USDA
United States Department of Agriculture
Agricultural Research Service

SCINet Newsletter: April 2024

Research Spotlight | News | Working Groups
Training | Support | Connect

RESEARCH SPOTLIGHT

Monte Carlo simulations on Atlas for soil content determinations

By: Galina Yakubova, Aleksandr Kavetskiy, Stephen A. Prior, and H. Allen Torbert
National Soils Dynamics Laboratory, Auburn, AL

The Monte Carlo Method, or multiple probability simulation, is a mathematical technique used to estimate possible outcomes of uncertain events. The Monte Carlo Method was applied for nuclear problems by John von Neumann and Stanislaw Ulam during work on the Manhattan Project. It is a very powerful technique for studying the propagation of ionized particles into a medium and the secondary effects under irradiation.

The USDA-ARS National Soils Dynamics Laboratory (NSDL) developed an in-situ method of neutron-gamma analysis for determining soil elemental content. This method is based on the registration of gamma rays issued from the studied soil upon neutron irradiation. Since each soil element issues gamma rays of specific energy, soil elemental content can be defined from the measured neutron stimulated gamma spectra. Monte Carlo computer simulations of neutron stimulated gamma spectra of different objects (e.g., soils, composts, calibration pits, and other agricultural media) is an important aspect of this research.

Conducting experiments with different measurement system designs and large samples (2 m x 2 m x 50 cm) with various soil elemental contents and textures is expensive, time consuming, labor-intensive, and often physically impractical. Thus, computer simulations are useful for incorporating different factors to define their effects on the resultant gamma spectra.

Developed and maintained by Los Alamos National Laboratory, MCNP software is used for analyzing the transport of neutrons (hence "NP" for *neutral particles*) and gamma rays by the Monte Carlo method (hence "MC"). MCNP simulations require a very large number of neutron events (on the scale of 10^9-10^{10} or greater) to create gamma spectra with an accuracy comparable to experimental measurements. By comparison, 10^7 neutrons per second and a measurement time of no less than 15 min are used in field experiments to acquire the soil gamma spectra. Thus, the application of high-performance computing (HPC) is suitable for these types of simulations. SCINet's Atlas HPC cluster at Mississippi State University is currently used for this NSDL research project.

MCNP computer simulations of neutron-stimulated gamma spectra on the Atlas HPC cluster make it possible to solve the following problems regarding soil content determinations:

- optimizing the design of the field measurement system;
- determining dose levels to soil and nearby environments during measurements (see Figure 1);
- developing methods and procedures for:
    - calibration of the measurement system for carbon and moisture content determinations;
    - correcting for soil moisture when measuring carbon content in the field;
    - spectra deconvolution for determining soil component contents;
    - neutron-gamma measurement for determining the carbon-to-nitrogen ratio in compost materials;
    - assessing potassium content and chlorine contamination in the field.

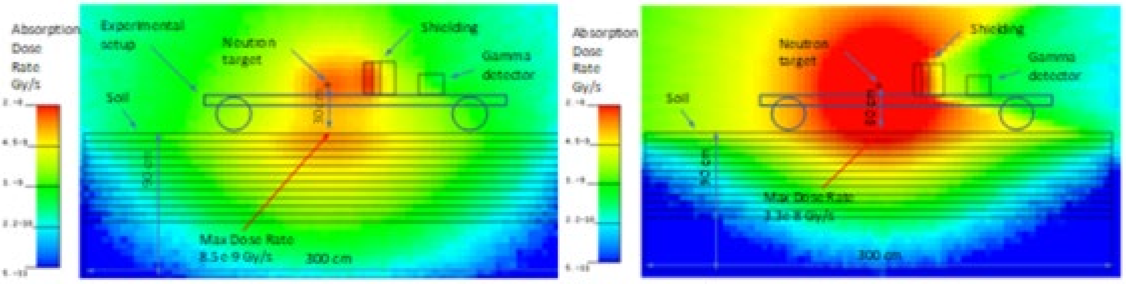


Figure 1. Gamma (left panel) and neutron (right panel) dose rate distributions in soil and the surrounding environment during neutron-gamma analysis at a flux of 2-10^7 neutron/s.

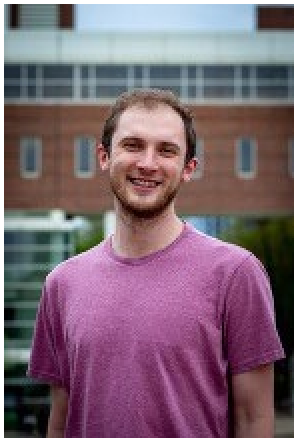
SCINet and AI COE Fellows

Please welcome our newest SCINet and ARS Artificial Intelligence Center of Excellence (AI-COE) fellows!

Dr. EfraÃn Duarte received his Ph.D. in Agronomic Sciences with a specialization in Soil Science from the University of ConcepciÃ³n, Chile. His doctoral research focused on the application of satellite remote sensing data and *in situ* measurements to model soil organic carbon and forest degradation using various machine learning algorithms and geospatial datasets. Prior to his Ph.D., Dr. Duarte earned a B.Sc. in Forestry Engineering from JosÃ© Cecilio del Valle University in Honduras and a M.Sc. in Project Management from Central American Technological

University (UNITEC), also in Honduras.

Currently, Dr. Duarte is a postdoctoral fellow working at the USDA-ARS Forage and Range Research Laboratory, under the mentorship of Dr. Alexander HernÃ¡ndez. Efrain is interested in assessing how different restoration practices on disturbed landscapes impact the spatial and temporal dynamics of soil moisture in fragile semiarid ecosystems of the western USA. To achieve this, Dr. Duarte is developing cutting-edge ecological datasets, while calibrating time series of satellite-derived soil moisture with field data and multispectral sensors onboard unoccupied aerial vehicles (UAVs). Dr. Duarte is implementing the translation of local-server geospatial workflows to high-performance computing clusters. This activity will effectively bridge gaps for USDA-ARS scientists and support staff to migrate their local geospatial analyses to SCINet using open-source software exclusively.



Dr. Austin Weigle is co-advised by Drs. Brenda Oppert, Chris Mattison, and Gerard Lazo. He is working on the SCINet project "Engineering Agriculture Products through the Application of AI/ML to Insect Enzyme and Substrate Interactions". For this project, he is combining molecular modeling, simulation, and machine learning to engineer allergen-degrading insect enzymes for optimal function in the human stomach. He is also identifying nontoxic wheat epitope sequences by resolving gliadin peptide folding mechanisms. The results of his research might translate to the design or identification of protein products that could reduce wheat- and nut-allergenicity.

Dr. Weigle completed his B.S. in Plant Biology (conc. Biochemical and Molecular Physiology) at Southern Illinois University, which he attended as a Chancellor's Scholar and a Division I high jumper on the Men's Track & Field Team. He then completed his Ph.D. in Chemistry (concentrating in Chemical Biology) at the University of Illinois Urbana-Champaign with Prof. Diwakar Shukla. For his dissertation research, Dr. Weigle used molecular dynamics to engineer substrate selectivity in a plant sugar transporter and study how lipid environments affect modeled membrane protein function. He enjoys combining data-driven approaches and rational design elements for solving interdisciplinary problems at the chemistry-biology interface. During his scientific training, Dr. Weigle received honorable mentions from the Barry M. Goldwater Foundation and the National Science Foundation. He is excited to apply his scientific trainings towards agriculture-related problems and is looking forward to developing his AI-related skillsets through his SCINet/AI-COE fellowship.

NEWS

AI Innovation Fund Awardees

Congratulations to the following five ARS scientists and their co-PIs for earning FY2024 Artificial Intelligence Innovation Fund awards. There were many excellent proposals submitted this year, making for a very competitive field and a difficult review process.

(Awardees are listed in alphabetical order by lead PI's last name.)

- **Ardeshir Adeli** and Yanbo Huang: *Deep Learning Based High-Resolution Field Level Soil Moisture Mapper from UAVs*
- **Christopher Anderson**, Shawn Bearson, and Paul Villanueva: *Rapid prediction of* Salmonella enterica *virulence from MALDI-TOF mass spectra using machine learning*
- **Carson Andorf**, Hye-Seon Kim, and Taner Sen: *The* Fusarium-*Host Interactome Discovery App*
- **Loren Honaas**: *Bad apples? Next Generation Postharvest Risk Assessment Tools*
- **William Meikle**: *Turning the 2014-2023 USDA-ARS Hive Frame Image Reservoir into a FAIR Dataset for Downstream AI and Machine Learning Research and Development in Precision Apiculture*

We look forward to sharing the results from these exciting research projects with the SCINet community!

SCINet/AI-COE Fellowship Mentors

Congratulations to the following five ARS scientists and their co-mentors for earning funding to host FY2024 SCINet/AI-COE postdoctoral fellows. This funding opportunity was extremely competitive this year. (Awardees are listed in alphabetical order by lead mentor's last name.)

- **Jason Abernathy**: *Application of machine learning to reproductive epigenetics for precision aquaculture*
- **Kathryn Bushley**, Ted Thannhauser, and Catherine Wram: *Secondary Metabolome of Nematode Parasitic Fungi for Biocontrol*
- **Michael Cosh**: *Developing AI-empowered climate mitigation solutions in agricultural soils*
- **Amy Hudson**: *Integrating Multiple Data Streams into Forecasts of Vector-borne Livestock Disease Emergence and Spread*
- **Kelly Thorp**, Jeff Arnold, Mike White, and Merilynn Schantz: *Modernizing the Soil and Water Assessment Tool (SWAT+) for use on SCINet computing infrastructure*

Internships Update

Many thanks to the 61 ARS scientists who applied to serve as AI-COE/SCINet graduate student internship mentors in 2024!

We have matched 29 interns with internship opportunities based on the mutual interests of each student and their prospective mentor. These interns will work on a wide variety of research projects, all of which include significant artificial intelligence/machine learning or data science components. Three interns are already working with ARS in spring internships.

We are planning an internships research symposium in the fall - we expect to announce details this summer!

Carpentries Instructor Training Program FY24 Cohort

In collaboration with The Carpentries, SCINet sponsors the enrollment of ARS scientists in the Carpentries Instructor Training Course. In this course, participants learn about evidence-based practices for effective and inclusive teaching, with a particular focus on teaching computational skills.

Congratulations to the following ARS scientists and postdoctoral fellows who were selected for the Carpentries Instructor Training Program:

- **David Bradshaw,** ORISE Fellow
- **Todd Skaggs,** Research Leader
- **Li Wang,** SCINet Fellow
- **Sarah Whitcomb,** Research Biologist

Need Proprietary Software on SCINet?

Does your research use proprietary software (i.e., software that needs a license) that is computationally demanding? Do multiple ARS researchers or their collaborators use this software, or do you imagine they would benefit from the software if it were made available to them? SCINet currently supports a select few proprietary software packages (e.g., Geneious for sequence data analysis), and is looking to identify other proprietary software that would be beneficial to provide to ARS researchers via SCINet.

Proprietary software can be installed on SCINet computing resources when the software vendor supports licensed use on our systems and when the software itself is compatible with our operating systems. From our conversations with ARS researchers at various ARS locations this year, we have learned about several proprietary software packages that might be suitable for use on SCINet. These include:

- Lasergene SeqMan Ngen
- Noldus EthoVision XT
- SnapGene
- STAR-CCM+
- Thermo Scientific Compound Discoverer.

If you are interested in using these or other proprietary software packages on SCINet, please let us know via this short software request form. Please note that the purpose of this form is to assess interest and size of the potential user base. Whether any of these software packages ultimately become available on SCINet depends on license suitability and availability of funding.

STAR-CCM+, potential SCINet software: Keshav Basnet and Thanos Papanicolaou of the USDA-ARS National Laboratory for Agriculture and Environment have contributed the following description of the STAR-CCM+ software and are particularly interested in connecting with potential users across ARS:

The direction of computational fluid dynamics (CFD) simulation in agro-environmental research is gearing towards more 3D flow dynamics of simple to very complex computational domains. Simcenter STAR-CMM+ is a multiphysics CFD package that enables scientists and engineers to model and explore the complexity and possibilities of products operating under real-world conditions. STAR-CCM+ contains an advanced mesh generator with automatic mesh refinement capabilities and offers flexibility in terms of resolving the boundary layers close to solid surfaces in very complex domains.

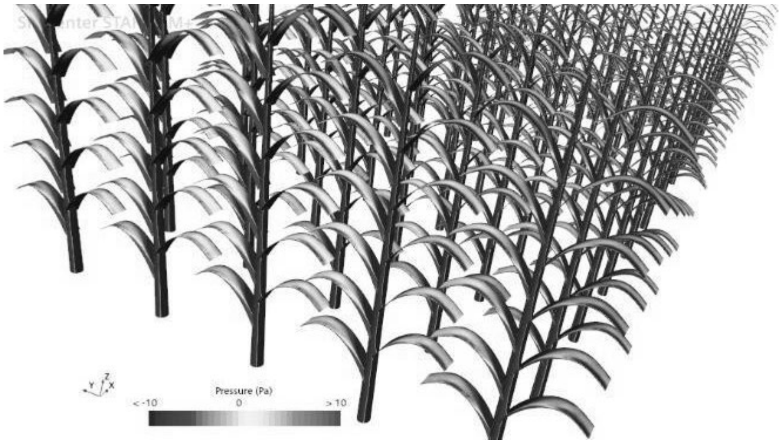


Figure 2. An example of CFD simulation for airflow past a corn field.

At present, the degree of CFD application by ARS scientists in the agro-environmental field is limited by computing power and software used. Having STAR-CCM+ available on SCINet would provide ARS researchers with easier access to CFD simulation software to better support their modeling needs. This new software availability would also expand the potential of CFD research in ARS and would offer more flexibility in accounting for the unique aspects of CFD projects related to air, water, and soil.

Again, if you are interested in using STAR-CCM+ or any other proprietary software on SCINet, please let us know.

WORKING GROUP UPDATES

Protein Function and Phenotype Prediction Working Group

Recent Publications from the Protein Function and Phenotype Prediction Working Group. Working group members are in bold and USDA Postdocs, most of whom are SCINet/AI-COE fellows, are listed with a (*).

- **Andorf, C.M.**, **Haley, O.C**.*, Hayford, R.K., Portwood, J.L. 2nd, **Harding, S**.*, Sen, S., Cannon, E.K., Gardiner, J.M., **Kim, H.-S.**, Woodhouse, M.R. (2024) PanEffect: a pan-genome visualization tool for variant effects in maize. Bioinformatics. 2024 Feb 9:btae073. https://doi.org/10.1093/bioinformatics/btae073.
- **Fagerquist, C.K**., Shi, Y., **Park, J.***. (2024) Unusual modifications of protein biomarkers expressed by plasmid, prophage and bacterial host of pathogenic E. coli

identified by top-down proteomic analysis. Rapid Communications in Mass Spectrometry. 38(1):1-16. https://doi.org/10.1002/rcm.9667.

- Hao, G., Naumann, T., Chen, H., Bai, G., McCormick, S., **Kim, H.-S.**, Tian, B., Trick, H., Naldrett, M.J., Proctor, R. (2023) *Fusarium graminearum* Effector FgNls1 Targets Plant Nuclei to Induce Wheat Head Blight. Molecular Plant-Microbe Interactions. https://doi.org/10.1094/MPMI-12-22-0254-R.
- **Poretsky, E.***, Cagirici, H.B., **Andorf, C.M.**, **Sen, T.Z**. (2024) Harnessing the predicted maize pan-interactome for putative gene function prediction and prioritization of candidate genes for important traits. Genes | Genomes | Genetics. https://doi.org/10.1093/g3journal/jkae059.
- Roger, A. Jaiswal, N., Roggenkamp, E., **Kim, H.-S.**, MacCready, J.S., Chilvers, M.I., Scofield, S.R., Iyer-Pascuzzi, A.S., Helm, M. (2024) Genome-informed trophic classification and functional characterization of virulence proteins from the maize tar spot pathogen *Phyllachora maydis*. Phytopathology bioRxiv DOI: 10.1101/2024.01.22.576543.

If you are interested in joining this group, please contact Hye-Seon Kim (hyeseon.kim@usda.gov) or Carson Andorf (carson.andorf@usda.gov).

Translational Omics Working Group

The Translational Omics Working Group is hosting an upcoming webinar:

Streamline unsupervised machine learning to survey and graph indel-based haplotypes from pan-genomes
Thursday, May 9, 2024, 11 AM - 12 PM ET

Xianran Li, USDA-ARS Wheat Health, Genetics, and Quality Research Unit, Pullman, WA

Dr. Xianran Li's research focuses on comprehending the interplay of genetics, environment, and their combined impact on crop performance. His lab investigates various aspects including genetics, genomics, enviromics, and bioinformatics. Xianran Li has authored/co-authored over 60 peer-reviewed publications and has delivered more than 40 talks in national and international conferences.

For more information or to join the working group's email list, please contact George Liu (George.Liu@usda.gov), Zhenbin Hu (Zhenbin.Hu@usda.gov), or Wenli Li (Wenli.Li@usda.gov)

Arthropod Genomics Research (AGR) Working Group

The SCINet Arthropod Genomics Working Group (AGR), along with the ARS Beenome100 Initiative, National Center for Biotechnology Information (NCBI), and the arthropod i5K Community, is hosting the Arthropod Genomics Symposium X (AGSx) Virtual Symposium 2024. AGSx comprises four separate 2-hour webinar sessions, with three co-organized and moderated by ARS AGR members and one by a team from NCBI. Each session highlights presentations from subject-matter experts and resource developers and facilitates open discussions and technology transfer for ARS and community workforce development. All sessions listed are free to attend, but registration at https://forms.office.com/g/6MCZUFWskb is required in order to receive Zoom webinar links.

AGSx 2024 sessions:

- **Session 1: Production of insects for food and feed.** Organized by Kristin Duffield, USDA-ARS, Peoria, IL USA & Brenda Oppert, USDA-ARS, Manhattan, KS USA. Tuesday, February 13, 2024, 11 AM - 1 PM ET.
- **Session 2: Insect Genome Biology and Evolution.** Organized by Lindsey Perkin, USDA-ARS, College Station, TX USA. Wednesday, April 10, 2024, 11 AM - 1 PM ET.
- **Session 3: Beenome100 and Comparative Bee Genomics.** Organized by Jay Evans, USDA-ARS, Beltsville, MD USA & Michael Branstetter, USDA-ARS, Logan, UT USA. Tuesday, April 30, 2024, 11 AM - 1 PM ET.
- **Session 4: NIH Comparative Genomics Resource: An NCBI Toolkit of Data and Tools Unlocking Arthropod Research** (Terence Murphy and team, National Center for Biotechnology Information, Bethesda, MD USA). Tuesday, May 21, 2024, 11 AM - 1 PM ET.

If you are interested in joining this group, please contact Brad Coates (brad.coates@usda.gov).

TRAINING

Genome Assembly Workshop

June 25-27, 2024, 12 PM - 4 PM ET

In this interactive, hands-on workshop, you will learn SCINet's computing resources to convert raw DNA sequencing data into a complete genome assembly. Along the way, you will also learn best practices for ensuring that your genome assemblies are robust and useful for downstream analyses. Anyone with at least a little command-line experience (you don't need to be an expert!) is welcome to join us!

In this workshop, we will cover the following topics:

- **Assembly Fundamentals**: Understand the assumptions and limitations in genome assembly due to biology, sequencing technology, and assembly tools.
- **Hands-On Coding**: Complete a genome assembly from start to finish: quality control, assembly, scaffolding, and polishing.
- **Evaluation of Assemblies**: Not all genomes are equal! Gain the expertise to critically assess the quality of a genome assembly.

To register for the Genome Assembly Workshop, please fill out the form.

Software Package/Environment Management Workshop

July 19, 2024, 12 PM - 5 PM ET

Package and computing environment management systems for R, Python, and Anaconda make it easy to install the software you need for your research projects. They also make it easy to accidentally use up all of the space in your home directory or end up with frustrating software conflicts! In this workshop presented by the SCINet Office, we will cover best practices for managing software packages and computing environments on SCINet's supercomputers. This will be a hands-on workshop that will provide you with the practical knowledge and skills you need to spend less time worrying about package management and more time focusing on your research! At least some experience with the command line will be helpful for working through the exercises.

To register for the Software Package/Environment Management Workshop, please fill out this form.

Coursera

The SCINet Office and the AI-COE are excited to provide training opportunities through Coursera. Coursera licenses are available to ARS scientists and support staff for training focused on scientific computing, data science, artificial intelligence, and related topics. Successful completion of courses and specializations result in widely recognized certificates and credentials. Please visit the SCINet Coursera Training Page to request a license. Licenses will be assigned on a rolling basis and are active for three months. Users may be able to extend their licenses upon request.

Please help us improve our training offerings!

What scientific computing training do you need? The SCINet Office's goal is to provide training opportunities and resources that meet the needs of ARS researchers, so we would be grateful if you could complete our short training request form and let us know how we can best help you learn the computing skills you need. Your feedback will help us decide where we should focus our efforts over the next year and beyond.

Training opportunities are continually being updated on the SCINet Upcoming Events webpage. For more information on any of the above trainings, registration questions, or suggestions, please email SCINet-training@usda.gov.

SUPPORT

Getting Started with SCINet Is As Easy as 1,2,3

If you do not already have a SCINet account, we hope you will consider joining the 2,300+ researchers who do. Follow the steps below to get your SCINet account.

1. Request a SCINet account to get started.
2. Read the SCINet FAQs covering general info, accounts/login, software, storage, data transfer, support/policy/O&M, parallel computing, and technical issues.
3. Register for a SCINet Forum account to connect to other users, ask questions, and learn how SCINet can enable your research.

P.S. Don't forget to complete your annual security training! This is required to maintain your account.

For technical assistance with your SCINet account, please email scinet_vrsc@usda.gov.

Support email addresses

All requests for help with user accounts, login problems, resource requests, or support for the Ceres HPC cluster should be sent to the SCINet Virtual Research Support Core (VRSC) at scinet_vrsc@usda.gov . Help requests specific to the Atlas HPC cluster should be sent to help-usda@hpc.msstate.edu.

Many emails are currently being sent to other SCINet email boxes. For the most expedient response to your support requests, be sure to send them to scinet_vrsc@usda.gov or to help-usda@hpc.msstate.edu for Atlas-specific requests.

SCINet User Tip: GPU options on Atlas

Graphics Processing Units, or GPUs, can dramatically accelerate computationally intensive data processing and analysis workflows. If you would like to apply the power of GPUs to your work, SCINet's Atlas supercomputer is the best place to do it! During the recent Atlas maintenance, the GPU nodes on Atlas were reconfigured to make using them simpler and more straightforward. There are three GPU types available on Atlas, and each is now contained in a dedicated partition/queue.

GPU type	**GPU memory**	**GPUs per node**	**No. of nodes**	**Slurm partition**
NVIDIA V100	32 GB	2	4 (8 GPUs total)	gpu-v100
NVIDIA A100, multi-instance	10 GB	56	2 (112 GPUs total)	gpu-a100-mig7
NVIDIA A100	80 GB	8	3 (24 GPUs total)	gpu-a100

To use the GPUs on any of these partitions, you must use Slurm's "gres" (general resource) argument to tell it how many GPUs you need. For example, to run a job on the multi-instance A100 partition, you might use the following arguments to request one GPU:

```
sbatch --partition gpu-a100-mig7 --gres=gpu:1 [ADDITIONAL SLURM ARGUMENTS]
```

To run your job with a different type of GPU, simply change the partition name. Note also that specifying the type of GPU is no longer required because each partition name is now specific to a single GPU type.

In order to share resources with other SCINet users, please avoid using the full A100 GPUs for jobs that do not require the GPU memory of a full A100.

Do you have tips to share? Email them to SCINet-Office@usda.gov to be included in future newsletters.

SCINet Corner

SCINet Corner is a VRSC-moderated virtual space for people to share knowledge, discuss best practices, learn about new opportunities, and explore resources to support progress on their projects.

The next SCINet Corner will be held on May 16, 2024, 1 PM - 2 PM ET. You can register for this and future SCINet Corners here.

Have a question that just can't wait? Want to see what other users are doing? Reach out to the ever-expanding SCINet Forum community for ideas, support, or just someone to bounce ideas off of at https://forum.scinet.usda.gov/.

CONNECT

The SCINet Community

Every newsletter highlights SCINet community members as a way to connect the ARS scientific computing community. To see all the SCINet community updates and review past newsletters, visit the Newsletter Archive.

Contribute

Do you use SCINet for your research? We would love to share your story! Email SCINet-Office@usda.gov to contribute content, ask questions, or provide feedback on the SCINet newsletter or website.

SCINet Office

Haitao Huang, Computational Biologist
Moe Richert, Web Developer
Lavida Rogers, Training Coordinator
Heather Savoy, Computational Biologist
Brian Stucky, Computational Biologist, Acting Chief Scientific Information Officer

SCINet Leadership Team

Brian Stucky, Acting Chief Scientific Information Officer
Rob Butler, SCINet Program Manager
Jeremy Edwards, Science Advisory Committee (SAC) Chair
Nora Lapitan, Acting Associate Administrator

Note: This newsletter is edited to comply with ARS editorial standards.

SCINet Website

Stay Connected with the USDA Agricultural Research Service
5601 Sunnyside Avenue, Beltsville, MD 20705