**Thursday, September 19**

|  |  |  |
| --- | --- | --- |
| 8:00 | Opening remarks and logistics | Deb Peters, USDA ARS, Acting Chief Science Information Officer |
| 8:15 | [Introduction to SCINet](#80g3ewwk4b1) | Steve Kappes, USDA ARS, Associate Administrator, Beltsville, MD |
| 8:30 | [Keynote: Harnessing AI to Transform Agricultural Research](#2zr2myr1m189) | Simon Liu, USDA ARS, Associate Administrator, Beltsville, MD |
| **SESSION I. Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) in the ARS** | | |
| 9:00 | [Overview of AI and ML in Agriculture](#tolq45m9zja) | Jerry Hatfield, USDA ARS, [National Laboratory for Agriculture and The Environment](https://www.ars.usda.gov/midwest-area/ames/nlae/), Ames, IA |
| 9:30 | [An AI Recommendation System for Agricultural Research](#hpfyfxpvdh1a) | Debra Peters, USDA ARS, Jornada Experimental Range, Las Cruces, NM |
| 9:50 | break |  |
| 10:10 | [The toolbox for field-scale decision making](#i4ub185kqs2r) | Ken Sudduth, USDA ARS, [Cropping Systems and Water Quality Research](https://www.ars.usda.gov/midwest-area/columbia-mo/cropping-systems-and-water-quality-research/), Columbia, MO |
| 10:30 | [Big data for big country: optimization, monitoring, and predictive analytics in western rangelands](#o0la6zici6jb) | Brandon Bestelmeyer, USDA ARS, Jornada Experimental Range, Las Cruces, NM |
| 10:50 | [Transforming Precision Sustainable Agriculture with AI/ML](#5m5d70nihw91) | Steven Mirsky, USDA ARS, Sustainable Agricultural Systems Lab, Beltsville, MD |
| 11:10 | [Overview of methods and software](#eovmav59g532) | Adam Rivers, USDA ARS Agricultural Microbiomes Group, Gainesville, FL |
| 11:40 | [Discussion: Why are you interested in AI for agriculture?](#x0ab9ppg9awu) | Moderator: Jerry Hatfield |
| 12:00 | Working Lunch and formation of discussion topics (participants purchase their own meals) |  |
| 1:30 – 3pm | [BREAKOUT GROUPS](#121ygf8detxd): Topics from lunch mtg: how is the ARS currently using AI/ML/DL? Have these talks sparked interest in other ways these approaches could be used? | Moderators: Marlen Eve, Jerry Hatfield, Jeff Silverstein |
| 3:00 | Break/[Discussion](#ye74e0mqdm5w) |  |
| **SESSION II. Challenges and limitations with AI** | | |
| 3:30 | [Deep learning (DL) in agriculture](#nhvej1uwcfjw) | Adam Rivers, USDA ARS Agricultural Microbiomes Group, Gainesville, FL |
| 4:00 | [Ethics, Bias, & Security Issues](#oj0ea9u0cxhs) | Anna Lenhart, Senior Consultant and Lead on AI Ethics, IBM Public Sector |
| 5:00 | [Discussion](#t8f6mct6bwml) |  |
| 5:30 | Poster session |  |
| 6:30 | Dinner on your own |  |

**Friday, September 20**

|  |  |  |
| --- | --- | --- |
| 8:00 | [Opening Remarks and Summary of Day 1](#6ckzijqhv0m2) |  |
| **SESSION III. High Performance Computing (HPC) and AI/ML/DL in agricultural problems** | | |
| 8:30 | [SCINet basics, introduction to SCINet resources w/training from Iowa State University](#mhpd475jke5d) (remote presentation) | Jim Coyle, Andrew Severin; Iowa State University High Performance Computing Group and USDA ARS SCINet Virtual Research Support Core (VRSC) |
| 9:00 | [Coupling machine learning and crop modeling to improve prediction in agriculture](#u33zzk4br6ue) | Sotirios V. Archontoulis, Iowa State University, Department of Agronomy, Ames, IA |
| 9:30 | [Automated Indexing and Other Machine Learning Applications at the National Agricultural Library](#whm083i5ibsl) | Paul Wester, Director, National Agricultural Library, Beltsville, MD |
| 10:00 | Break |  |
| 10:30 | [The Future of Machine Learning in Nutrition Research](#oaso791kl9uw) | Danielle Lemay, USDA ARS, [Western Human Nutrition Research Center](https://www.ars.usda.gov/pacific-west-area/davis-ca/whnrc/), [Immunity and Disease Prevention Research](https://www.ars.usda.gov/pacific-west-area/davis-ca/whnrc/idp/), Davis, CA |
| 11:00 | [Discussion and Q/A with speakers](#ox4owkm8jq36) |  |
| 12:00 | Lunch (participants purchase their own meals) |  |
| **SESSION IV. Looking forward and completion of products** | | |
| 1:00-3 | [Discussion](#hd8xx490mcop) and [Breakout groups](#mbxf5qu7za8b): what is the unexplored potential for AI/ML/DL in the ARS? Topics based on, but not limited to, both days’ talks plus own experience and needs. develop outline of white paper/journal article (Perspectives on the role of AI in agriculture) with writing tasks, dates, and authors | Moderators: Marlen Eve, Jerry Hatfield, Jeff Silverstein |
| 3:00 | Break |  |
| 4:00 | [Closing Remarks](#lcbej242n4z7) and Collection of Participant Feedback |  |

**8:15am Steve Kappes - Intro to SCINet**

SCINet/Big Data effort is now 6 years old

Initial funding was for 5 years and now SCINet is funded through the 1% assessment

Version 1 of SCINet was to set up the system

Version 2 now we are expanding and improving - storage, high speed connections, more compute nodes

Genomics community has done a good job of using the HPC but we have many more types of scientists that could use this resource, this is where Deb Peters (natural resources) comes in as CSIO

We need to know what tools and capabilities you need so that leadership can provide, but you as the scientists need to take ownership of it

Also, we don’t just want to serve the needs of the big labs only. We want to meet the needs of the small labs as well

We also need to do a much better job of getting our data and metadata stored and made publicly available

Questions

*How is all of this impacted by the fact that IT is becoming One USDA?*

For the time being, the USDA is hand off on SCINet

**8:30 Simon Liu - Keynote: Harnessing AI to Transform Agricultural Research**

[Link to TOC](#bi00qycha9kr)

What is AI? Academics might say a branch of computer science, but the government has an official definition defined by NIST

Major areas of AI study: comp sci, psych, philosophy, math, etc.

Why AI now? Matures because of 3 major forces: computer storage, algorithm advancement, data explosion

Big milestones: 1999 invention of the GPU for image based AI, 2002 amazon brings cloud storage to masses, 2004 new algorithm for coping with the data explosion - MapReduce, 2012 big advances in Deep Learning (images)

Many countries have national strategies on AI

US National AI Strategic Plan

Continue to invest

Address ethical/legal/societal implications

Building the AI workforce

More

EO13859 from the Trump Admn - Maintaining American Leadership in Artificial Intelligence

Top AI Investment Motivators

Automate repetitive/manual tasks

Improve customer experience

More

Industry recognizes that AI can improve upon other analytics techniques in many fields including agriculture

Enterprise plans to deploy AI - only 9% of respondents aren’t interested in deploying AI techniques

In Government, there are many applications for AI

satellite operations/space exploration

Financial management - hand writing recognition, budget formulations, investment

Management

Criminal justice - fingerprint/facial/retinal recog

Health - emergency medical response

Library - voice recog, national language processing

Proposed AI roadmap for ARS

AI awareness

Actively pursue AI

AI is operational

AI becomes systemic

AI is transformational

ARS AI Applications

Crop breeding and trait dissection, smart irrigation, face recog, crop health, plant

geno/phenotyping, much more

ARS is currently in the Active/Operational stage - we need to push more operational AI

Next Steps for AI at ARS  
 Need support structure

Develop AI workforce

Establish AI Technical Infrastructure

HPC, Storage, Network, AI tools/techniques, SCINet can serve as this technical

infrastructure

Cloud Services

McKinsey graphic on problem type matched with AI technique

We need to have an enterprise AI data strategy

ARS Mission: deliver scientific solutions to national and global agricultural challenges

**9am Jerry Hatfield - Overview of AI and ML in Agriculture, Opportunities for Research**

[Link to TOC](#bi00qycha9kr)

Uses of AI in Ag: planting, nutrient management, weed control, more

These applications rely on differences in reflectance

What is needed: improve algorithms and incorporation into decision tools

Can we develop management strategies to optimize our genetic resources across a range of environments?

Opportunities - phenotyping, soil variation, soil-plant-atmo interactions, environmental quality

AI application to seasonal changes in field variability

The future

Form transdisciplinary teams to use AI/ML tools to evaluated ag systems

Use research scale info to enhance our understanding of GxExM framework

Partner with producers to extend this info to field/farm assessment and decisions

Questions

*What are the AI data gathering efforts beyond images?*

Working on building different types of sensors

We need to “take the shackles off our imaginations” in terms of really thinking what data we need to collect

*How do we add value for our downstream stakeholders (growers, etc) with AI techniques?*

We need to related the data we’re collecting back to things like yield quality

**9:30 - Deb Peters - An AI Recommendation System for Agricultural Research**

[Link to TOC](#bi00qycha9kr)

Deb runs a long term ecological research LTER site - 100k hectares

In the past there were 2 weather stations on the site and someone would manually collect the data and bring it back to the lab, process it, and get it to the scientists

The number of weather stations on site has grown - still a manual process up to about 15 stations

Problem: Now there are close to 100 weather stations collecting data - way too much for someone to collection, process QA/QC

The solution- replacing human behavior with an automated system “human-guided ML process”:

Automating the QA/QC process with a series of scripts

When the computer gets stuck because it hasn’t seen something before, they still

incorporate human input into the automated system so that the computer is constantly learning

QA/QC’d data goes to national and local lab repositories for access by scientists

Scientists spend too much time re-doing what other people have already done

Questions

*What kind of AI technique do you use for the AI Recommendation System?*

Machine Learning- not using a canned software, writing the code ourselves in R

**10:10am - Ken Sudduth - The toolbox for field-scale decision making**

[Link to TOC](#bi00qycha9kr)

Farmers are embracing information technology for better decision making

Using artificial neural networks to estimate crop yield

How important to yield are a variety of environmental variables

For any given site year the neural network outperforms other “traditional” statistical techniques

ML for N management - modifying existing N management recommendation tools

the traditional recommendation tools don’t capture the variable nature of the most

economically optimized N application

Decision tree used to adjust the existing tools

Questions

*What kind of neural network are you using?*

Traditional NN with ~10 nodes

Convolution may improve performance

*How do you approach interpretability with NNs?*

Sensitivity testing

**10:30am Brandon Bestelmeyer - Big data for big country: optimization, monitoring, and predictive analytics in western rangelands**

[Link to TOC](#bi00qycha9kr)

Spatio-temporal variability has a big impact on management outcomes

Change can be noisy, abrupt, spatially variable - constantly need to observe and adjust

to sustain resources

Easy to mismanage resources

Need more data

Can we predict future events/variability based on existing data?

Big data and citizen science to predict soil classes worldwide

Large database of soil observations + landPKS mobile app data

Spatial query of potential soil components plus variety of covariates

Predict soil class using ML

Explanation and prediction of VSV outbreaks using ML

ML demonstrates multiple factors at different scales are important

<https://rangelands.app/> jornada plant cover monitoring methods applied to 50k sites generating big data

Next steps at Jornada

Predict and manage cattle distribution

ML for predicting likelihood of restoration success in space and time

DL to assess rangeland health from high res imagery

Challenges and opportunities for ARS  
 Lack of validation data leading to poor models

Can collect more data, including crowdsourced

Limited data accessibility

SCINet for AI libaries, data access, storage

Limited expertise to frame questions and develop models

Recruit ag data scientists

Questions

*Have you made efforts to work with native populations and their lands?*

Yes, we have a tribal liaison through the SW Climate Hub. We are working on developing partnerships

**10:50 Steve Mirksy - Transforming Precision Sustainable Agriculture with AI/ML**

[Link to TOC](#bi00qycha9kr)

Grand challenge: feed the world despite destabilizing factors

We hope precision ag will get us there

Precision sustainable ag requires geospatial solutions

Sensing technology

Real-time data integration

Analytics (AI/ML) and visualization

How do we build the datasets we need?

On-farm monitoring

Tech (remote sensing)

Decision support tools

communications/data sharing platforms

Researcher, farmer, and agricultural professionals networks

Phenocams for stress monitoring

Determining the moment when plants become stress/destressed through AI on imagery

and feeding that information into decision support/management tools

Weed species distinction and mapping

ML to improve cover crop biomass predictions

ML to characterize Q&Q biomass

Resolution matters - satellite imagery can be limited - need drone data

Question

*Do you sensors measure air quality?*

Not currently working with air quality sensors

*Can you develop reusable AI tools that others can use?*

Collaborating with microsoft, ESRI, ag producers, more to make the tools they are developing relevant and useful

**11:10 Adam Rivers - Overview of methods and software - AI, ML OH MY! A Roadmap to Methods and Concepts**

[Link to TOC](#bi00qycha9kr)

AI: tech that has ability to reason and make decisions when you give it a set of information

Machine Learning is a type of AI- predicting, classifying, clustering, simplifying, many methods exist

Types of ML techniques

Classification

Regression - prediction of 1 or more values

Clustering - unsupervised method

Dimensionality reduction

What makes a good ML problem?

You want to classify, regress, cluster, or simplify

You have a large number of independent variables

You have enough data

ML - types of learning methods

Supervised - labeled data

Unsupervised - no labels

semi-supervised/active learning - some labels but not all

Reinforcement - gives quality info about each decision and iterates

ML terms

Label - predicted value

Feature - ind variable

Example - single data point

Training and evaluation - must split your dataset for training, validating, and testing

learning is the process of optimization

The variance bias tradeoff - overfitting not very useful, underfitting not capturing all the vairance

Learning curves - training set size vs error

Most of the real work of ML is data cleaning

Need generally normal distributed data

ML by method

Regression methods

Neighbor-based methods - PageRank, Blast, K-nearest, more

Regularization methods - L1, L2, Elastic Net, more

Decision Tree methods

Ensemble methods

Bayesian methods

Neural Network methods - many different kinds

Key technical points for leaders

Know your goal - ex: predict or classify?

Use ML when you want to: classify, regress, cluster, or simplify

Know the families of ML methods

Lookout for pitfalls

over/under fitting

Unclean data

More

Questions

*How do you make sure your selecting the right method for your data?*

Look at the weights after you fit a model, plot the learning curves to see if you need a more or less complicated model, diagnostics

*I think there is great value for using ML on smaller datasets. You don’t necessarily need big data, what are your thoughts?*

Some methods are limited to big data but not all

**11:40am - DISCUSSION - Why are you interested in AI for agriculture?**

[Link to TOC](#bi00qycha9kr)

Water quality interest - 1 side want higher quality to make better products - the other breeders that want to get to the higher quality. How do we take the physical traits which we can measure and translate that into genetic markers for our breeders?We need to deconvolute the physical traits somehow using ML

As RLs without AI expertise we have to understand it well enough to be able to support our scientists using these methods

Ecological modeling and prediction of invasive species spread. Combining with transportation models. Trying to make this info more useful for land managers and regulatory agencies

From an ONP perspective, not sure how much AI tools are actually being used because of the various ways that scientists describe their methods. Also, do we have the right tools to get the right people at ARS - what do we do in terms of position descriptions?

From library perspective, sitting on 100yrs of digital info - how can we mine that to link this info to current research?

Interest in the health of humans in addition to the environment. Trying to move away from belief based health and move toward science based nutritional info

How to use ML for nutrient modeling? Some necessary software is not available on SCINet. Matlab, ArcGIS, ENVI

Simon Liu: What is the output of this conference? Would like to see this conference come up with a bunch of papers for a special issue journal/magazine - IT Professional - instead of just a conference proceeding paper. The special issue could be used in many different ways by managers, and capitol hill. Can count as 1 of the 2 required pubs per year. 3000 words plus 3-5 figures for the long, short article is 1200 words with 2-3 figures. THe presentations have shown many case studies that could be published in the special issue. Proposed submission date - submit articles by Dec 1 and get the special issue published in March 2020. 1 overview, 1 tutorial what is AI/ML, few case studies.

There is real power in special issues. A few years back ARS had a special issue in Remote Sensing that put ARS in the spotlight as a leader in remote sensing research

This conference was sponsored by SCINet. The other workshops are creating working groups to continue workshop efforts

**1:30pm** **Breakout Groups**

[Link to TOC](#bi00qycha9kr)

What are the challenges and solutions for moving ARS up the AI progress diagram?

4 groups

Support structure - Marlen Eve - transfering info to stakeholders

Workforce development - Jeff Silverstein

Technical Infrastructure - Brian Scheffler - hardware and software

Data/Model Strategy - Jerry Hatfield

**Breakout Group Report Out: Workforce Development (GLen Moglen)**

Strategies for both new hires and people already at the agency

Challenges:

For new hires

speed and volume of hiring,

do the right positions descriptions exist- we probably need new descriptions

Should this be cat 3 or 4

Should new hires have domain science background in addition to AI expertise,

These AI experts would function best in RUs as opposed to areas

For existing employees

Trainings

Informal work sessions/information sharing

Development of a list of existing online AI resources/tutorial

Incentives for trained people to information share with their groups

AI and existing IT support separate from OneUSDA consolidation

More support in the virtual research support core

**Breakout Group Report Out: Data/Model Strategy (Jerry Hatfield)**

In terms of data we need to learn how to support collaborative teams

How to replace our statisticians- many are retiring- new generation of replacements could include AI expertise

Supply chain of data - how do we build more effective metadata and make it more accessible

Need more data sharing

Need more and higher quality metadata, especially important for meta-analysis, data can be useful in a larger context than how it was collected

**Breakout Group Report Out: Support Structure (Marlen Eve)**

AI ownership

Hard to patent/license data products - public data turned into private apps/tools that cost

Data cites

How does ARS get credit

Data transfer agent?

New OTT responsibility for managing/protecting data/tools for some maturation term

tools/data registered API with embedded cookie that reports when the data has been

used

AI Community of Practice

Develop an LTAR WG around AI/ML to model the sharing of info resources

Ag Library develop best practices, code repository for AI

Develop standards and communication strategies

Develop something like the climate hubs structure to enable appropriate deployment of \

AI and SCINet

AI priorities and strategies

Needs of stakeholders are diverse and uncoordinated, have more leadership at the ONP

level to create more formalized strategic alliances with stakeholders/other agencies

What is “user friendly” AI data?

**Breakout Group Report Out: Technical Infrastructure (Brian Scheffler)**

Need to manage the entire data workflow and connect separate data-related issues

We need hands-free data transfer and automated processes

Communication - across RUs and disciplines

Training

Development of Standards of practice

How to handle data analysis in partnerships - for data on university computers who owns it and security issues

How can we talk about hardware needs when it’s not clear what is available on SCINet

What about the farmers in the field and getting their data into our systems

**3pm DISCUSSION**

[Link to TOC](#bi00qycha9kr)

Deb: over the next 5 years there will be 11 SCINet funded ORISE postdoc who will help RUs implement some of the recommendations from these workshops, 2 year positions, want to indoctrinate them into ARS after the 2 year period

Steve Kappes: must change the attitude SCINet doesn’t have \_\_\_\_ so I can’t use it, to I need \_\_\_\_\_\_\_ on SCINet and then find others who need the same and work with Deb Peters and VRSC to find the solution

Simon: 2210’s taken by OCIO IT, 1550 Computer Scientist could be better for ARS to use, we can leverage postdocs to get the scientists we need quickly- the process for hiring has been streamlined, there is no open competition required. 2020 goal to recruit 1400 people including postdocs and contractors

*Do postdocs need to go through the process of getting prioritized?*

Simon: You can hired postdocs as long as you have funding

**Deb: How do we move from individuals doing AI to ARS AI enterprise?**

**For tomorrow each group needs to prioritize the top two challenges and top two tasks to meet these challenges**

**3:30pm Adam Rivers - Deep Learning**

[Link to TOC](#bi00qycha9kr)

Neural Network Basics

Data observations come in

Multiple hidden layers computing, many connections between individual components

and layers

1 output layer

Each individual component (neuron) has two parameters: weight and bias

How does learning (training) occur?

Randomly set weights/biases

Run samples to get predictions

Calculate cost

Backpropagate error

Output the gradient of the cost function

Adjust the parameters and run again

Why add layers?1 layer represents, but doesn’t learn, better generalization with more layers (depth)

Regularization is still needed with many layers - randomly select a set of nodes (dropout) to see how the NN output changes

Convolution neural network for 2 or 3D data (operating on pixels in an image for example)

Recurrent NN and long-term short-term memory (LTSM)

Attention Methods

Generative adversarial networks

Autoencoders - used to reduce dimensions and handle missing data, can reproduce input data into smaller dimension/size (ex to make files smaller than gzip)

Agricultural applications for DL: phenotyping, GIS image processing, crop census, multispectral processing, animal health, edge computing

There are very cheap sensors and tools for applying deep learning

**4pm Anna Lenhart - AI Ethics, Bias, Security Issues**

[Link to TOC](#bi00qycha9kr)

AI Ethics and Risk Zones

9 risk zones: Spread of misinformation, online addiction, economic inequalities, machine

ethics and algorithm bias, surveillance state, Data Control & monetization, implicit trust

and user understanding, hateful and criminal actors (hackers), energy use and the environment

AI is used in college admissions, loan qualification, granting asylum - sometimes the algorithm decisions are not fair

There’s a lot of human input to AI which contributes to bias

Sample bias - data doesn’t represent environment fully

Historical data that encodes stereotypes

Choice of AI algorithm

Feature selection

Handling of missing values

Decision to use proxies

Lack of diversity in tech

Guidelines

Categories of AI principles: <https://bit.ly/34bcivV>, high level guidelines

IBMs AI Ethics Guiding Principles

Accountability - whoever creates the AI system is accountable for the actions of that

System, can the user potentially misuse the tools and how to build this into the algorithm

Value Alignment - collaborative development, AI can’t just be developed using the values

of the programmers

Explainability -Blacbox - could mean proprietary, it’s a policy problem. Could mean very deep algorithm to be understood, this is a research problem

Fairness - “i don’t include \_\_\_\_\_ in my model, so it’s fair” is not good enough, for example your model could still be racist or discriminatory. “Fairness through awareness” trying to test the fairness of models.

User Data Rights - privacy issues

EO on Maintaining American Leadership in AI

Algorithm Accountability Act 2019 - bill in congress - algorithm impact assessments, think EPA environmental impact assessment. NY State has passed similar bill already

General Data Protection Regulation (GDPR) 2016 - US companies that work internationally have to follow. Other similar policies being developed by states will likely result in a federal policy over the next few years

Overview of risk mitigation techniques

Design - techniques that leverage social science and diversity into algorithms

Technical - statistical metrics to identify biases

Universal best practices

Type of model - supervised vs unsupervised, deep vs shallow

Type of data -

Assessing Risk - risk assessments/rubrics

More human in the loop = less risky

Working with vulnerable/minority populations = more risky

Black box = more risky

Will behavior change without human intervention? = more risky

More diversity on your team = less risky

You can start on ethiccampus.org, free tools

AI ethics + Design Thinking

Framing risk

Considering data user rights

Mapping stakeholders & systems to understand explainability, fairness, values,

Accountability

Data deep dives - where is discriminatory bias encoded

Unintended consequences for all types of errors

Map out your stakeholder to think about explainability

In the SNAP eligibility algorithm example the users (employees managing benefits) they

likely need to know why two similar looking employees had different algorithm results -

one eligible, one not. Simply Releasing the full algorithm code won’t be helpful to there benefits managers

Technical Toolkits: AIF 360 from IBM, Google Fairness tools

Issue with “Noise attacks” can throw off a DL algorithm - strategic addition of noise can change the output of a DL algorithm where the human eye couldn’t tell the difference. Example given of a panda photo plus a bit of noise then identified by DL as a gibbon, still looks like panda to human

Fact sheets/”nutrition labels” for AI algorithms would include info on purpose, basic performance, safety, security, lineage

Questions

*Explainability matters for us for different reasons, what are the new tools for interpretability?*

Lime- what if, change one thing run model again

DARPA is working on explainability layers but the best method currently may still be tweak and run

*Does the DL noise hack issue mean that we need to keep track of a chain of custody of our images?*

We’re still trying to figure out how to protect against this. The best course of action is to protect your data from hackers as best you can

If the magnitude of consequences is high maybe AI isn’t worth it

*How to reconcile the push for public data with the risk of hacking?*

**DISCUSSION**

[Link to TOC](#bi00qycha9kr)

*Can we define use cases for DL and get someone knowledgeable to train algorithm and get others at ARS to use? This is a workflow method for getting ARS people trained and implementing AI techniques*

It’s a great idea

*We must be careful with free DL algorithms (CNNs) that were developed on non-agricultural imagery. They can very easily incorrectly identify crops*

**09/20/2019**

**Opening Remarks**

[Link to TOC](#1qcrkphyefbg)

For special issue articles send the following info to Deb/Kerrie by Fri 9/27:

Goal of the research, AI method used, key result, are you willing to lead the paper yes/no

Deb and Jerry will look through this info and decide how to go forward. Could be more than 1 special issue

Get your travel reimbursement documentation submitted right away after the conference

**8:30am - Andrew Severin & Jim Coyle - SCINet basics, introduction to SCINet resources w/training from Iowa State University (remote presentation)**

[Link to TOC](#1qcrkphyefbg)

SCINet components: high-speed network, high-performance computing system, Virtual Research Support Core (VRSC)

VRSC- maintains system and helps with knowledge transfer and applying informatics to your research, guide efficient use of SCInet, enable researchers to translate big data to informative data (won’t do your analysis for you but can help you understand where to start/set up a workflow), acts as a knowledge repository, supports SCINet-funded workshops

1 node on a supercomputer (HPC) has is about 10 times as powerful as your personal computer and has more RAM, the ARS HPC has hundreds of nodes

An HPC cluster like Ceres at ARS is comprised of a login node, data transfer node, hundreds of compute nodes controlled by a job scheduler, storage

Ceres isn’t a very large supercomputer but we are expanding and research groups are able to purchase private nodes as well

Why use Ceres?

Large data, sharing data, collaboration, data integration from multiple researchers, want

to create a saved/portable environment for running code that never becomes outdated

(singularity containers)

There are hundreds of software packages on Ceres already and you are able to request additional software to be installed by contacting the VRSC [scinet\_vrsc@iastate.edu](mailto:scinet_vrsc@iastate.edu)

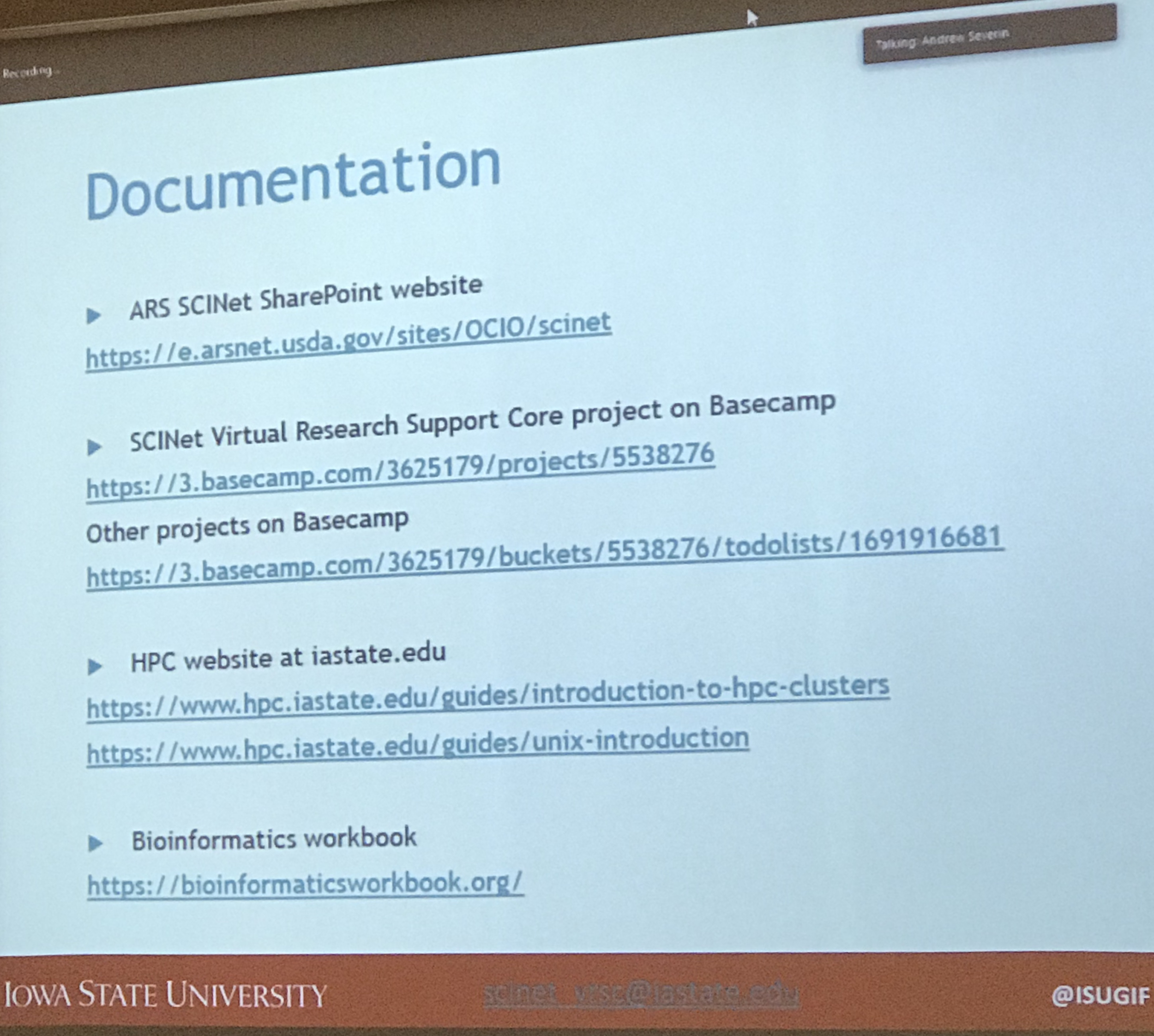
[www.bioinformaticsworkbook.org](http://www.bioinformaticsworkbook.org) created by the VRSC contains tutorials, data management guidance, more. Developing a similar resource for the geospatial community. Suggestions for additional tutorials/info to add to these resources are welcome - contact the VRSC

Ways to get help with SCINet/Ceres:

Post to Basecamp - not for individual issues like logging in

Contact VRSC, [scinet\_vrsc@iastate.edu](mailto:scinet_vrsc@iastate.edu)

See picture below for links to SCINet documentation and tutorials



Questions

*Does VRSC has AI expertise to help researchers?*

No

*What technical infrastructure would we need to do more AI research on Ceres?*

Likely more GPUs, only 1 private GPU currently

*Does VRSC offer support for parallelization of research codes?*

Yes

*What are the hurdles for data transfer of large data between ARS locations?*

Higher speed connections and routing issues

Simon: we have a highway with cars limited to drive 10mph, how to get them up to 60 mph

Jim Coyle working on this, Andrew doesn’t know the details

**9am - Sotorios Archontoulis - Coupling machine learning and crop modeling to improve prediction in agriculture**

[Link to TOC](#1qcrkphyefbg)

Why were simulation models developed?

To break complicated things into components and for simulating things we can’t measure

Models can help us with process-based understanding

Why link ML with crop modeling?

More accurate predictions, faster/cost effective, utilize more data

Can we predict corn yield and N leaching at planting time?

Not enough data to train ML models - use the Ag model to increase size of database

How much data is needed for ML? Which input variables are the most important? Which ML technique is best suited? Can we use ensembles of ML techniques?

ML can be used as a predictive and explanatory tool

Techniques may be limited by data availability

Questions

*Do the farmers implement your modeling/ML recommendations?*

They didn’t make explicit recommendations but they provided information for the farmers to make their own decisions

*How good is the simulated data that is being input to the ML models?*

Modeled yield data error ~15%, N leaching modeled error ~25%, good enough

**9:30am - Paul Wester - Automated Indexing and Other Machine Learning Applications at the National Agricultural Library**

[Link to TOC](#1qcrkphyefbg)

Indexing and cataloging has traditionally been manual and very human intensive - volume of library resources now requires automation

ML indexing of articles starting in 2012 with only 5 staff, ability to index half million

articles per year, staff still used to check the quality of the ML indexing

The ML indexing process works using the NAL Thesaurus (NALT) terms

Ag Data Commons - working on ML technique to automatically find “more like this” articles (neural net)

ARS Program Portfolio Analysis

Natural language processing collab with ONP/OSQR for identifying research

collaborators, research overlap, more

Questions/COmments

*Simon Liu: We had many vendors offering the Library solutions, but the major problem in the beginning was false positives/negatives. Vendors will promise anything, but you need to have metrics for them to meet or their solutions will likely under-deliver*

*Can we use NAL’s portfolio analysis to find other ARS scientists who work on similar research?*

We are starting with a couple of national programs and are almost ready to demo for ONP. Over the next year we plan to expand to all the research proposals. We plan to have visualizations as well so you could for example see maps of researchers working on similar research - visualize the research network.

*Why do we have to submit the same information about our pubs twice, once to ARIS, once to NAL?*

NAL are working to streamline this so that you only have to submit once and it will be in the ARIS system and also available as full text at NAL

**10:30am - Danielle Lemay - The Future of Machine Learning in Nutrition Research**

[Link to TOC](#1qcrkphyefbg)

In the human nutrition studies they conduct, hundreds of samples from each participant are collected. Using ML to predict lactose intolerance

Trying to predict body mass from dietary and physical data

Can nutrient/nutrition be captured through images?

Not really, there is wide variation in nutrient content for meals that look very similar

But for written food diaries, AI could be very useful by producing labels for the written

food dietary entries that a participant could easily verify quickly

Challenges for AI in Nutritional Research

Need more high-quality data

Could implement more AI-assisted observational studies

How AI/ML can enhance nutrition research

Improve dietary intake assessments

Identify specific features important to health outcomes

Improved models to predict what individual people should eat

**11:15am Discussion and Q/A with speakers**

[Link to TOC](#1qcrkphyefbg)

Potential papers - “short publicity articles” - showcase current AI ag research and larger context

* How does using AI in agriculture impact our stakeholders
* Theme of linking traditional statistics and modeling to new AI approaches - the

opportunities and challenges around doing this. Limitations of statistical approaches,

opportunities of AI approaches

* AI approached for individual sites / farm level (compared to commercial approaches)
* AI approaches for regional patterns/questions
* AI for different data structures text vs tabular vs images
* Taxonomy of approaches that span across nutrition, crop/range management, etc. This could be included in the overview
* Temporal variability - how does AI deal with changes over time

Specific Suggestions for Paper Content

* In overview paper: Make sure to include discussion of the use of AI techniques and the impact on the end user (farmer), do these techniques impact the ability of farmers to understand/use the tools we develop and that they approve of, also that the farmers don’t end up with a lot of different tools they need to use
* How to use AI research to the benefit of our USDA regulatory agencies
* Overview paper to present a vision of the future
* Making sure we present AI not as the end product but as a tool we integrate into our research toolkit

*Where is the impact of AI/ML in agriculture?* *The talks were educational but didn’t come away with a sense of where the major impact is of using AI*

Regional explanatory results, can also predict, better information to regulatory stakeholders

Improvement to individual management decisions

What else?

*It may be a better idea to bin our papers around outcomes instead of AI approaches.*

**1pm - Discussion Prior to Breakout Groups**

[Link to TOC](#1qcrkphyefbg)

We want to make sure we explain the benefit to our stakeholders in these papers and

We want to include an ask

*Explain the choice of IT magazine for the special issue publication?*

Deb will send an issue of the magazine out so people can get a feel for it - IT Professional

The mag is for IT professionals, very readable, peer reviewed, Simon is involved and has already been talking to the editors, will be a quick turn around

Special issue topic will be “Harnessing AI to Transform Agriculture”

*How many articles?*

1 overview, 1 tutorial - may not need it, 3-4 case articles

*A suggested paper flow:*

3 major components: define the problem, discuss the impact/potential impact, conclude with the elements still missing to address the full problem

*The ability to perform comprehensive meta-analysis is needed. How to integrate different data sources, have common tags/metadata*

Info sheet distributed to the Group:

AI serves ~~three~~ four purposes

· Showcase current Ag research to attract funding (outward looking)

· Inward looking, building collaborations within ARS

· Attract outside people to work with us

· shows public benefit

Potential cases: ag-problem driven

· Impacts on stakeholders (how to take AI research to make it actionable)

· On-farm or site-based questions

· Regional patterns/questions

· Larger context – how to extend to larger area; scalable approaches

· Temporal variability (climate) –

· AI for integrated solutions/data fusion

Overview topics:

- Vision for the future for different areas, show potential for AI as a tool

- Lifecycle of AI in Ag research: highlight challenges, aggregated maturity levels, to increase acceptance (congress, farmers)

Methods

· Structure of the data text, tabular, image (structured, unstructured)

· Traditional statistics/AI/ML hybrid

- Taxonomy of approaches – types of challenges across different areas of research

**1:30pm - Breakout Groups for discussing/outlining papers**

[Link to TOC](#1qcrkphyefbg)

Site/Farm Level Paper

Regional/Continental Scale Patterns Paper

Overview Paper

**On Farm Report Out**

3 topics that could be success stories

ML for interpreting water stress measurements and making irrigation decisions

N management systems

Animal welfare (image analysis)

Carcass quality

Individuals have been assigned to follow up on these topics and they plan to meet again

**Regional Report Out**

Overarching theme- how AI can allow use of big data for precise predictions/estimates and tools that can refine interventions

VSV, ecological niche modeling

Precise estimates of crop performance (Mirsky)

Soil class estimates

Conclude with challenges and the ask

**AI for Data Integration/Fusion**

Bringing in the NAL work, but not NAL-only pub

Group didn’t meet, but will continue discussions together and with Deb (Jennifer, Cyndy, Danielle Lemay)

**4pm - Simon Liu - Concluding Remarks**

[Link to TOC](#1qcrkphyefbg)

This group can serve as a model for USDA progress