Tech curriculum (in both Colleges of Engineering and Sciences). PI Liu is developing a course, "Domain-Aware Statistical Learning," and air quality prediction will be converted into a course project. Co-PI Xie has been teaching graduate level "Computational Statistics" and will integrate suitable research topics as course components. Co-PI Wang can naturally integrate the project outcomes into his course, "Air Pollution Meteorology". (ii) Three PhD students will be trained by this project. Developing the proposed ML system will constitute a major part of their PhD dissertations, potentially re-shaping their future career path. These activities will collectively promote awareness of extreme natural events among a large pool of students, and attract many of them into the field of extreme weather research.

• **Outreach**. The team will provide timely progress updates and seek feedback/suggestions from NOAA Program Managers, which will enable our team to quickly identify and reach out to potential collaborators/communities through the existing operations of NOAA. Technical details and validation results will be published on journals and presented at conferences. The team will also leverage the well-established outreach programs of Georgia Tech to promote the project outcomes and collect feedback for improvement. For example, the Brook Byers Institute for Sustainable Systems (BBISS) of Georgia Tech, in collaboration with the Center for Sustainable Communities Research and Education (SCoRE), has already built community-engaged programs, through which our team can engage more faculty, students, and community partners in long-term research-education collaborations.

7 Diversity, Equity, Inclusion, and Accessibility (DEIA)

The team resonates with NOAA's value in DEIA, and will foster DEIA as core to our project success. We have formed an interdisciplinary team (see Sec.1.2), which will be fully committed to increasing inclusive work environments where everyone feels valued. The team will recruit and retain a diverse and highly skilled workforce (e.g., Graduate Research Assistants from underrepresented groups) by providing strong and equal support, including the opportunities to attend research conferences, productive and healthy research discussions, fair recruitment and retention processes transparent to the diverse communities. Finally, to receive professional guidance, the team plans to collaborate with the Institute Diversity, Equity, and Inclusion team at Georgia Tech to assess our team dynamics and organize diversity review of the proposed activities during the course of this project.

Data Management Plan

In accordance with NOAA's policy on the dissemination and sharing of research results, the team will make computer code, data sets (excluding the proprietary/confidential data from the participating networks), model outputs and publications publicly accessible.

- **Data Sharing Protocol & Guidance (P&G)**

The team will establish the Protocol & Guidance (P&G) for data sharing and management to reduce collaboration barriers, and ensure that the activities are compliant with relevant legislation as well as the policies and procedures of participating researchers/students. The P&G will define the following areas: the types of data to be used/shared across the team, the security and format of the data, the purpose of the research, the roles and responsibilities of each member, the condition of use of the data generated (both before and after it has been initially created).

- **A Summary of Major Datasets, Data Storage and Dissemination**

This project is expected to generate the following three major datasets:

◇ Category A (existing data shared among participating researchers/students)
◇ Category B (additional data generated by the project activities)
◇ Category C (research papers and technical reports)

The sharing of any datasets under Category A must be compliant with the policies and procedures of the participating members (details will be finalized once the data sharing P&G has been established). The participating members who share the data will have the authority to determine the appropriate approaches to share their data.

Datasets under Category B and C will be made as accessible as possible to the public in accordance with the NOAA's policy on the dissemination and sharing of research results. The data dissemination plan is given as follows:

◇ ALL developed methodologies will be published in technical reports, peer-reviewed journals and conference proceedings. In particular, ALL computer code and data (excluding the proprietary and confidential datasets) will be published as supplementary information to the published article. The published articles will clearly describe how these data sets can be accessed and used to reproduce the reported results.

◇ The papers and associated software will be accessible by the public on GitHub.

◇ Research performed by students involved in this project will be documented (e.g., in their dissertations) with detailed descriptions on how/what data sets are used to generate the results reported in the documents.

◇ Curriculum materials, including lecture slides, numerical examples and lecture videos, will be hosted on the project website and made available to the public.

Re-distribution of the data will not be allowed without permission.

- **Software and Computer Code**

All software and computer code will be made available on GitHub under the GNU General Public License (GPL) version 2 (http://www.gnu.org/licenses/old-licenses/gpl-2.0.txt). This is an open-source license, which permits others to modify the source code, share modifications with the community, and continue developing the software.

- **Data Retention**

The retention period of datasets under Category B and C will be at least five years after the project and a longer retention period is expected.

References

[1] Baker, B. (2023). The history and development of the NOAA FENGSHA dust emission scheme. *UFS Webinar*.

[2] Bottero, L., Calisto, F. Graziano, G., Pagliarino, V., Scauda, M., Tiengo, S., and Azeglio, S. (2020). Physics-informed machine learning simulator for wildfire propagation. *arXiv:2012.06825*.

[3] Carslaw, D. (2019). The openair manual open-source tools for analysing air pollution data. *https://davidcarslaw.com/files/openairmanual.pdf*.

[4] Coen, J. M., Cameron, J., Michalakes, E., Patton, P., Riggan, P. J., and Yedinak, K. (2013). Wrf-fire: Coupled weather-wildland fire modeling with the weather research and forecasting model. *Journal of Applied Meteorology and Climatology*, 52:16–38.

[5] Cressie, N. and Wikle, C. (2011). *Statistics for Spatio-Temporal Data*. John Wiley & Sons, Hoboken, New Jersey.

[6] Dong, Z., Cheng, X. Y., and Xie, Y. (2022). Spatio-temporal point processes with deep non-stationary kernels. *arXiv:2211.11179*.

[7] Dong, Z., Repasky, M., Cheng, X. Y., and Xie, Y. (2023a). Deep graph kernel point processes. *NeurIPS (Spotlight Paper at Temporal Graph Learning Workshop)*.

[8] Dong, Z., Zhang, H., Zhu, S. X., Xie, Y., and Van Hentenryck, P. (2023b). Multi-resolution spatio-temporal prediction with application to wind power generation. *arXiv:2108.13285*.

[9] Hodges, J. L. and Lattimer, B. Y. (2019). Wildland fire spread modeling using convolutional neural networks. *Fire Technology*, 55:2115–2142.

[10] Huot, F., Hu, R. L., Goyal, N., Sankar, T., Ihme, M., and Chen, Y.-F. (2022). Next day wildfire spread: A machine learning dataset to predict wildfire spreading from remote-sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13.

[11] International Business Machine (2024). Environmental Intelligence Suite. *https://www.ibm.com/docs/en/environmental-intel-suite*.

[12] Liu, X. and Yeo, K. M. (2023). Inverse models for estimating the initial condition of spatio-temporal advection-diffusion processes. *Technometrics*, 65:432–445.

[13] Liu, X., Yeo, K. M., Hwang, Y. D., Singh, J., and Kalagnanam, J. (2016). A statistical modeling approach for air quality data based on physical dispersion processes and its application to ozone modeling. *Annals of Applied Statistics*, 10:756–785.

[14] Liu, X., Yeo, K. M., and Lu, S. Y. (2022). Statistical modeling for spatio-temporal data from stochastic convection-diffusion processes. *Journal of American Statistical Association*, 117:1482–1499.

[15] Liu, Y., Qian, Y., Rasch, P. J., Zhang, K., Leung, L.-R., Wang, Y., Wang, M., Wang, H., Huang, X., and Yang, X.-Q. (2024). Fire–precipitation interactions amplify the quasi-biennial variability in fires over southern Mexico and Central America. *Atmospheric Chemistry and Physics*, 24(5):3115–3128.

[16] Liu, Y., Zhang, K., Qian, Y., Wang, Y., Zou, Y., Song, Y., Wan, H., Liu, X., and Yang, X.-Q. (2018). Investigation of short-term effective radiative forcing of fire aerosols over North America using nudged hindcast ensembles. *Atmospheric Chemistry and Physics*, 18(1):31–47.

[17] Liu, Y. Q., Liu, Y., Fu, J., Yang, C., Dong, X., Tian, H. Q., Tao, B., Yang, J., Wang, Y. H., Zou, Y. F., and Ke, Z. M. (2021). Projection of future wildfire emissions in western usa under climate change: contributions from changes in wildfire, fuel loading and fuel moisture. *International Journal of Wildland Fire*, 31:1–13.

[18] Mandel, J., Beezley, J. D., and Kochanski, A. K. (2011). Coupled atmosphere-wildland fire modeling with wrf 3.3 and sfire 2011. *Geoscientific Model Development*, 4:591–610.

[19] Morgan A. Crowley, Jeffrey A. Cardille, J. C. W. and Wulder, M. A. (2019). Multi-sensor, multi-scale, bayesian data synthesis for mapping within-year wildfire progression. *Remote Sensing Letters*, 10(3):302–311.

[20] NOAA-Global Systems Laboratory (2024). GEFS-Aerosols. *https://gsl.noaa.gov/focus-areas/atmospheric-chemistry-and-smoke-prediction/gefs-aerosols*.

[21] NOAA Research (2024). HRRR-Smoke Graphics. *https://rapidrefresh.noaa.gov/hrrr/HRRRsmoke/*.

[22] Omatu, S. and Seinfeld, J. H. (1989). *Distributed Parameter Systems: Theory and Applications*. Clarendon Press.

[23] Osher, S. and Fedkiw, R. (2003). *Level Set Methods and Dynamic Implicit Surfaces*. Springer.

[24] PR Newswire (2012). Singapore's National Environment Agency Collaborates With IBM To Improve Environmental Services Related To Air, Weather and Public Health. *https://www.prnewswire.com/news-releases/singapores-national-environment-agency-collaborates-with-ibm-to-improve-environmental-services-related-to-air-weather-and-public-health-161201205.html*.

[25] Qian, E., Karamer, B., Peherstorfer, B., and Willcox, K. (2020). Lift&learn: Physics-informed machine learning for large-scale nonlinear dynamical systems. *Physica D: Nonlinear Phenomena*, 406:132401.

[26] Radke, D., Hessler, A., and Ellsworth, D. (2019). Firecast: Leveraging deep learning to predict wildfire spread. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 4575–4581. International Joint Conferences on Artificial Intelligence Organization.

[27] Rothermel, R. C. (1972). A mathematical model for predicting fire spread in wildland fires. *USDA Forest Service Research Paper, https://www.fs.usda.gov/research/treesearch/32533*.

[28] Sarangi, C., Qian, Y., Leung, L. R., Zhang, Y., Zou, Y., and Wang, Y. (2023). Projected increases in wildfires may challenge regulatory curtailment of $pm_{2.5}$ over the eastern us by 2050. *Atmospheric Chemistry and Physics*, 23(2):1769–1783.

[29] Sarkka, S. and Hartikainen, J. (2012). Infinite-dimensional kalman filtering approach to spatio-temporal gaussian process regression. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, pages 993–1001.

[30] Schmidt, C. (2020). Chapter 13 - monitoring fires with the goes-r series. In Goodman, S. J., Schmit, T. J., Daniels, J., and Redmon, R. J., editors, *The GOES-R Series*, pages 145–163. Elsevier.

[31] Sigrist, F., Kunsch, H. R., and Stahel, W. A. (2015). Stochastic Partial Differential Equation based Modelling of Large Space-Time Data Sets. *Journal of the Royal Statistical Society: Series B*, 77(1):3–33.

[32] Sousa, M. J., Moutinho, A., and Almeida, M. (2020). Wildfire detection using transfer learning on augmented datasets. *Expert Systems with Applications*, 142:112975.

[33] Tang, Y., Martin, C. R., Li, Y., Chai, T. F., Pagowski, M., Wang, H. L., Kleist, D. T., Baker, B., Campbell, P., Huang, J. P., McQueen, J., Montuoro, R., Tong, D., Stajner, I., Jung, Y. S., Kumar, R., and Kondragunta, S. (2024). Impacts of the 2023 canadian fires on us air quality simulated by NOAA UFS-AQM with aerosol data assimilation. *104th AMS Annual Meeting*, https://ams.confex.com/ams/104ANNUAL/meetingapp.cgi/Paper/430756.

[34] Wei, G. Z., Krishnan, V., Xie, Y., Sengupata, M., Zhang, Y. C., Liao, H. T., and Liu, X. (2024a). A statistical model for multi-source remote-sensing data streams of wildfire aerosols optical depth. *INFORMS Journal on Data Science*, https://doi.org/10.1287/ijds.2021.0058.

[35] Wei, G. Z., Qiu, F., and Liu, X. (2024b). Convolutional non-homogeneous poisson process and its application to wildfire ignition risk quantification for power delivery networks. *Technometrics*, https://doi.org/10.1080/00401706.2024.2365729.

[36] Wei, G. Z. and Liu, X. and Barton, R. (2023). An extended PDE-based statistical spatio-temporal model that suppresses the Gibbs phenomenon. *Environmetrics*, https://doi.org/10.1002/env.2831.

[37] Wong, D. C., Pleim, J., Mathur, R., Binkowski, F., Otte, T. adn Gilliam, R., Pouliot, G., Xiu, A., Young, J. O., and Kang, D. (2012). WRF-CMAQ two-way coupled system with aerosol feedback: software development and preliminary results. *Geoscientific Model Development*, 5:299–312.

[38] Wong, D. C., Willison, J., Pleim, J. E., Sarwar, G., Beidler, J., Bullock, R., Herwehe, J. A., Gilliam, R., Kang, D., Hogrefe, C., Pouliot, G., and Foroutan, H. (2024). Development of the mpas-cmaq coupled system (v1.0) for multiscale global air quality modeling. *Geoscientific Model Development*, 17(21):7855–7866.

[39] Xu, C., Vazquez, D., Yao, R., Qiu, F., and Xie, Y. (2023). Spatio-temporal wildfire prediction using multi-modal data. *IEEE Selected Areas in Information Theory*, 4:302–313.

[40] Yeo, K. M., Hwang, Y. D., Liu, X., and Kalagnanam, J. (2019). Development of hp-inverse model by using generalized polynomial chaos. *Computer Methods in Applied Mechanics and Engineering*, 347:1–20.

[41] Yoo, M. S. and Wikle, C. (2024). A baysian spatio-temporal level set dynamic model and application to fire front propagation. *The Annals of Applied Statistics*, 18:404–423.

[42] Zhu, S. X., Yao, R., Xie, Y., Qiu, F., Qiu, Y. M., and Wu, X. (2022). Quantifying grid resilience against extreme weather using large-scale customer power outage data. *arXiv:2109.09711*.

[43] Zou, Y., Wang, Y. H., Ke, Z. M., Tian, H. Q., Yang, J., and Liu, Y. Q. (2019). Development of a REgion-Specific Ecosystem Feedback Fire (RESFire) Model in the Community Earth System Model. *Journal of Advances in Modeling Earth Systems*, 11:417–445.

CURRICULUM VITAE

Xiao Liu
David M. McKenney Family Associate Professor
H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology
email: xiao.liu@isye.gatech.edu webpage: https://sites.google.com/site/liuxiaosite1/
ORCID: https://orcid.org/0000-0003-4510-2847

• Academic and Industry Positions

Aug 2023-present	David M. McKenney Family Associate Professor H. Milton Stewart School of Industrial and Systems Engineering Georgia Institute of Technology
Jun 2017-July 2023	Assistant Professor and John L. Imhoff Endowed Chair (2022-2023) Department of Industrial Engineering, University of Arkansas
Oct 2015-May 2017	Research Staff Member (permanent research scientist position at IBM) IBM Thomas J. Watson Research Center, Yorktown Heights, New York
July 2012-Oct 2015	Research Staff Member (permanent research scientist position at IBM) IBM Smarter Cities Research Collaboratory Singapore, Singapore
Jun 2013-July 2016	Adjunct Assistant Professor Industrial and Systems Engineering, National University of Singapore

• Education

Ph.D.	Industrial and Systems Engineering, National University of Singapore, 2009
B.Eng.	Mechanical Engineering, Harbin Institute of Technology Hong Kong University of Science and Technology (exchange student)

• Research Areas

○ Domain-Aware Statistical Learning	[*physics-informed statistical modeling for nonlinear dynamics, statistical modeling for advection-diffusion processes, adaptive reduced-order models, inverse models, etc.]
○ Resilience & Environment	[*wildfires and power grid resilience, remote-sensing data modeling and environmental processes, reliability, survival analysis, etc.]
○ Data Science & Applied Statistics	[*ensemble trees for recurrence data, gradient boosted trees for Gaussian process, structural boosting trees for edge detection, etc.]

• Selected publications relevant to this proposal (*papers published by my PhD students)

1. *Wei, G.Z., Qiu, F., and Liu, X., (2024), "Convolutional Non-Homogeneous Poisson Process and its Application to Wildfire Ignition Risk Quantification for Power Delivery Networks", ***Technometrics***, available online: https://doi.org/10.1080/00401706.2024.2365729, arXiv: https://arxiv.org/abs/2301.00067.
2. *Wei, G.Z., Krishnan, V., Xie, Y., Sengupata, M., Zhang, Y.C., Liao, H.T., and Liu, X., (2024), "A Statistical Model for Multi-Source Remote-Sensing Data Streams of Wildfire Aerosols Optical Depth," ***INFORMS Journal on Data Science***, https://doi.org/10.1287/ijds.2021.0058, arXiv: https://arxiv.org/abs/2206.11766.
3. *Liu, X.C., Phan, D., Hwang, Y., Klein, L. Liu, X., Yeo, K., (2024), "Optimal Sensor Allocation with Multiple Linear Dispersion Processes", under revision, arXiv: https://arxiv.org/abs/2401.10437.
4. Liu, X., and Yeo, K.M. (2023), "Inverse Model for Estimating the Initial Condition of Spatio-Temporal Advection-Diffusion Processes", ***Technometrics***, 65, 432-445. arXiv: https://arxiv.org/abs/2302.04134.

5. *Wei, G.Z., Liu, X., and Barton, R. (2023), "An extended PDE-based statistical spatio-temporal model that suppresses the Gibbs phenomenon", ***Environmetrics***, https://doi.org/10.1002/env.2831.
6. *Liu, X.C., Liu, X., Kaman, T, Lu, X., and Lin, G. (2023), "Statistical Learning for Nonlinear Dynamical Systems", ***Technometrics***, 65, 564-578 (https://doi.org/10.1080/00401706.2023.2203175).
7. Liu, X., Yeo, K. M., Lu, S. Y. (2022), "Statistical Modeling for Spatio-Temporal Data from Stochastic Convection-Diffusion Processes", ***Journal of the American Statistical Association* (*Theory and Methods*)**, 117, 1482-1499, arXiv: https://arxiv.org/abs/1910.10375.
8. Liu, X., and Pan, R. (2020), "Analysis of Large Heterogeneous Repairable System Reliability Data with Static System Attributes and Dynamic Sensor Measurement in Big Data Environment", ***Technometrics***, 62, 206-222.
9. Yeo, K.M., Hwang, Y.D., Liu, X., and Kalagnanam, J. (2019), "Development of *hp*-inverse model by using generalized polynomial chaos", ***Computer Methods in Applied Mechanics and Engineering***, 347, 1-20.
10. Liu, X., Gopal, V., and Kalagnanam, J. (2018), "A Spatio-Temporal Modeling Framework for Weather Radar Image Data in Tropical Southeast Asia", ***Annals of Applied Statistics***, 12(1), 378–407.
11. Liu, X., Yeo, K.M., Hwang, Y.D., Singh, J. and Kalagnanam, J. (2016) "A Statistical Modeling Approach for Air Quality Data Based on Physical Dispersion Processes and Its Application to Ozone Modeling", ***Annals of Applied Statistics***, 10(2), 756-785.
12. Yeo, K., Hwang, Y., Liu, X. and Kalagnanam, J. (2016), "Stochastic Optimization Algorithm for Inverse Modeling of Air Pollution", ***DFD16 (Fluid Dynamics) Meeting of the American Physical Society***.
13. Singh, J., Yeo, K., Liu, X., Hosseini, R., and Kalagnanam, J. (2015), "Evaluation of WRF model seasonal forecasts for tropical region of Singapore", ***Advances in Science and Research***, 12, 69-72.

• U.S. Patent related to this project

1. "Airborne particulate source detection system". US20160377430A1.
2. "Detection Algorithms for Distributed Emission Sources of Abnormal Events". US20170147927A1.

• Industry R&D Experiences related to this project

1. Technical Lead, Engineering Predictive Environmental Analytics System (PEAS) with IBM Research and the National Environmental Agency (NEA) Singapore, 2012-2015

• Honors & Award related to this project

1. 2022 NSF CAREER Award (Title: Domain-Award Statistical Learning---Harnessing the Convergence of Engineering Knowledge and Data Science)
2. 2018 Statistics in Physical Engineering Sciences (SPES) Award, American Statistical Association (ASA).
3. [team honor] 2015 IBM Outstanding Technical Achievement Award for the Predictive Environmental Analytics System (PEAS) with the National Environmental Agency Singapore, 2012-2015.
4. [team honor] 2015 IBM Chairman's Environmental Award.

• Professional Services

1. Program co-Chair, Institute of Industrial and Systems Engineer (IISE) Annual Conference, 2025.
2. Program Committee, AAAI Conference on Artificial Intelligence (2019-2021).
3. Program Committee, IEEE BigData (2017-2022).

IDENTIFYING INFORMATION:

NAME: Xie, Yao

ORCID iD: https://orcid.org/0000-0001-6777-2951

POSITION TITLE: Coca-Cola Foundation Chair and Professor

PRIMARY ORGANIZATION AND LOCATION: Georgia Institute of Technology, Atlanta, Georgia, United States

Professional Preparation:

ORGANIZATION AND LOCATION	DEGREE (if applicable)	RECEIPT DATE	FIELD OF STUDY
Stanford University, Stanford, Georgia, United States	PHD	01/2012	Electrical Engineering (minor in Mathematics)
University of Florida, Gainesville, Georgia, United States	MS	07/2006	Electrical and Computer Engineering
University of Science and Technology of China, Hefei, Not Applicable, N/A, China	BS	07/2004	Electrical Engineering and Information Science

Appointments and Positions

2023 - present Coca-Cola Foundation Chair and Professor, Georgia Institute of Technology, Atlanta, Georgia, United States

2019 - 2023 Associate Professor, Georgia Institute of Technology, H. Milton School of Industrial and Systems Engineering (ISyE), Atlanta, Georgia, United States

2017 - 2023 Harold R. and Mary Anne Nash Early Career Professor, Georgia Institute of Technology, H. Milton School of Industrial and Systems Engineering, Atlanta, Georgia, United States

2013 - 2019 Assistant Professor, Georgia Institute of Technology, H. Milton School of Industrial and Systems Engineering (ISyE), Atlanta, Georgia, United States

2012 - 2013 Research Scientist, Duke University, Electrical and Computer Engineering, Durham, North Carolina, United States

Products

Products Most Closely Related to the Proposed Project

1. Xu C, Zuniga Vazquez D, Yao R, Qiu F, Xie Y. Spatio-temporal wildfire prediction using multi-modal data. IEEE Selected Areas in Information Theory (JSAIT). 2023 May; 4:302-313.
2. Dong Z, Cheng X, Xie Y. Spatio-temporal point processes with deep non-stationary kernels. International Conference on Learning Representations (ICLR). 2023.
3. Dong Z, Zhu S, Xie Y, Mateu J, Rodriguez-Cortes F. Non-stationary spatio-temporal point process modeling for high-resolution COVID-19 data. Journal of the Royal Statistical Society:

Series C (Applied Statistics). 2023 May; 72(2):368–386.

4. Zhu S, Bukharin A, Xie L, Yang S, Keskinocak P. Early detection of COVID-19 hotspots using spatio-temporal data. IEEE Jounal Selected Topics in Signal Processing (JSTSP). 2023; 72(2):368–386.
5. Dong Z, Zhu S, Zhang H, Xie Y, Van Hentenryck P. Multi-resolution spatio-temporal prediction with application to wind power generation. arXiv:2108.13285. 2023.

Other Significant Products, Whether or Not Related to the Proposed Project

1. Xu C, Jiang H, Xie Y. Conformal prediction for multi-dimensional time series by ellipsoidal sets. International Conference on Machine Learning (ICML). 2024.
2. Xu C, Xie Y. Conformal Prediction for Time Series. IEEE Trans Pattern Anal Mach Intell. 2023 Oct;45(10):11575-11587. PubMed PMID: 37819805.
3. Smith JR, Xie Y, Josef C, Kamaleswaran R. Online critical-state detection of sepsis among ICU patients using Jensen-Shannon divergence. American Medical Informatics Association (AMIA) Annual Symposium. 2022.
4. Wang J, Moore R, Kamaleswaran R, Xie Y. Improving sepsis prediction model generalization with optimal transport. Machine Learning for Heath (ML4H) Symposium. 2022.
5. Wei S, Kong X, Xavier A, Zhu S, Xie Y, Qiu F. Assessing electricity service unfairness with transfer counterfactual learning. t NeurIPS 2023 Workshop on Causal Representation Learning.. 2023.

Certification:

I certify that the information provided is current, accurate, and complete. This includes but is not limited to current, pending, and other support (both foreign and domestic) as defined in 42 U.S.C. § 6605.

I also certify that, at the time of submission, I am not a party to a malign foreign talent recruitment program.

Misrepresentations and/or omissions may be subject to prosecution and liability pursuant to, but not limited to, 18 U.S.C. §§ 287, 1001, 1031 and 31 U.S.C. §§ 3729-3733 and 3802.

Certified by Xie, Yao in SciENcv on 2024-11-26 21:43:48

Curriculum Vitae
YUHANG WANG

School of Earth and Atmospheric Sciences
Georgia Institute of Technology
311 Ferst Drive
Atlanta, GA 30332-0340

Phone: (404) 894-3995
Fax: (404) 894-5638
Email: ywang@eas.gatech.edu
Web: http://apollo.eas.gatech.edu

EDUCATION

Ph.D., Atmospheric Chemistry (1997), Harvard University
M.A., Atmospheric Chemistry (1994), Harvard University
M.S., Atmospheric Science (1992), University of Wyoming
B.S., Atmospheric Sciences (1990), Peking University

EMPLOYMENT

3/2010-present: Professor, Georgia Institute of Technology
7/2002-2/2010: Associate professor, Georgia Institute of Technology
1/2000-6/2002: Assistant professor, Rutgers University
10/1997-12/1999: Research faculty (research scientist II), Georgia Institute of Technology

SELECTED RELEVANT PUBLICATIONS (from >185 journal papers)

Liu, Y., Qian, Y., Rasch, P. J., Zhang, K., Wang, Y., Wang, M., Wang, H., and Yang, X.-Q., Fire-precipitation interactions amplify the quasi-biennial variability of fires over southern Mexico and Central America, *Atmos. Phys. Chem., 24,* 3115–3128, https://doi.org/10.5194/acp-24-3115-2024, 2024.

Zhang, A., Wang, Y., & Zou, Y., Positive feedback to regional climate enhances African wildfires, *iScience, 26,* 108533. https://doi.org/10.1016/j.isci.2023.108533, 2023.

Sarangi, C., Qian, Y., Leung, L. R., Zhang, Y., Zou, Y., and Wang, Y., Projected increases in wildfires may challenge regulatory curtailment of PM2.5 over the eastern US by 2050, *Atmos. Chem. Phys., 23,* 1769–1783, https://doi.org/10.5194/acp-23-1769-2023, 2023.

Shuman, J., and 86 others, Reimagine fire science for the Anthropocene, *PNAS NEXUS*, https://doi.org/10.1093/pnasnexus/pgac115, 2022.

Ke, Z., Wang, Y., Zou, Y., Song, Y., and Liu, Y., Global wildfire plume-rise data set and parameterizations for climate model applications, *J. Geophys. Res., 126,* e2020JD033085, https://doi.org/10.1029/2020JD033085, 2021.

Liu, Y. ,Liu, Y., Fu, J., Yang C.-E., Dong, X., Tian, H., Tao, B., Yang, J., Wang, Y., Zou, Y., Ke, Z., Projection of future wildfire emissions in western USA under climate change: contributions from changes in wildfire, fuel loading and fuel moisture, *International J. of Wildland Fire*, https://doi.org/10.1071/WF20190, 2021.

Song, Y., and Wang Y., Global wildfire outlook forecast with neural networks, *Remote Sens., 12,* 2246, doi:10.3390/rs12142246, 2020.

Cao, D., Park, TW., Song, Y., Wang, Y., You, Z. Lin, Y., and Deng, Y., A dynamical pathway bridging African biomass burning and Asian summer monsoon, *Clim. Dyn.,* https://doi.org/10.1007/s00382-021-05788-8, 2021.

Zou, Y., Wang, Y., Qian, Y., Tian, H., Yang, J., and Alvarado, E., Using CESM-RESFire to understand climate–fire–ecosystem interactions and the implications for decadal climate variability, *Atmos. Chem. Phys., 20*, 995–1020, https://doi.org/10.5194/acp-20-995-2020, 2020.

Zou, Y., Y. Wang, Z. Ke, H. Tian, J. Yang, and Y. Liu, Development of a REgion-Specific ecosystem feedback Fire (RESFire) model in the Community Earth System Model, *J. Adv. Mod. Earth Sys., 11,* https://doi.org/10.1029/2018MS001368, 2019.

Zhou, L., Y. Wang, Y. Chi, S. Wang, and Q. Wang, Contrasting post-fire dynamics between Africa and South America based on MODIS observations, *Remote Sens., 11,* doi:10.3390/rs11091074, 2019.

Liu, Y., K. Zhang, Y. Qian, Y. Wang, Y. Zou, Y. Song, H. Wan, X. Liu, and X.-Q. Yang, Investigation of short-term effective radiative forcing of fire aerosols over North America using nudged hindcast ensembles, *Atmos. Chem. Phys., 18,* 31-47, https://doi.org/10.5194/acp-18-31-2018, 2018.

Zhang, Y., H. Forrister, J. Liu, J. Dibb, B. Anderson, J. Schwarz, A. Perring, J. Jimenez, P. Campuzano-Jost, Y. Wang, A. Nenes, R. Weber, Top-of-atmosphere radiative forcing affected by brown carbon in the upper troposphere, *Nature Geosci., 10,* 486-488, DOI: 10.1038/NGEO2960, 2017.

Zhang, H., Y. Wang, T.-W. Park, and Y. Deng, Quantifying the relationship between extreme air pollution events and extreme weather events, *Atmos. Res., 188,* 64-79, doi:10.1016/j.atmosres.2016.11.010, 2017.

Liu, X., Y. Zhang, L. G. Huey, R. J. Yokelson, Y. Wang, J. L. Jimenez, P. Campuzano-Jost, A. J. Beyersdorf, D. R. Blake, Y. Choi, J. M. S. Clair, J. D. Crounse, D. A. Day, G. S. Diskin, A. Fried, S. R. Hall, T. F. Hanisco, L. E. King, S. Meinardi, T. Mikoviny, B. B. Palm, J. Peischl, A. E. Perring, I. B. Pollack, T. B. Ryerson, G. Sachse, J. P. Schwarz, I. J. Simpson, D. J. Tanner, K. L. Thornhill, K. Ullmann, R. J. Weber, P. O. Wennberg, A. Wisthaler, G. M. Wolfe, and L. D. Ziemba, Agricultural fires in the southeastern us during SEAC4RS: Emissions of trace gases and particles and evolution of ozone, reactive nitrogen, and organic aerosol, *J. Geophys. Res. Atmos., 120,* doi: 10.1002/2016JD025040, 2016.

Zeng, T., Z. Liu, and Y. Wang, Large fire emissions in summer over the southeastern US: Satellite measurements and modeling analysis, *Atmos. Environ., 127,* 213-220, 2016.

Yang, J., H. Tian, B. Tao, W. Ren, S. Pan, Y. Liu, and Y. Wang, A growing importance of large fires in conterminous United States during 1984–2012, *J. Geophys. Res. Biogeosci., 120,* doi:10.1002/2015JG002965, 2015.

Yang, J., H. Tian, B. Tao, W. Ren, C. Lu, S. Pan, Y. Wang, and Y. Liu, Century-scale patterns and trends of global pyrogenic carbon emissions and fire impacts on the terrestrial carbon balance*, Global Biogeochem. Cycles, 29*, doi:10.1002/2015GB005160, 2015.

Yang, J., H. Tian, B. Tao, W. Ren, J. Kush, Y. Liu, and Y. Wang, Spatial and temporal patterns of global burned area in response to anthropogenic and environmental factors: Reconstructing global fire history for the 20th and early 21st centuries, *J. Geophys. Res. Biogeosci., 119,* doi:10.1002/2013JG002532, 2014.

Zeng, T., and Y. Wang, Biomass burning induced nationwide summer peaks of OC/EC ratios in the continental United States, *Atmos. Environ*, *45*, 578-586, doi:10.1016/j.atmosenv.2010.10.038, 2011.

CURRENT AND PENDING (OTHER) SUPPORT INFORMATION

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person.

*NAME: Liu, Xiao

PERSISTENT IDENTIFIER (PID) OF THE SENIOR/KEY PERSON: https://orcid.org/0000-0003-4510-2847

*POSITION TITLE: David M. McKenney Family Associate Professor

*ORGANIZATION AND LOCATION: Georgia Institute of Technology, Atlanta, Georgia, United States

Proposals/Active Projects

***Proposal/Active Project Title:** CAREER: Domain-aware Statistical Learning

***Status of Support:** Current

Proposal/Award Number:

***Source of Support:** National Science Foundation

***Primary Place of Performance:** Georgia Institute of Technology

***Proposal/Active Project Start Date: (MM/YYYY):** 10/2022

***Proposal/Active Project End Date: (MM/YYYY):** 09/2027

***Total Anticipated Proposal/Project Amount:** $500,176

*** Person Months per budget period Devoted to the Proposal/Active Project:**

Year	Person Months
2023	1
2024	1
2025	1
2026	1
2027	1

***Overall Objectives:** To create a new Structure-Exploiting-Preserving (SEP) domain-aware statistical learning paradigm that enables the direct embedding of governing physics into data-driven models during model construction. To improve data literacy among the general public by improving awareness of the increasing availability of data and the capability of interpreting those data through local community activities.

***Statement of Potential Overlap:** None

***Proposal/Active Project Title:** AccelNet-Design: International Networks Towards Future U.S. Urban Resilience (Resilient-NET)

***Status of Support:**	Current
Proposal/Award Number:	
***Source of Support:**	National Science Foundation
***Primary Place of Performance:**	Georgia Institute of Technology
***Proposal/Active Project Start Date: (MM/YYYY):**	09/2022
***Proposal/Active Project End Date: (MM/YYYY):**	07/2025
***Total Anticipated Proposal/Project Amount:**	$249,905

*** Person Months per budget period Devoted to the Proposal/Active Project:**

Year	Person Months
2023	0.01
2024	0.01

***Overall Objectives:** To catalyze new cross-disciplinary global collaborations, synergize complementary scientific expertise, and provide critical access to one-of-its-kind pilot projects, data, platforms, and research capabilities for U.S. researchers to make global impacts and strengthen their leadership roles on emerging challenges in building resilience into future urban socio-technical systems.

***Statement of Potential Overlap:** This is the AccelNet Design-Track project.

***Proposal/Active Project Title:**	AccelNet Implementation Phase 1: International Networks Towards Future Resilience of U.S. Urban Socio-Technical Systems
***Status of Support:**	Pending
Proposal/Award Number:	
***Source of Support:**	NSF
***Primary Place of Performance:**	Georgia Institute of Technology
***Proposal/Active Project Start Date: (MM/YYYY):**	09/2025
***Proposal/Active Project End Date: (MM/YYYY):**	08/2029
***Total Anticipated Proposal/Project Amount:**	$1,500,000

*** Person Months per budget period Devoted to the Proposal/Active Project:**

Year	Person Months
2026	1
2027	1
2028	1
2029	1

***Overall Objectives:** To build comprehensive and strategic links among disciplinarily and geographically diverse U.S. and international networks, called Resilient-NET, to synergize research capabilities, data and best practices towards resilient and equitable future urban societies.

***Statement of Potential Overlap:** None

***Proposal/Active Project Title:** A Physics-Informed Machine Learning System for Real-Time Air Quality Forecasting during Wildfire Events

***Status of Support:** Pending

Proposal/Award Number:

***Source of Support:** NOAA

***Primary Place of Performance:** Georgia Institute of Technology

***Proposal/Active Project Start Date: (MM/YYYY):** 08/2025

***Proposal/Active Project End Date: (MM/YYYY):** 07/2028

***Total Anticipated Proposal/Project Amount:** $1,050,000

*** Person Months per budget period Devoted to the Proposal/Active Project:**

Year	Person Months
2026	1
2027	1
2028	1

***Overall Objectives:** This project will develop and test computationally-fast, physics-informed, and multi-source data-driven Machine Learning (ML) algorithms and software for generating low-latency hourly regional air quality predictions (updated every 10 minutes) for the next 24 hours during hazardous wildfire episodes. The ML system will consist of a complete suite of three components: (i) a fire spread and source estimation module, (ii) a smoke propagation module, and (iii) an air quality (Particulate Matters, PM) prediction module.

***Statement of Potential Overlap:** N/A

***Proposal/Active Project Title:** Travel: Travel Support for 2025 IISE Annual Conference and Expo, Atlanta, GA; May 31-June 3, 2025

***Status of Support:** Pending

Proposal/Award Number:

***Source of Support:** NSF

***Primary Place of Performance:** Arizona State University

***Proposal/Active Project Start Date: (MM/YYYY):** 02/2025

***Proposal/Active Project End Date: (MM/YYYY):** 08/2025

***Total Anticipated Proposal/Project Amount:** $50,000

*** Person Months per budget period Devoted to the Proposal/Active Project:**

Year	Person Months
2025	0.01

***Overall Objectives:** Support participants travel to IISE 2025

***Statement of Potential Overlap:** None

***Proposal/Active Project Title:** Compound Challenges and Collective Solutions: Building Resilience in Communities under Multi-hazard Threats in Greater Los Angeles

***Status of Support:** Pending

Proposal/Award Number:

***Source of Support:** NSF

***Primary Place of Performance:** Georgia Institute of Technology

***Proposal/Active Project Start Date: (MM/YYYY):** 01/2025

***Proposal/Active Project End Date: (MM/YYYY):** 12/2027

***Total Anticipated Proposal/Project Amount:** $999,648

*** Person Months per budget period Devoted to the Proposal/Active Project:**

Year	Person Months
2025	1
2026	1
2027	1

***Overall Objectives:** To develop evidence-based strategies that prioritize equity and community empowerment. To identify critical vulnerabilities and adaptive capacities of infrastructures and communities. The ultimate goal is to provide actionable insights and tools for policymakers, urban planners, and community leaders to build robust, resilient systems that can withstand and adapt to the increasing frequency and intensity of extreme events, ensuring the safety and well-being of all residents

***Statement of Potential Overlap:** First of all, this is not a networking proposal. This is a research proposal laser-focused on developing new mathematical modeling tools and policies for cascading events, especially flooding, in California communities.

***Proposal/Active Project Title:**	Enabling Real-time Wildfire Management Capabilities using Explainable Machine Learning (ExpML) -- An integrated system for Fire Spread Prediction, Data Fusion, and Dynamic Deployment of Suborbital Platform
***Status of Support:**	Pending
Proposal/Award Number:	
***Source of Support:**	NASA
***Primary Place of Performance:**	Georgia Institute of Technology
***Proposal/Active Project Start Date: (MM/YYYY):**	10/2024
***Proposal/Active Project End Date: (MM/YYYY):**	09/2027
***Total Anticipated Proposal/Project Amount:**	$1,279,364

*** Person Months per budget period Devoted to the Proposal/Active Project:**

Year	Person Months
2025	1
2026	1
2027	1

***Overall Objectives:** This pending proposal focuses on fire spread modeling, and especially the optimal deployment of suborbital platforms for data collection. It does NOT address emission estimation, smoke propagation, nor air quality prediction at all.

***Statement of Potential Overlap:** N.A.

Certification:

I certify that the information provided is current, accurate, and complete. This includes but is not limited to current, pending, and other support (both foreign and domestic) as defined in 42 U.S.C. § 6605.

I also certify that, at the time of submission, I am not a party to a malign foreign talent recruitment program.

Misrepresentations and/or omissions may be subject to prosecution and liability pursuant to, but not limited to, 18 U.S.C. §§ 287, 1001, 1031 and 31 U.S.C. §§ 3729-3733 and 3802.

Certified by Liu, Xiao in SciENcv on 2024-12-02 08:52:24

CURRENT AND PENDING (OTHER) SUPPORT INFORMATION

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person.

*NAME: Xie, Yao

*POSITION TITLE: Professor

*ORGANIZATION AND LOCATION: Georgia Institute of Technology, Atlanta, Georgia, United States

Proposals/Active Projects

***Proposal/Active Project Title:** Structure discovery in complex dynamic networks

***Status of Support:** Current

Proposal/Award Number:

***Source of Support:** Office of Naval Research

***Primary Place of Performance:** Georgia Institute of Technology

***Proposal/Active Project Start Date: (MM/YYYY):** 07/2024

***Proposal/Active Project End Date: (MM/YYYY):** 06/2027

***Total Anticipated Proposal/Project Amount:** $200,000

*** Person Months per budget period Devoted to the Proposal/Active Project:**

Year	Person Months
2024	1
2025	1
2026	1

***Overall Objectives:** To develop change-point detection for dynamic networks.

***Statement of Potential Overlap:** None.

***Proposal/Active Project Title:** Collaborative Research: ATD: a-DMIT: a novel Distributed, MultI-resolution, Topology-aware online monitoring framework of massive spatiotemporal data

***Status of Support:** Current

Proposal/Award Number:

***Source of Support:** NSF

***Primary Place of Performance:** Georgia Tech

***Proposal/Active Project Start Date: (MM/YYYY):** 07/2023

***Proposal/Active Project End Date: (MM/YYYY):** 06/2026

***Total Anticipated Proposal/Project Amount:** $127,396

*** Person Months per budget period Devoted to the Proposal/Active Project:**

Year	Person Months
2023	0.1
2024	0.1
2025	0.1

***Overall Objectives:** Develop computationally efficient algorithm for non-parametric anomaly detection from streaming data.

***Statement of Potential Overlap:** None.

***Proposal/Active Project Title:** Collaborative Research: IMR: MM-1A: MapQ: Mapping Quality of Coverage in Mobile Broadband Networks using Latent Gaussian Process Models

***Status of Support:** Current

Proposal/Award Number: 55484

***Source of Support:** NSF

***Primary Place of Performance:** Georgia Institute of Technology

***Proposal/Active Project Start Date: (MM/YYYY):** 01/2023

***Proposal/Active Project End Date: (MM/YYYY):** 12/2025

***Total Anticipated Proposal/Project Amount:** $321,978

*** Person Months per budget period Devoted to the Proposal/Active Project:**

Year	Person Months
2023	0.3
2024	0.3
2025	0.3

***Overall Objectives:** Develop novel statistical methods that synthesize data on cellular networks and signals, together with performance measurements, to generate fine-grained maps of mobile broadband coverage and quality of coverage, while also guiding future measurement efforts to provide most benefit.

***Statement of Potential Overlap:** None

***Proposal/Active Project Title:** MFB: Novel Graph Neural Networks to Understand, Predict, and Design Allosteric Transcription Factors

***Status of Support:** Current
Proposal/Award Number:
***Source of Support:** NSF
***Primary Place of Performance:** Georgia Tech
***Proposal/Active Project Start Date: (MM/YYYY):** 08/2022
***Proposal/Active Project End Date: (MM/YYYY):** 07/2025
***Total Anticipated Proposal/Project Amount:** $1,485,927

*** Person Months per budget period Devoted to the Proposal/Active Project:**

Year	Person Months
2023	1
2024	1
2025	1

***Overall Objectives:** To develop invertible graph neural networks for allosteric transcription factors.

***Statement of Potential Overlap:** None.

***Proposal/Active Project Title:** Bridging Statistical Hypothesis Tests and Deep Learning for Reliability and Computational Efficiency
***Status of Support:** Current
Proposal/Award Number: 34709
***Source of Support:** NSF
***Primary Place of Performance:** Georgia Institute of Technology
***Proposal/Active Project Start Date: (MM/YYYY):** 01/2022
***Proposal/Active Project End Date: (MM/YYYY):** 12/2024
***Total Anticipated Proposal/Project Amount:** $1,199,954

*** Person Months per budget period Devoted to the Proposal/Active Project:**

Year	Person Months
2022	1
2023	1
2024	1

***Overall Objectives:** Develop fundamental theory and methods for deep learning for hypothesis testing.

***Statement of Potential Overlap:** None.

***Proposal/Active Project Title:** A Physics-Informed Machine Learning System for Real-Time Air Quality Forecasting during Wildfire Events

***Status of Support:** Pending

Proposal/Award Number:

***Source of Support:** NOAA

***Primary Place of Performance:** Georgia Institute of Technology

***Proposal/Active Project Start Date: (MM/YYYY):** 08/2025

***Proposal/Active Project End Date: (MM/YYYY):** 07/2028

***Total Anticipated Proposal/Project Amount:** $1,050,000

*** Person Months per budget period Devoted to the Proposal/Active Project:**

Year	Person Months
2026	1
2027	1
2028	1

***Overall Objectives:** This project will develop and test computationally-fast, physics-informed, and multisource data-driven Machine Learning (ML) algorithms and software for generating low-latency hourly regional air quality predictions (updated every 10 minutes) for the next 24 hours during hazardous wildfire episodes. The ML system will consist of a complete suite of three components: (i) a fire spread and source estimation module, (ii) a smoke propagation module, and (iii) an air quality (Particulate Matters, PM) prediction module.

***Statement of Potential Overlap:** This proposal.

***Proposal/Active Project Title:** CIF:Medium:Generative AI Modeling High-Dimensional Distributions with Applications in Power Systems Outage Analysis

***Status of Support:** Pending

Proposal/Award Number:

***Source of Support:** National Science Foundation

***Primary Place of Performance:** Georgia Institute of Technology

***Proposal/Active Project Start Date: (MM/YYYY):** 07/2025

***Proposal/Active Project End Date: (MM/YYYY):** 06/2028

***Total Anticipated Proposal/Project Amount:** $783,508

*** Person Months per budget period Devoted to the Proposal/Active Project:**

Year	Person Months
2025	1

Year	Person Months
2026	1
2027	1

***Overall Objectives:** This project aims to develop rigorous mathematical models integrating generative AI into scenario generation for critical infrastructure, particularly power systems, to ensure trustworthy and explainable outputs with algorithmic guarantees.

***Statement of Potential Overlap:** There is no scientific, budgetary or effort overlap.

Certification:

I certify that the information provided is current, accurate, and complete. This includes but is not limited to current, pending, and other support (both foreign and domestic) as defined in 42 U.S.C. § 6605.

I also certify that, at the time of submission, I am not a party to a malign foreign talent recruitment program.

Misrepresentations and/or omissions may be subject to prosecution and liability pursuant to, but not limited to, 18 U.S.C. §§ 287, 1001, 1031 and 31 U.S.C. §§ 3729-3733 and 3802.

Certified by Xie, Yao in SciENcv on 2024-11-27 11:16:57

CURRENT AND PENDING (OTHER) SUPPORT INFORMATION

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person.

*NAME: Wang, Yuhang NONE

PERSISTENT IDENTIFIER (PID) OF THE SENIOR/KEY PERSON: https://orcid.org/0000-0002-7290-2551

*POSITION TITLE: Professor

*ORGANIZATION AND LOCATION: Georgia Institute of Technology, Atlanta, Georgia, United States

Proposals/Active Projects

***Proposal/Active Project Title:** Development and utilization of multi-sensor data record for long term trend studies and model evaluations (PI, Ray Wang)

***Status of Support:** Current

Proposal/Award Number:

***Source of Support:** NASA

***Primary Place of Performance:** Atlanta, GA

***Proposal/Active Project Start Date: (MM/YYYY):** 05/2021

***Proposal/Active Project End Date: (MM/YYYY):** 04/2025

***Total Anticipated Proposal/Project Amount:** $571,865

*** Person Months per budget period Devoted to the Proposal/Active Project:**

Year	Person Months
2021	0.5
2022	0.5
2023	0.5

***Overall Objectives:** (1) Validate SAGE III retrieved ozone and water vapor products; (2) Create an updated GOZCARDS ozone data set using SAGE III and other viable limb scattering measurements; (3) Investigate the lower stratospheric ozone trends; (4) Evaluate lower stratospheric ozone change in CCMI-2 and CMIP6 historical simulations.

***Statement of Potential Overlap:** N/A

***Proposal/Active Project Title:** A Physics-Informed Machine Learning System for Real-Time Air Quality Forecasting during Wildfire Events

***Status of Support:** Pending

Proposal/Award Number:

***Source of Support:** NOAA

***Primary Place of Performance:** Georgia Institute of Technology

***Proposal/Active Project Start Date: (MM/YYYY):** 08/2025

***Proposal/Active Project End Date: (MM/YYYY):** 07/2028

***Total Anticipated Proposal/Project Amount:** $1,050,000

*** Person Months per budget period Devoted to the Proposal/Active Project:**

Year	Person Months
2026	1
2027	1
2028	1

***Overall Objectives:** This project will develop and test computationally-fast, physics-informed, and multisource data-driven Machine Learning (ML) algorithms and software for generating low-latency hourly regional air quality predictions (updated every 10 minutes) for the next 24 hours during hazardous wildfire episodes. The ML system will consist of a complete suite of three components: (i) a fire spread and source estimation module, (ii) a smoke propagation module, and (iii) an air quality (Particulate Matters, PM) prediction module.

***Statement of Potential Overlap:** This proposal

***Proposal/Active Project Title:** Assessing the global impacts of biomass burning emissions on tropospheric ozone

***Status of Support:** Pending

Proposal/Award Number:

***Source of Support:** NASA

***Primary Place of Performance:** Atlanta, GA

***Proposal/Active Project Start Date: (MM/YYYY):** 02/2025

***Proposal/Active Project End Date: (MM/YYYY):** 01/2028

***Total Anticipated Proposal/Project Amount:** $631,690

*** Person Months per budget period Devoted to the Proposal/Active Project:**

Year	Person Months
2025	1
2026	1
2027	1

***Overall Objectives:** Biomass burning (BB) emits large amounts of gaseous and particulate matter that significantly change air quality and tropospheric oxidants. While the potentially significant impact of BB on tropospheric ozone due to BB emissions of reactive nitrogen ($NOx=NO+NO_2$) and volatile organic compounds (VOCs) has been known, the observations from suborbital and ground-based field experiments such as ATom (Atmospheric Tomography Mission) and FIREX-AQ (Fire Influence on Regional to Global Environments and Air Quality) shed new insights into chemical and physical processes of BB emissions and their impacts on atmospheric composition. While insightful, observation-based analysis has limitations in quantitatively assessing the global impact of BB emissions on tropospheric ozone, partly due to the constraints in data availability. To address this, we propose using the GEOS-Chem model simulations to reassess the impact of BB emissions on tropospheric ozone.

***Statement of Potential Overlap:** N/A

***Proposal/Active Project Title:** Enhancing the modeling capability of fire absorbing aerosols in the GEOS system for assessing brown carbon radiative forcing

***Status of Support:** Pending

Proposal/Award Number:

***Source of Support:** NASA

***Primary Place of Performance:** Atlanta, GA

***Proposal/Active Project Start Date: (MM/YYYY):** 01/2025

***Proposal/Active Project End Date: (MM/YYYY):** 12/2028

***Total Anticipated Proposal/Project Amount:** $935,559

*** Person Months per budget period Devoted to the Proposal/Active Project:**

Year	Person Months
2025	1
2026	1
2027	1
2028	1

***Overall Objectives:** Modeling and EPIC (Earth Polychromatic Imaging Camera) retrievals of fire brown carbon (BrC) will be developed and analyzed synergistically in a modeling-analysis framework anchored by aircraft and satellite observations. The improved modeling and EPIC retrievals of BrC will be used to assess the radiative forcing of fire soluble BrC and dark BrC in light of that of black carbon. The proposed research will significantly advance GEOS modeling, observation-constrained analysis, and our understanding of fire BrC and its impact on the climate system.

***Statement of Potential Overlap:** N/A

Certification:

I certify that the information provided is current, accurate, and complete. This includes but is not limited to current, pending, and other support (both foreign and domestic) as defined in 42 U.S.C. § 6605.

I also certify that, at the time of submission, I am not a party to a malign foreign talent recruitment program.

Misrepresentations and/or omissions may be subject to prosecution and liability pursuant to, but not limited to, 18 U.S.C. §§ 287, 1001, 1031 and 31 U.S.C. §§ 3729-3733 and 3802.

Certified by Wang, Yuhang in SciENcv on 2024-11-27 13:17:23

Budget Justification

The project period is for a duration of 36 months, 8/01/2025 – 7/31/2028

Senior Personnel

The Principal Investigator (PI), Xiao Liu, is an Associate Professor in the School of Industrial and Systems Engineering at Georgia Tech and is committed to 1 month of support each year of the project. PI Liu will oversee ALL proposal activities. The Co-Principal Investigator (Co-PI), Yao Xie, is a Professor in the School of Industrial and Systems Engineering at Georgia Tech and is committed to 1 month of support each year of the project. The Co-Principal Investigator (Co-PI), Yuhang Wang, is a Professor in the School of Earth & Atmospheric Sciences at Georgia Tech and is committed to 1 month of support for each year of the project. The total salary requested for the three-year project period is $155,821.

Year 1	Year 2	Year 3	Total
$50,413	**$51,925**	**$53,483**	**$155,821**

Other Personnel

This project will support three Graduate Research Assistants (GRAs). Two will be from the School of Industrial and Systems Engineering (CoE) and one from the School of Earth & Atmospheric Sciences (CoS) at Georgia Tech. In Year 1 and Year 2, funds are requested to support each of the GRAs for 9.5 months in each year. In Year 3, each GRA will be supported for 9 months. The GRAs will be committed to the research tasks described in the proposal. The total salary requested for the three-year project period is $335,423.

Year 1	Year 2	Year 3	Total
$110,516	**$113,831**	**$111,076**	**$335,423**

**All salaries escalated at 3% annually starting Year 1.*

Fringe Benefits Rate

The fringe benefits rate is 31.7% for the PI, and 7.9% for the GRA. The total fringe benefits for the project period is $75,894.

Year 1	Year 2	Year 3	Total
$24,712	**$25,453**	**$25,729**	**$75,894**

Travel

(Domestic) Travel budget ($9,613) is requested for the PI, Co-PIs and the GRAs to participate potential meetings with NOAA, stakeholders and activities that help to disseminate the results.

The total travel budget is summarized as follows:

Year 1	Year 2	Year 3	Total
$6,958	**$496**	**$2,160**	**$9,613**

Other Direct Costs

The tuition rate is $1,596 per month per student. Tuition is escalated at 3% annually starting Year 1. Total tuition requested is $142,194

Year 1	Year 2	Year 3	Total
$46,850	**$48,256**	**$47,088**	**$142,194**

Total Direct Costs:

Year 1	Year 2	Year 3	Total
$239,449	**$239,961**	**$239,535**	**$718,945**

Facilities and Administrative Costs

Indirect costs are charged on direct costs (excludes equipment, participant support, tuition, and sub-award amount greater than $25,000) at a rate of 57.4%.

Total Indirect Costs:

Year 1	Year 2	Year 3	Total
$110,551	**$110,039**	**$110,465**	**$331,055**

TOTAL PROJECT COSTS:

Year 1	Year 2	Year 3	Total
$350,000	**$350,000**	**$350,000**	**$1,050,000**

DEPARTMENT OF THE NAVY
OFFICE OF NAVAL RESEARCH
875 NORTH RANDOLPH STREET
SUITE 1425
ARLINGTON, VA 22203-1995

April 2, 2024

NEGOTIATION AGREEMENT

INSTITUTION: **GEORGIA INSTITUTE OF TECHNOLOGY**
GEORGIA TECH RESEARCH CORPORATION
ATLANTA, GEORGIA 30332

The Facilities and Administrative (F&A) rates contained herein are for use on grants, contracts and/or other agreements issued or awarded to the Georgia Institute of Technology/Georgia Tech Research Corporation (GIT/GTRC) by all Federal Agencies of the United States of America, in accordance with the provisions and cost principles mandated by 2 CFR Part 200. These rates shall be used for forward pricing and billing purposes for GIT/GTRC's Fiscal Years 2026 and 2027.

Section I: RATES - TYPE: PREDETERMINED (PRED)

F&A Rates:

TYPE	FROM	TO	RATE	BASE	APPLICABLE TO	LOCATION
Pred	7/1/2025	6/30/2027	57.4%	(a)	Organized Research (1)	On Campus
Pred	7/1/2025	6/30/2027	26.0%	(a)	Organized Research (1)	Off Campus
Pred	7/1/2025	6/30/2027	66.5%	(a)	Organized Research (2)	On Campus
Pred	7/1/2025	6/30/2027	35.2%	(a)	Organized Research (2)	Off Campus
Pred	7/1/2025	6/30/2027	36.7%	(a)	Other Sponsored Activities	On Campus
Pred	7/1/2025	6/30/2027	52.8%	(a)	Sponsored Instruction	On Campus

DISTRIBUTION BASES

(a) Modified Total Direct Costs (MTDC) means all direct salaries and wages, applicable fringe benefits, materials and supplies, services, travel, and up to the first $25,000 of each subaward (regardless of the period of performance of the subawards under the award). MTDC excludes equipment, capital expenditures, charges for patient care, rental costs, tuition remission, scholarships and fellowships, participant support costs and the portion of each subaward in excess of $25,000.

APPLICABLE TO

(1) Applies to DOD contracts awarded before November 30, 1993, all Non-DOD Instruments, and all DOD grants (See Section II, paragraph E). (Capped)
(2) Applies to only DOD contracts awarded on or after November 30, 1993, in accordance with and under the authority of DFARS 231.303(1) (See Section II, paragraph E). (Uncapped)

SECTION II - GENERAL TERMS AND CONDITIONS

A. **LIMITATIONS**: Use of the rates set forth under Section I is subject to availability of funds and to any other statutory or administrative limitations. The rates are applicable to a given grant, contract or other agreement only to the extent that funds are available and consistent with any and all limitations of cost clauses or provisions, if any, contained therein. Acceptance of any or all of the rates agreed to herein is predicated upon the following conditions: (1) that no costs other than those incurred by the institution were included in this indirect cost pool as finally accepted and that such costs are legal obligations of the institution and allowable under governing cost principles; (2) that the same costs that have been treated as indirect costs are not claimed as direct costs; (3) that similar types of costs have been accorded consistent accounting treatment; and (4) that the information provided by the institution which was used as a basis for acceptance of the rates agreed to herein, and expressly relied upon by the Government in negotiating and accepting the said rates is not subsequently found to be materially incomplete or inaccurate.

B. **ACCOUNTING CHANGES**: The rates contained in Section I of this agreement are based on the accounting system in effect at the time the agreement was negotiated. Changes to the method(s) of accounting for costs, which affect the amount of reimbursement resulting from the use of these rates require the prior written approval of the authorized representative of the cognizant agency for indirect costs. Such changes include but are not limited to changes in the charging of a particular type of cost from indirect to direct. Failure to obtain such approval may result in subsequent cost disallowances.

C. **PREDETERMINED RATES**: The predetermined rates contained in this agreement are not subject to adjustment in accordance with the provisions of 2 CFR Part 200, subject to the limitations contained in Part A of this section.

D. **USE BY OTHER FEDERAL AGENCIES**: The rates set forth in Section I are negotiated in accordance with and under the authority set forth in 2 CFR Part 200. Accordingly, such rates shall be applied to the extent provided in such regulations to grants, contracts, and other agreements to which 2 CFR Part 200 applies, subject to any limitations in part A of this section. Copies of this document may be provided by either party to other federal agencies to provide such agencies with documentary notice of this agreement and its terms and conditions.

E. **APPLICATION OF INDIRECT COST RATES TO DEPARTMENT OF DEFENSE (DOD) CONTRACTS**: In accordance with DFARS 231.303, no limitation may be placed on the reimbursement of otherwise allowable indirect cost incurred by an institution of higher education under a DoD contract awarded on or after November 30, 1993, unless the same limitation is applied uniformly to all other organizations performing similar work. It has been determined by DoD that such limitation is not being uniformly applied. Accordingly, the rates cited (2) of Section I, as explained under the title, "APPLICABLE TO" do not reflect the application of the 26% limitation on administrative indirect costs imposed by 2 CFR Part 200, whereas (1) does so.

F. **DFARS WAIVER**: Signature of this agreement by the authorized representative of GIT/GTRC and the Government acknowledges and affirms the University's request to waive the prohibition contained in DFARS 231.303(1) and the Government's exercise of its discretion contained in DFARS 231.303(2) to waive the prohibition in DFARS 231.303(1). The waiver request by GIT/GTRC for Other Sponsored Activities and Sponsored Instruction is made to simplify the University's overall management of DoD cost reimbursements under DoD contracts.

G. **SPECIAL REMARKS**:

(1) In accordance with 2 CFR Part 200, Subpart 200.414(g), GIT/GTRC has requested an extension of its Fiscal Year FY2025 rates. Therefore, the rates identified in Section I are an extension of the GIT/GTRC's FY2025 rates.

(2) The Government's agreement with the rates set forth in Section I is not an acceptance of the GIT/GTRC's accounting practices or methodologies. Any reliance by the Government on cost data or methodologies submitted by GIT/RI is on a non-precedence-setting basis and does not imply Government acceptance.

Accepted:

FOR GEORGIA INSTITUTE OF TECHNOLOGY: GEORGIA TECH RESEARCH CORP.	FOR THE U.S. GOVERNMENT:
DocuSigned by: Cynthia Hope 39CCA48AF46844D...	TINGLE.BETTY.JOHNSON.1204289359 Digitally signed by TINGLE.BETTY.JOHNSON.1204289359 Date: 2024.04.03 09:17:28 -04'00'
Cynthia Hope, VP for Research, VP GTRC/GTARC	Betty J. Tingle Contracting Officer
4/2/2024 | 8:34 PM EDT	
Date	Date

For information concerning this agreement contact:
Betty Tingle, Office of Naval Research, Phone: (703) 696-7742, E-mail: betty.j.tingle.civ@us.navy.mil