Citizen Science at UND

Travis Desell
Department of Computer Science, University of North Dakota

Biology Seminar

March 6, 2015

University of North Dakota

1. What is Citizen Science?

2. A Case for Volunteer Computing

3. DNA@home

4. Wildlife@Home

4. Future Work

5. Questions?

What is Citizen Science?

What is Citizen Science?

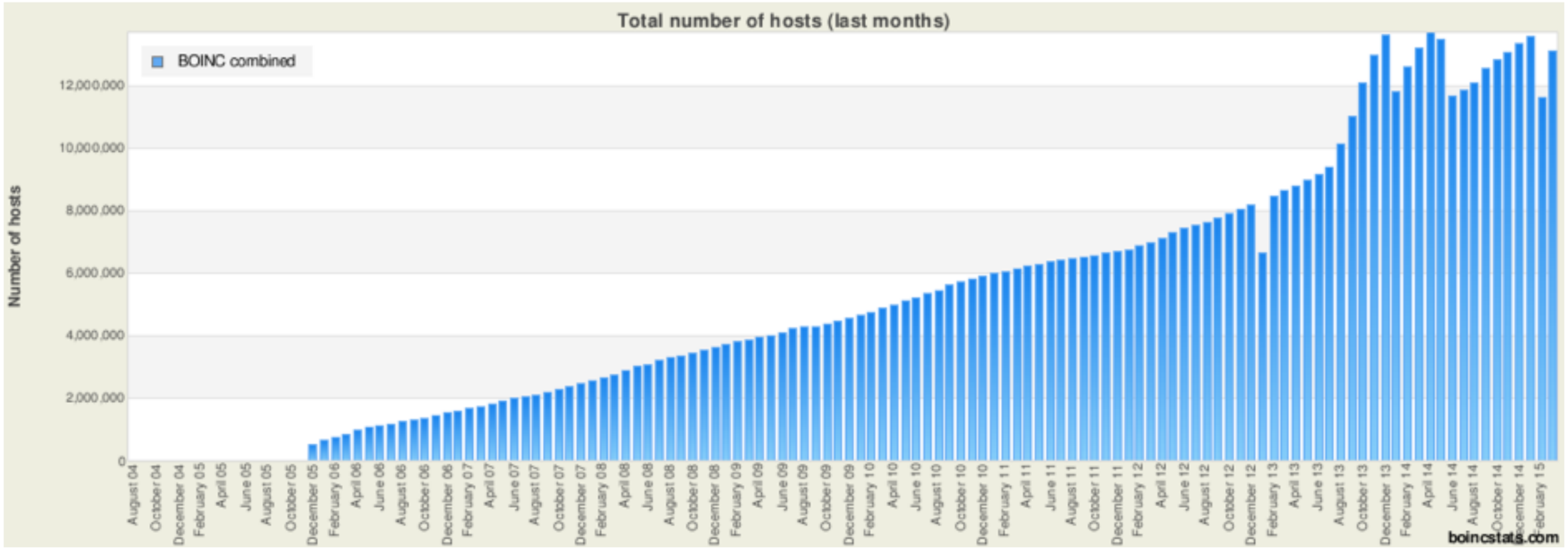
Directly involving the public in science!

Crowd sourcing: people volunteer their brains to provide or analyze scientific data.

Volunteer computing: people volunteer their computers to run tasks to solve scientific problems.

A Case for Volunteer Computing

Combined BOINC Statistics

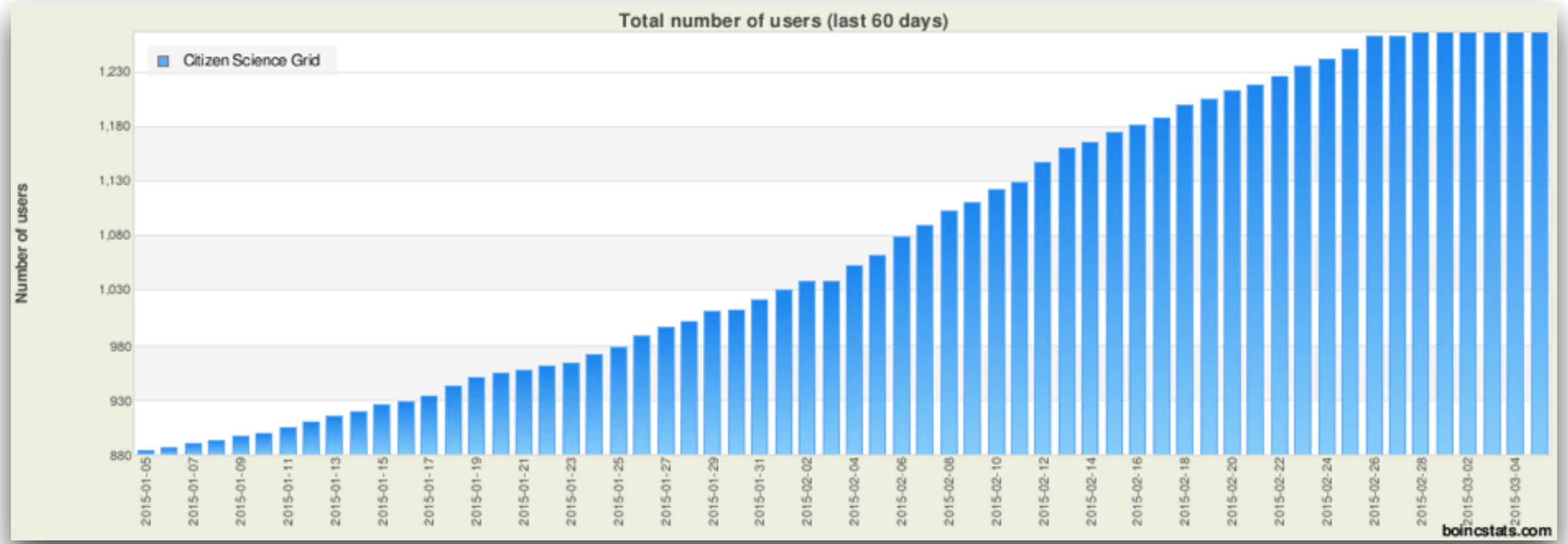


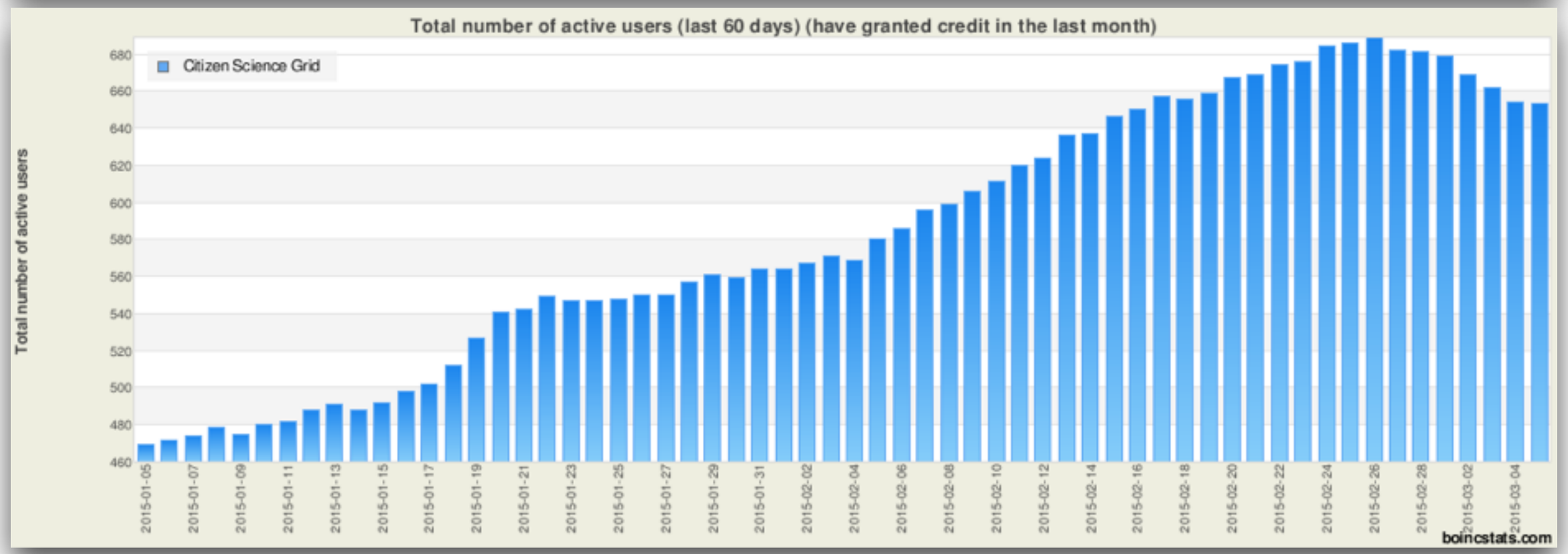
Currently, 285,705 active users are providing around 160,732 TeraFLOPS of computing power (as of last night).
Over 3,311,372 users have participated in BOINC.

The worlds fastest supercomputer (top500.org) currently has 3,120,000 cores and provides 33,862.7 TeraFLOPS. The second has 560,640 cores and provides 17,590.0 TeraFlops.

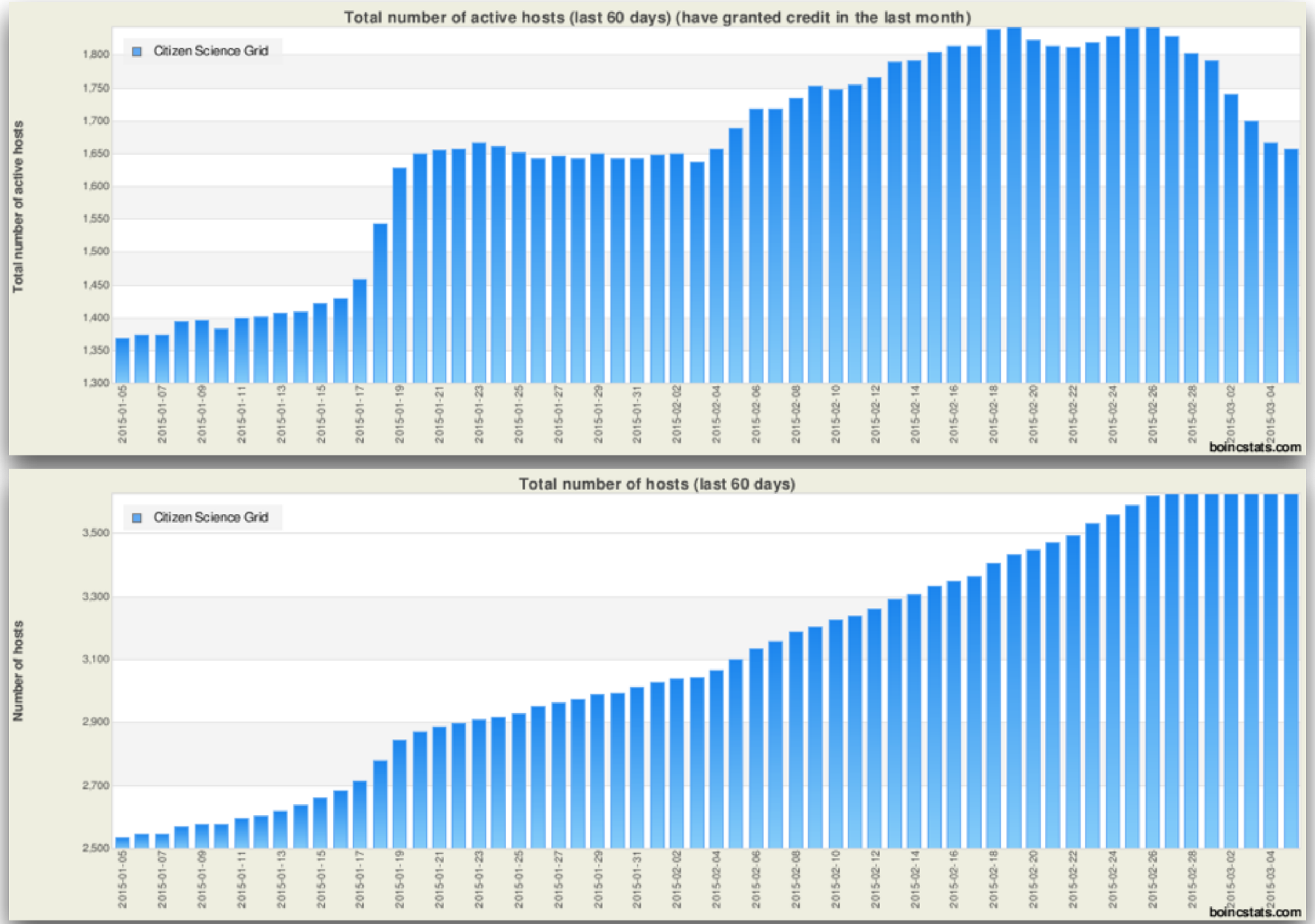
Figures from: http://boincstats.com/stats/project_graph.php?pr=bo&view=hosts

Citizen Science Grid Users





Citizen Science Grid Hosts



Citizen Science Grid Statistics

In the last couple months, ~1000 volunteers have volunteered ~2000 computers to participate in DNA@Home in our current analysis (more on that in a bit).

The DNA@Home application is available for 32 and 64 bit versions of Linux, OS X and Windows.

We are currently gearing up to send out more Wildlife@Home work and are developing a new version of the SubsetSum@Home application for use on GPUs.

What's Volunteer Computing Good For?

What's Volunteer Computing Good For?

Volunteered computers can't easily talk to each other (firewalls, security, etc), and even if they could the latency is very high.

This limits things to "Bag-of-Tasks" (embarassingly parallel) problems.

However, some algorithms can fit in this model with some modifications, such as numerical optimization (for example evolutionary algorithms, below).

Problems like many simulations which require tightly coupled communication between processors do not work well. Luckily, we have a cluster for that! (But that's another lecture.)

Travis Desell, David P. Anderson, Malik Magdon-Ismail, Heidi Newberg, Boleslaw Szymanski and Carlos A. Varela. **An Analysis of Massively Distributed Evolutionary Algorithms**. *In the Proceedings of the 2010 IEEE Congress on Evolutionary Computation (IEEE CEC 2010)*. pages 1-8. Barcelona, Spain. July 2010.

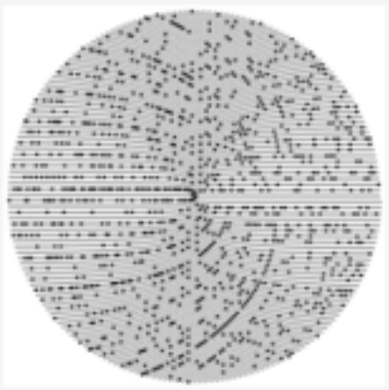
Citizen Science Grid

The University of North Dakota Citizen Science Grid is run by Travis Desell, an Assistant Professor in UND's Computer Science Department. It is hosted by UND's Computational Research Center and Information Technology Systems and Services. The CSG is dedicated to supporting a wide range of research and educational projects using volunteer computing and citizen science, which you can read about and visit below.

Volunteer Your Computer | Volunteer Your Brain

DNA@Home

The goal of DNA@Home is to discover what regulates the genes in DNA. Ever notice that skin cells are different from a muscle cells, which are different from a bone cells, even though all these cells have every gene in your genome? That's because not all genes are "on" all the time. Depending on the cell type and what the cell is trying to do at any given moment, only a subset of the genes are used, and the remainder are shut off. DNA@home uses statistical algorithms to unlock the key to this differential regulation, using your volunteered computers.



SubsetSum@Home

The Subset Sum problem is described as follows: given a set of positive integers S and a target sum t, is there a subset of S whose sum is t? It is one of the well-know, so-called "hard" problems in computing. It's actually a very simple problem computationally, and the computer program to solve it is not extremely complicated. What's hard about it is the running time – all known exact algorithms have running time that is proportional to an exponential function of the number of elements in the set (for worst-case instances of the problem).

Wildlife@Home

Wildlife@Home is *citizen science project* aimed at analyzing video gathered from various cameras recording wildlife. Currently the project is looking at video of sharp-tailed grouse, *Tympanuchus phasianellus*, and two federally protected species, interior least terns, *Sternula antillarum*, and piping plovers, *Charadruis melodus* to examine their nesting habits and ecology.

Notice for DNA@Home and SubsetSum@Home Users

DNA@Home, SubsetSum@Home and Wildlife@Home are now sub-projects of Citizen Science Grid. All workunits for these sub-projects will be sent out from the Citizen Science Grid project. You can link your old DNA@Home and SubsetSum@home accounts to your account on Citizen Science Grid by visiting the link accounts webpage. This will copy the credit over from the old projects to your account here. You'll need to detach your BOINC client from these old projects and attach to Citizen Science Grid.

User of the Day

Defender
I like chicks!

News RSS

[wildlife] error checking on video watching page

I've updated the video watching page to disable the finished button when the observations have missing data or there are no observations. Let me know if you're having any problems with it!

Travis Desell on Friday, February 27th
leave a comment

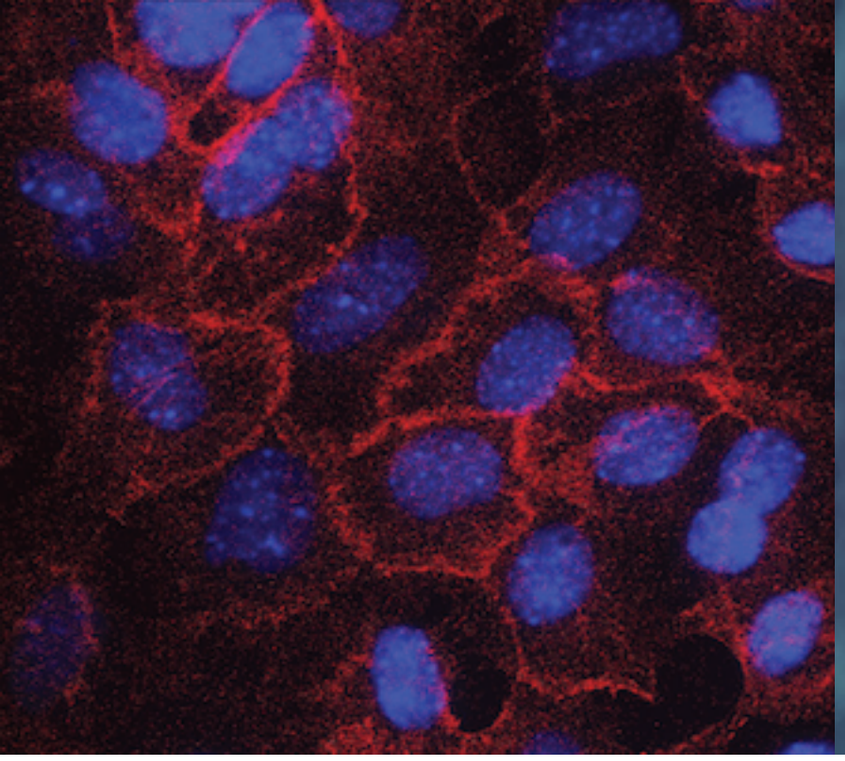
[wildlife] New Badges

Hello fellow Biologists!

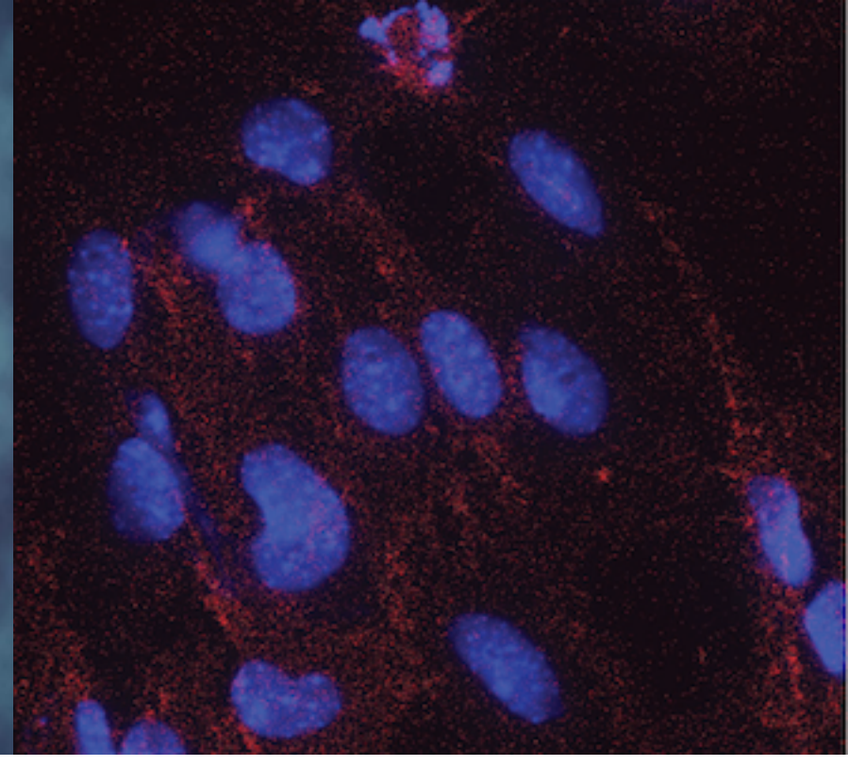
I'm Kelly Sagen, the newest member of the Wildlife@Home team! Currently, I am being reigned in as "the Badge Creator!" That's right! Many of you are putting your passion for biology to work and have earned the highest level of badges available and now new badges are on the way!

So, let me introduce a little more about myself. I'm currently wrapping up my Bachelors degree at the University of North Dakota and have been involved with two research projects through Species Pattern and Community Ecology (SPaCE) and San Diego Zoo Global's Institute for Conservation Research.

Through it all, I've had a passion for photography



E-cadherin protein (stained in red)
before Snail expression.

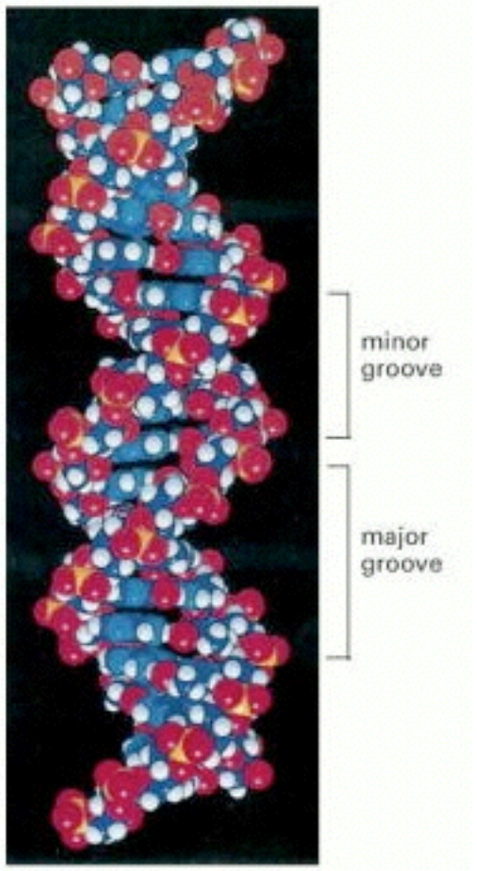


E-cadherin protein (stained in red)
after Snail expression.

DNA@Home

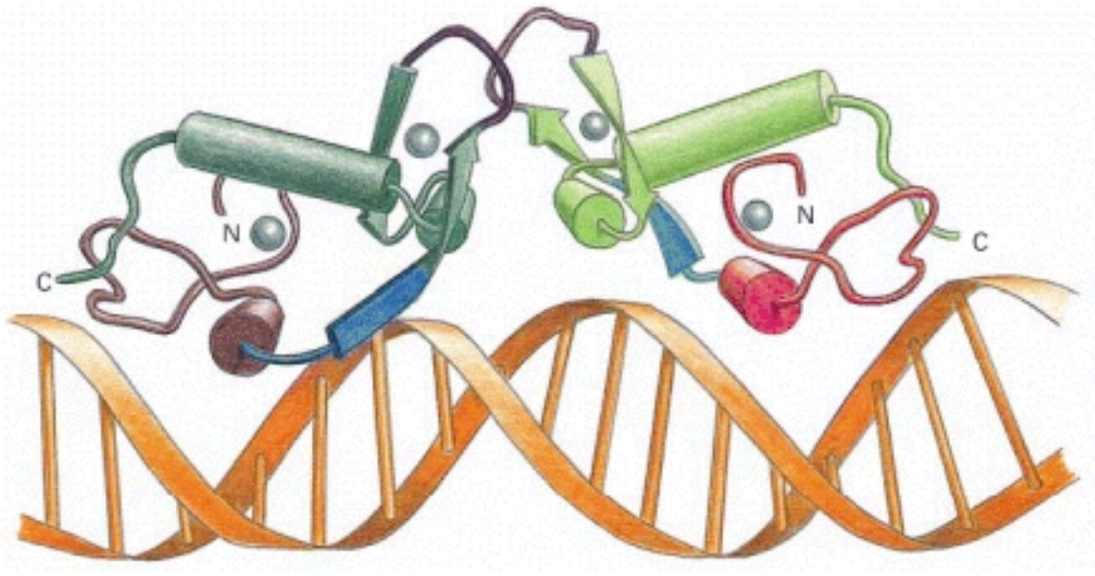
Travis Desell, Archana Dhasarathy & Sergei Nechaev
Departments of Computer Science & Basic
Sciences (Medical School)
University of North Dakota
http://volunteer.cs.und.edu/csg/dna

DNA@Home



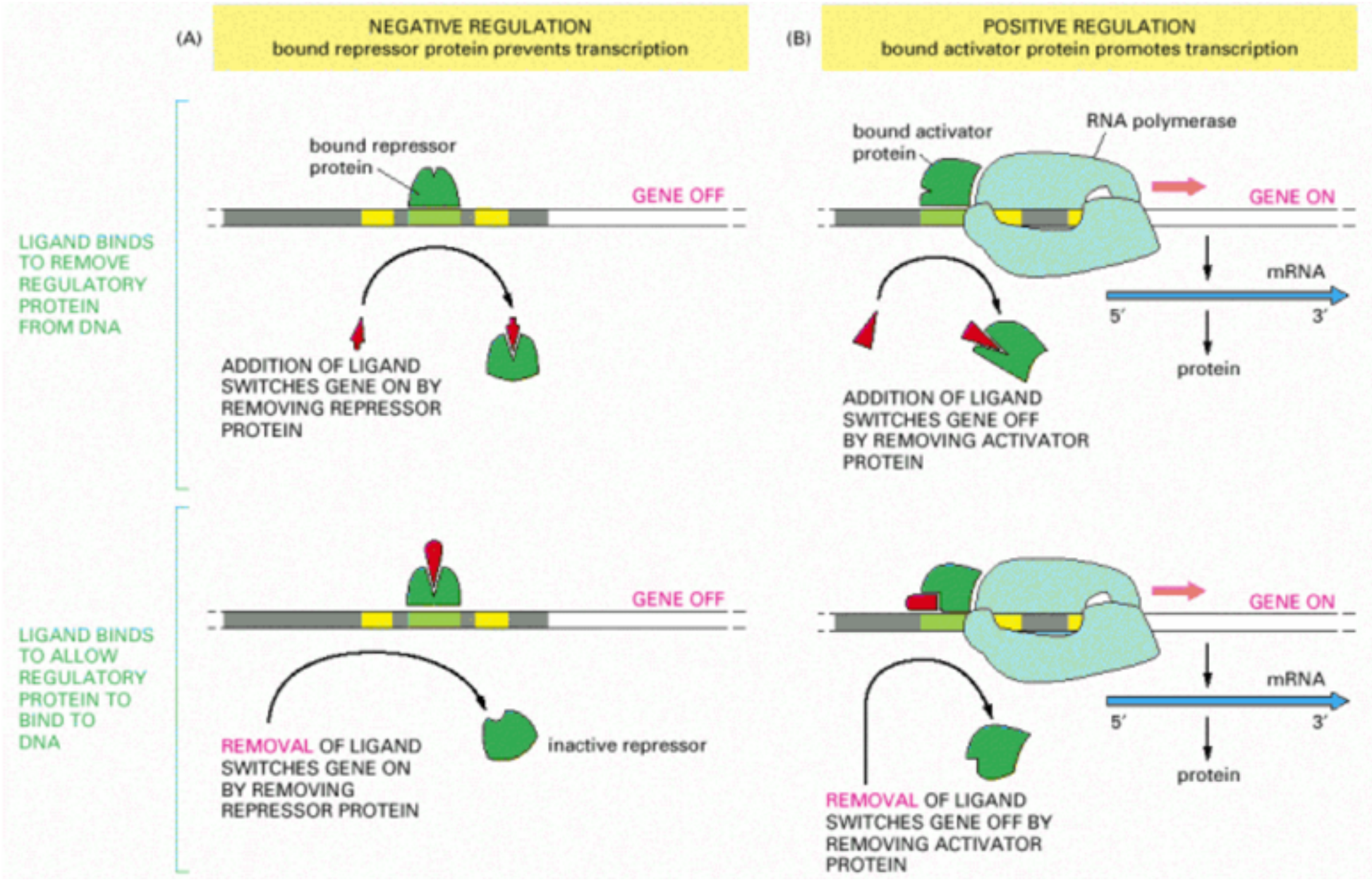
- Find protein binding sites using Gibbs sampling
- Use random walks (Markov chains) which result in sites distributed according to their actual probability of being the correct binding site

- Previously analyzed samples from *Mycobacterium tuberculosis* and *Yersinia pestis*.
- Currently analyzing HG19 regions related to SNAIL and SLUG transcription factors



What is a Binding Site?

Alberts, Johnson, Lewis, Raff, Roberts, & Walter, Molecular Biology of the Cell 4th Edition, 2002



Binding sites are sequences of DNA before a gene that proteins bind to. Different proteins will cause the gene to either ‘turn on’ or ‘turn off’.

Finding Binding Sites

```
...1234567890... … ...1234567890...
GGCCGGTGCTATTACG … GCACGGAGTTATGCGA S. cerevisiae
GGTCGGTGCTATCACG … TCGCGGAGGTATAGGA S. paradoxus
GGCCTGTGTTATTTCG … GCGCGGTGTTATACGA S. mikatae
AACCGGTGTTATTACA … GCGCGGAGTTATAAAG S. kudriavzevii
AGACGGTGTTATGGCA … ACGCGGAGGTATGCGG S. bayanus
```

- Biology is messy -- binding sites are not exact sequences.
- Multiple species with the same genes will have similar binding sites.
- We need to find ‘motifs’ which have the best probability of matching sequences of DNA across species.

Forward Motif Model

		Position											
		1	2	3	4	5	6	7	8	9	10	11	12
Base	A	0.1	0.1	0.2	0.1			0.7	0.6		0.5	0.05	0.9
	C	0.5	0.1	0.2	0.3			0.1	0.2		0.25	0.05	0.05
	G	0.2	0.1	0.3	0.3			0.1	0.1		0.25	0.6	0.05
	T	0.2	0.7	0.3	0.3			0.1	0.1		0.0	0.3	0.0
		Base Probability											

Reverse Motif Model

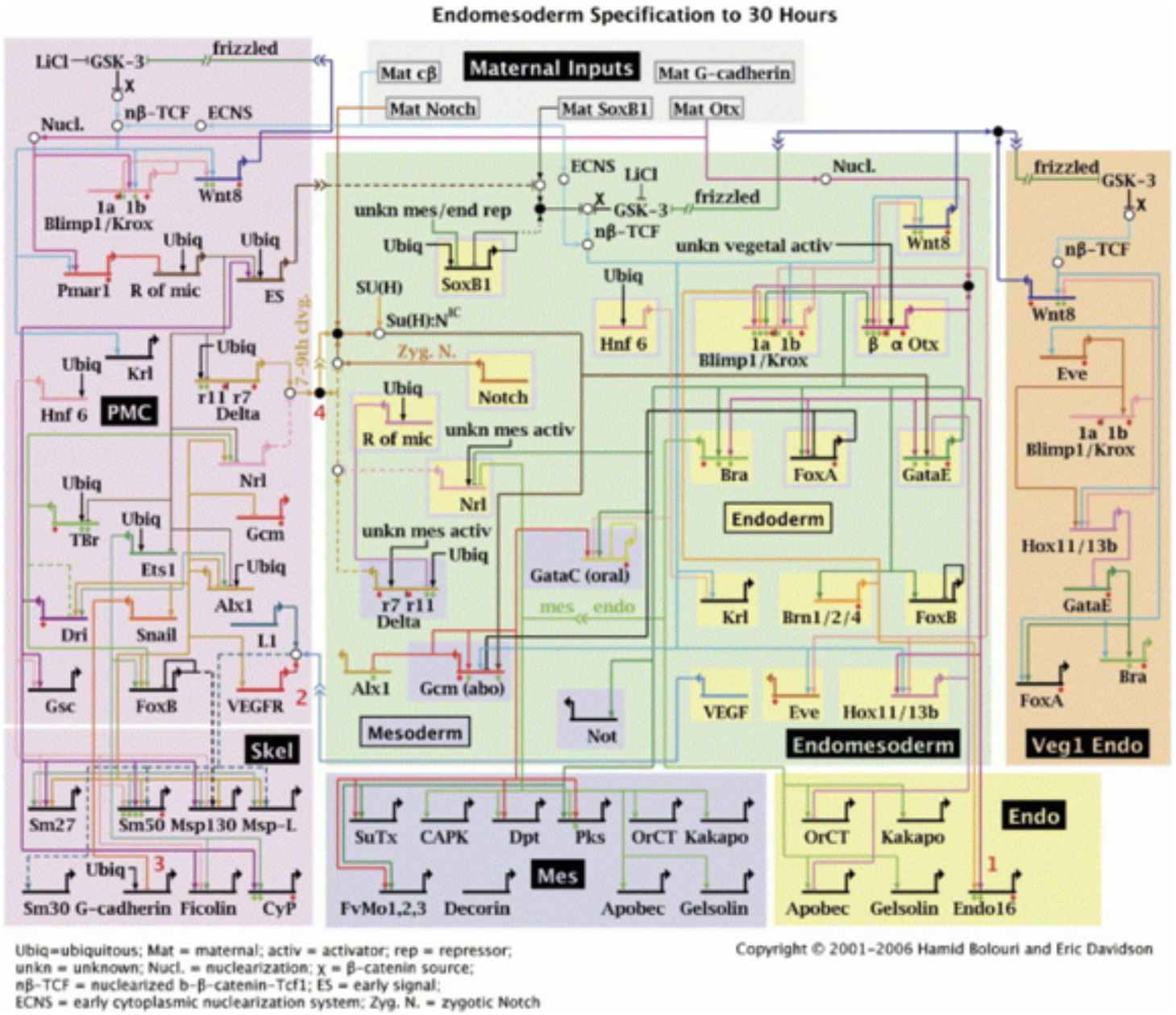
		Position											
		1	2	3	4	5	6	7	8	9	10	11	12
Base	A	0.0	0.3	0.0		0.1	0.1			0.3	0.3	0.7	0.2
	C	0.05	0.6	0.25		0.1	0.1			0.3	0.3	0.1	0.2
	G	0.05	0.05	0.25		0.2	0.1			0.3	0.2	0.1	0.5
	T	0.9	0.05	0.5		0.6	0.7			0.1	0.2	0.1	0.1
		Base Probability											

Palindromic Motif Model

		Position											
		1	2	3	4	5	6	7	8	9	10	11	12
Base	A	0.25	0.85	0.1	0.0					0.25	0.6	0.05	0.3
	C	0.1	0.05	0.2	0.5					0.25	0.1	0.05	0.35
	G	0.35	0.05	0.1	0.25					0.5	0.2	0.05	0.1
	T	0.3	0.05	0.6	0.25					0.0	0.1	0.85	0.25
		Base Probability											

Objective - Regulatory Circuits

Howard-Ashby, Materna, Brown, Tu, Oliveri, Cameron, & Davidson, Dev Biol, 2006

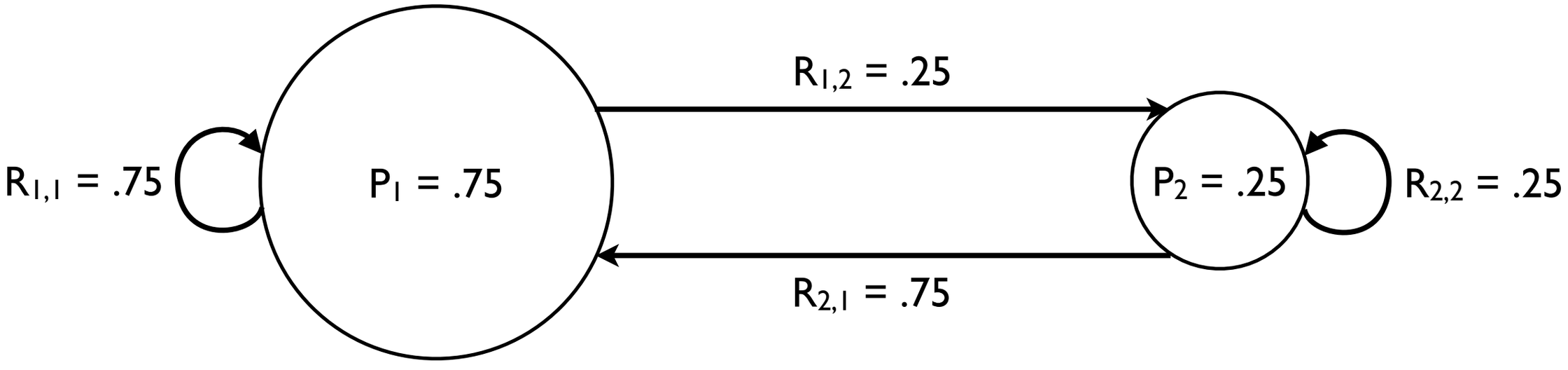


Turning a gene on causes new proteins to be produced, what binding sites will that activate?

Turning a gene off stops production of proteins, which other binding sites will that activate?

Gibbs Sampling

A simple set of states and their transition probabilities.



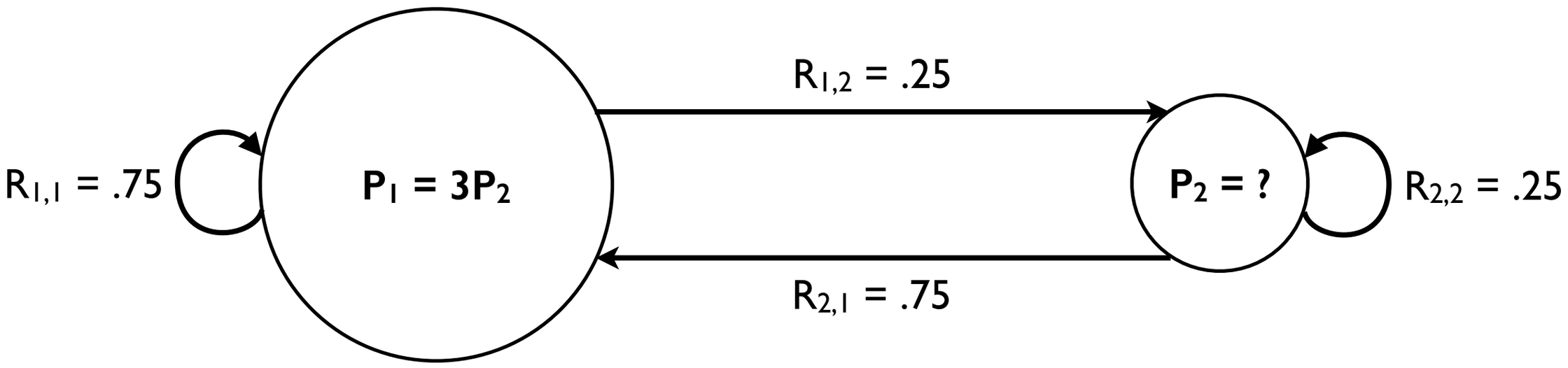
Gibbs sampling is a variant of Markov Chain Monte-Carlo (MCMC) sampling. It performs random walks where each step taken must satisfy a *detailed balance equation*:

$$P_i * R_{i,j} = P_j * R_{j,i}$$

Where P_i is the probability of state *i* being a solution, and P_j is the probability of state j being a solution. $R_{i,j}$ and $R_{j,i}$ are *transition probabilities*, the probability that the state will move from state *i* to state *j* and *j* to *i*, respectively.

To perform Gibbs sampling, it is sufficient to know the relative probabilities of P_i and P_j as it may not be possible to calculate their exact probabilities.

Gibbs Sampling



Given the *detailed balance equation*:

$$P_i * R_{i,j} = P_j * R_{j,i}$$

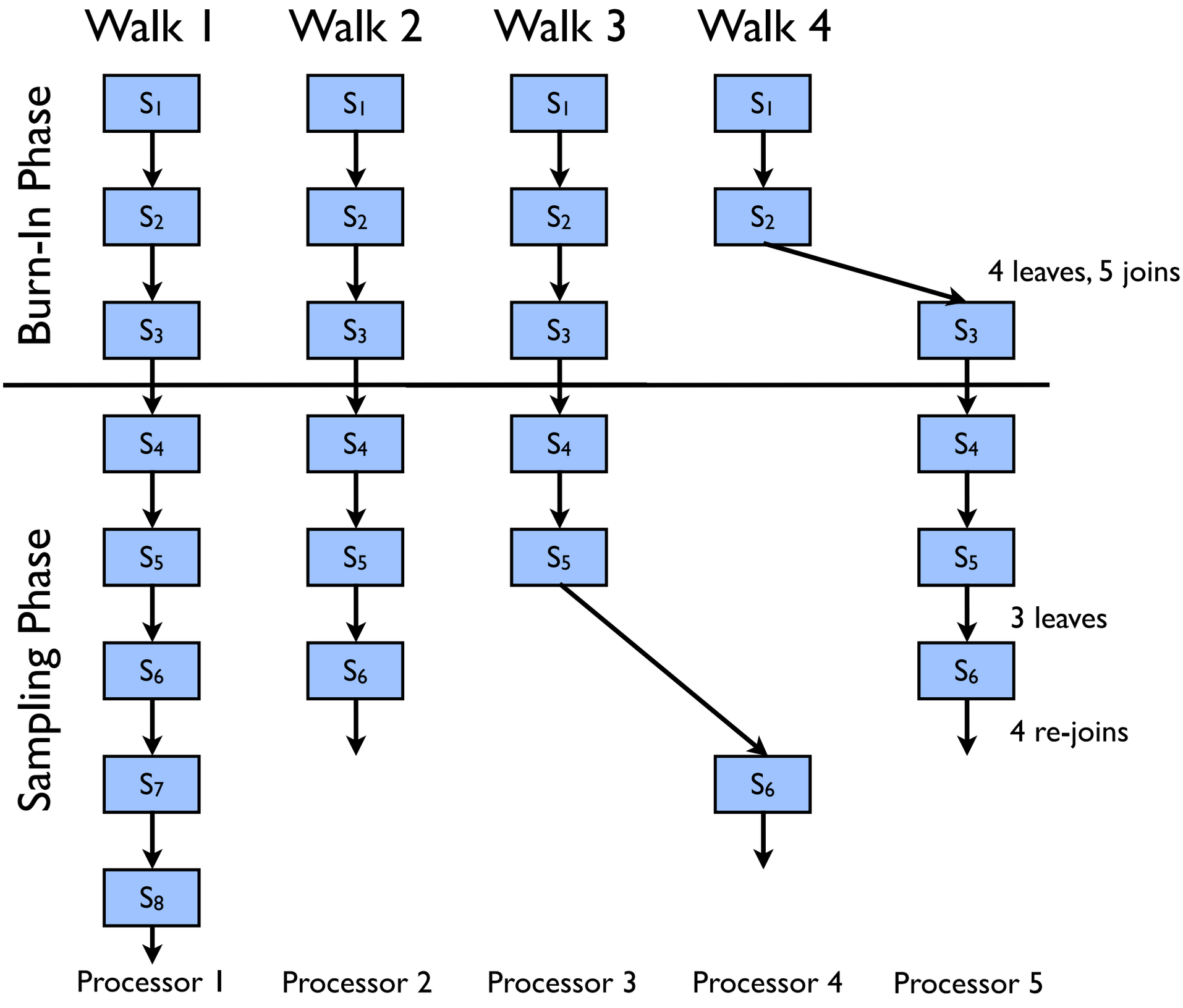
We can determine the same transition probabilities if only the relative probabilities of P_1 and P_2 are known:

$$0.25 * 3 * P_2 = 0.75 * P_2$$

If we perform a long enough random walk between the above states 1 and 2, they will be sampled according to their actual probability distribution: State 1 will be sampled 3 times as much as state 2.

Using gibbs sampling we can find regions of over-represented sequences and calculate their probability of being a transcription factor.

Gibbs Sampling on BOINC



DNA@Home uses parallel Gibbs sampling walks.

Arrows represent *workunits*, or tasks, where hosts receive an initial state with depth *x*, *Sx*, and report a final state with depth *y*, *Sy*.

Workunits have fixed walk lengths (in this case 1). When a walk completes its burn-in period, samples are taken.

Processors can join and leave, restarting from walks of previously left processors.

DNA@Home Results

- A burn-in of 1,000,000 steps and 30,000,000 samples on an average CPU for the *Mycobacterium tuberculosis* data set would take ~2,893 days.
- For 3,000 parallel walks using a burn-in period of 1,000,000 steps, it takes ~7 days for DNA@Home to accumulate 30,000,000 samples -- a ~400x speedup.
- Recent results with HG19, SNAIL and SLUG, gathered using over 2,000 volunteered computers, are currently being processed for publication.

Further Reading

Travis Desell, Lee A. Newberg, Malik Magdon-Ismail, Boleslaw K. Szymanski and William Thompson. Finding Protein Binding Sites Using Volunteer Computing Grids. In the *2011 2nd International Congress on Computer Applications and Computational Science (CACS 2011)*.

Wildlife@Home

Travis Desell & Susan Ellis-Felege
Departments of Computer Science & Biology
University of North Dakota
http://volunteer.cs.und.edu/csg/wildlife

What is Wildlife@Home?

- A *citizen science* project that combines both crowd sourcing and volunteer computing.
- Users volunteer their brain power by observing videos and reporting observations.
- Users volunteer their computer power by downloading videos and performing.
- A scientific web portal to robustly analyze and compare results from users, experts and the computer vision techniques.

Between 2012 and now, Dr. Ellis-Felege has gathered over 100,000 hours of avian nesting video from the following species:

1. Sharp-tailed grouse (*Tympanuchus phasianellus*), an important game bird and wildlife health indicator species.
2. Piping plovers (*Charadrius melodus*), a federally listed threatened species.
3. Interior least terns (*Sternula antillarum*), a federally listed endangered species.

More video is incoming (ducks!), and we have recently received over 2 million motion sensor camera images from a new Hudson Bay project.

Sharp-tailed Grouse

Piping Plover

All three current species are ground nesting birds.

Sharp-tailed grouse nest in the dense grass (top left). Nests were monitored in areas of high oil development, moderate oil development and no oil development (protected state land).

Piping plover and interior least tern are shore nesting species (top right). Nests were monitored along the Missouri River in North Dakota.

What's the point?

1. Current cameras that use automated motion detection miss some predators and are not robust enough).
2. Camera footage allows Dr. Ellis-Felege to manage and evaluate studies with large enough sample sizes for statistical significance.
3. Answer biological questions about parental investment and predator-prey interactions for these ground nesting species.
4. Examine the effect of oil development on wildlife in western North Dakota, which is experiencing a boom in fracking.

Most grouse video is sleeping birds and grass blowing in the wind.
But occasionally, interesting things happen.

Piping plover and tern video is more interesting, with active bi-parental involvement and less obscuring vegetation.

There are many challenges:

1. Dramatically changing weather conditions
2. Dawn/Day/Dusk/Night lighting conditions
3. Model species (sharp tailed grouse and piping plover) and some predators have cryptic coloration (camouflage).
4. Moving vegetation and insects can cause false negatives.

From all this video, we want to determine:

1. Bird Presence
2. Nest Defense
3. Predation Events
4. Nest Success
5. Other events of interest

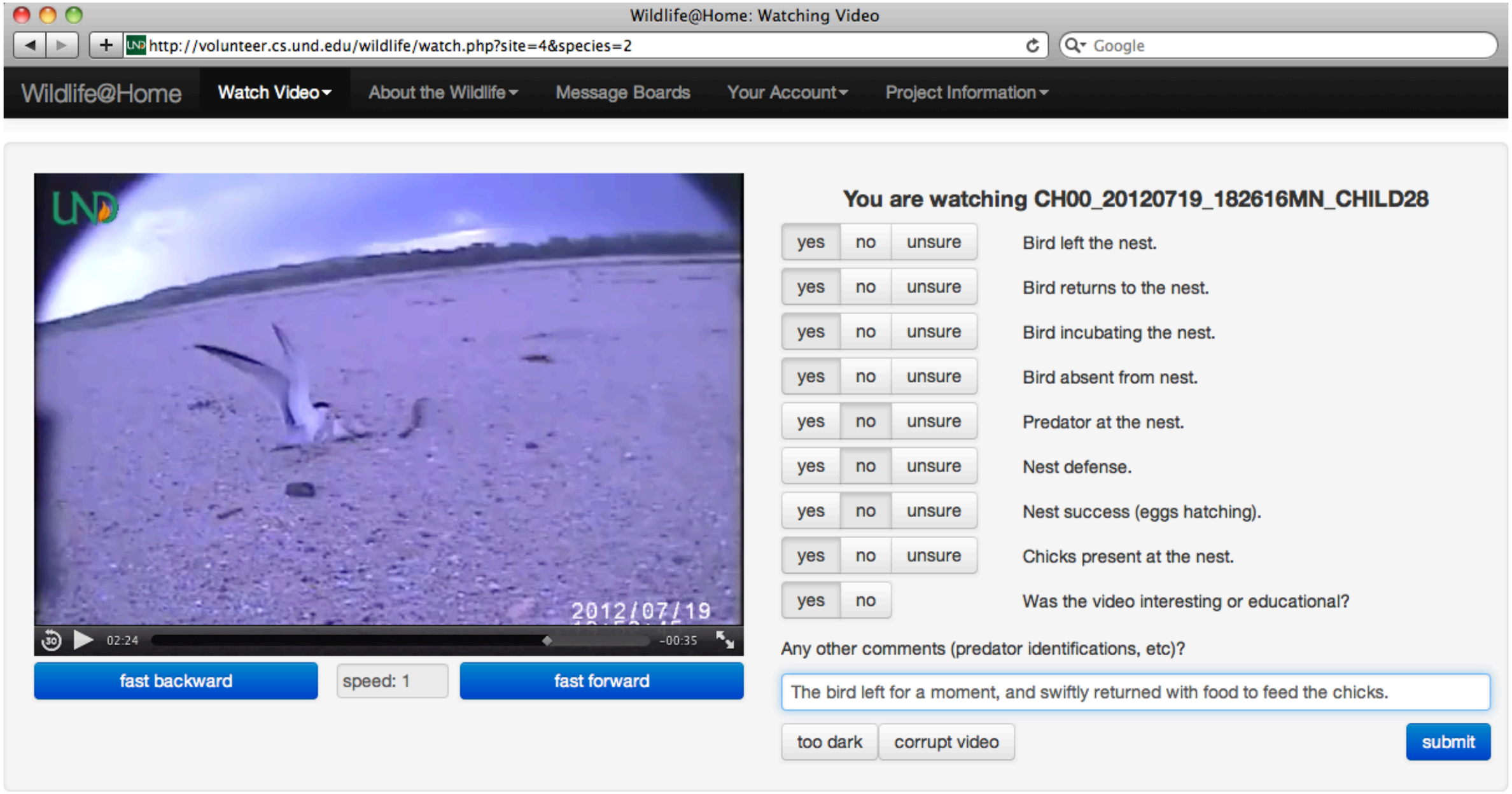
Analyzing all this video requires both a massive amount of computing power as well as a massive amount of brain power.

Computer vision techniques will need to be run, trained and verified, and updated based on human feedback.

Live Demo

A Tale of Two Interfaces

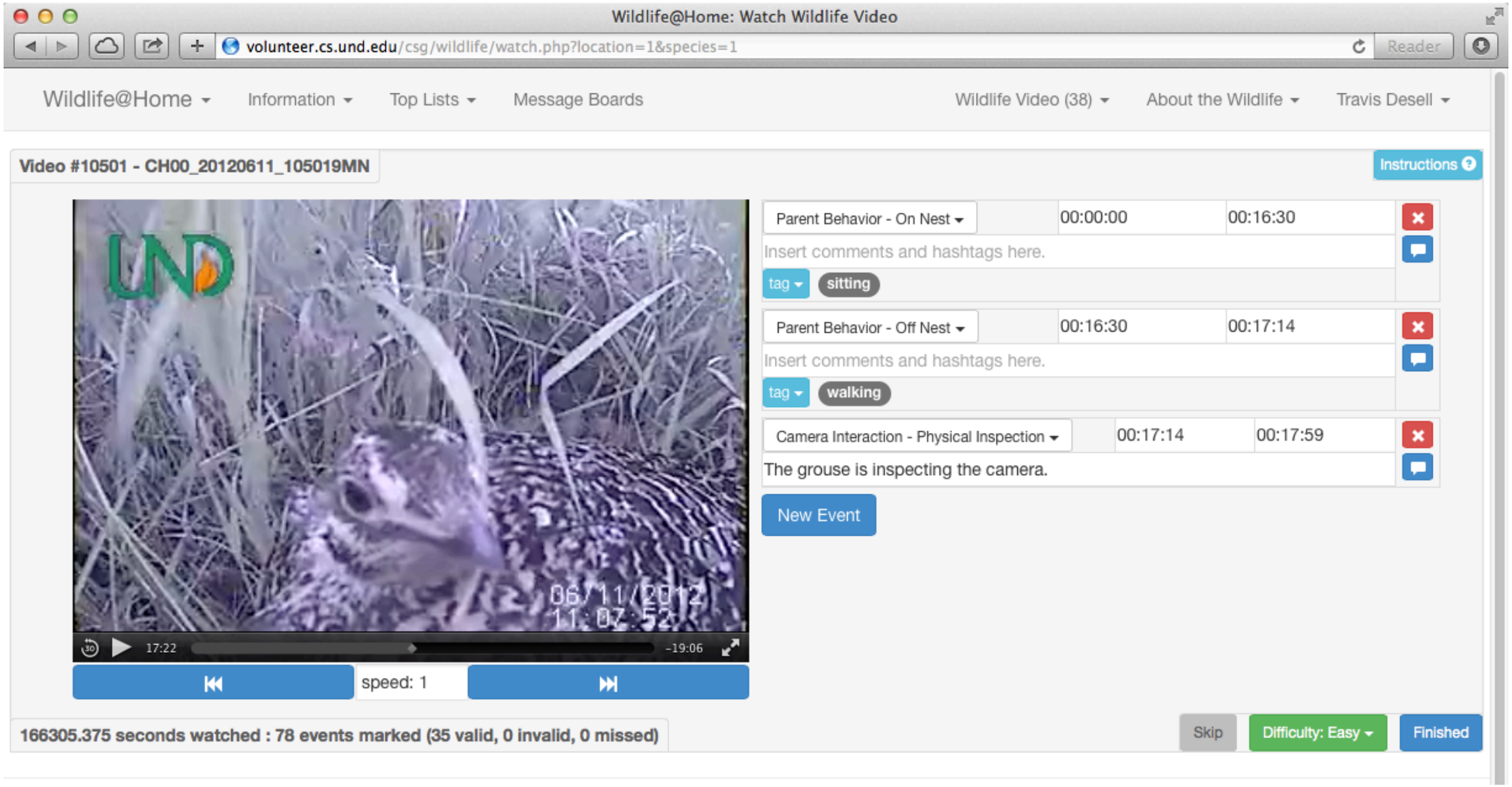
A Tale of Two Interfaces



Originally, Wildlife@Home has a simple interface where users could select yes, no or unsure to specify if an event happened at any time during the video.

As we'll see, this simplicity actually had it's costs.

A Tale of Two Interfaces



The interface is significantly more complex, but allows for very accurate specification of when events occur and also a direct comparison to what Dr. Ellis-Felege's experts report.

A Tale of Two Interfaces

Duration (s)	Completed	Observations	Valid	Invalid	Inconvclusive	Valid (%)
< 180	89,645	220,320	206,193	13,129	618	93.58
181 ... 300	8,942	18,715	17,930	649	75	95.80
301 ... 600	6,446	14,022	12,899	1,033	50	91.99
601 ... 1200	3,785	8,396	7,569	744	55	90.15
Total	108,818	261,453	244,591	15,555	798	93.55

Results gathered over 9 months, from August 2013 to April 2014:

- 206 users provided 261,453 observations for 108,818 video segments (~2.4 views to reach a quorum for a video segment)
- 261,453 observations total over 7,411.2 hours of video watched by volunteers. Only 798 were marked inconclusive, and 15,555 marked invalid.
- In the later months of the original interface, video segments were also generated with durations greater than 3 minutes, due to feedback from the users and an interest in seeing how well volunteers would perform on longer video segments. Additional video segments were generated with 5, 10 and 20 minute durations.

A Tale of Two Interfaces

Event Type	Total	TP	TN	FP	FN	Accuracy (%)
Bird Leave/Return	12501	154	8504	287	3556	69
Bird Presence	21230	9407	1338	9270	1215	51
Bird Absence	9540	1092	4680	2173	1595	61
Predator Presence	414	4	393	11	6	96
Nest Defense	33	0	33	0	0	100
Chick Presence	708	12	418	252	26	61

Of the 108,818 video segments marked by volunteers, 25,549 corresponded to videos that were marked by the projects experts.

- True positives (TP) were when a quorum of volunteers marked an event as occuring a video segment, and the times of the video segment overlapped with the time of a similar expert event.
- False positives (FP) were when the marked event did not overlap with the time of a similar expert event.
- True negatives (TN) were when the event was not marked and an expert did not mark the event during that time.
- False negatives (FN) were when the event was not marked and an expert did mark an event during that time.

A Tale of Two Interfaces

Event Type	Total	TP	TN	FP	FN	Accuracy (%)
Bird Leave/Return	12501	154	8504	287	3556	69
Bird Presence	21230	9407	1338	9270	1215	51
Bird Absence	9540	1092	4680	2173	1595	61
Predator Presence	414	4	393	11	6	96
Nest Defense	33	0	33	0	0	100
Chick Presence	708	12	418	252	26	61

Predator presence and nest defense were very accurate, at 96% and 100%.

Bird Leave/Return were fairly accurate at 69%.

Bird absence was not great at 61%.

Bird presence was especially poor at 51% (essentially random guesses).

There were not enough nest success events for comparison.

A Tale of Two Interfaces

5 second buffer

Event	Misses	Type Mismatch	Matches
Parent Behavior - Not In Video	221 (0.23)	23 (0.02)	708 (0.74)
Chick Behavior - In Video	13 (0.93)	0 (0.00)	1 (0.07)
Territorial - Predator	8 (0.53)	1 (0.07)	6 (0.40)
Territorial - Non-Predator Animal	14 (0.93)	0 (0.00)	1 (0.07)
Camera Interaction - Attack	12 (0.57)	9 (0.43)	0 (0.00)
Camera Interaction - Physical Inspection	22 (0.55)	7 (0.18)	11 (0.28)
Camera Interaction - Observation	9 (0.64)	3 (0.21)	2 (0.14)
Error - Video Error	12 (0.09)	7 (0.05)	120 (0.86)
Error - Camera Issue	12 (0.09)	47 (0.34)	78 (0.57)
Parent Behavior - On Nest	484 (0.11)	152 (0.04)	3686 (0.85)
Parent Behavior - Off Nest	315 (0.31)	16 (0.02)	701 (0.68)

10 second buffer

Event	Misses	Type Mismatch	Matches
Parent Behavior - Not In Video	177 (0.19)	26 (0.03)	749 (0.79)
Chick Behavior - In Video	13 (0.93)	0 (0.00)	1 (0.07)
Territorial - Predator	8 (0.53)	1 (0.07)	6 (0.40)
Territorial - Non-Predator Animal	13 (0.87)	1 (0.07)	1 (0.07)
Camera Interaction - Attack	10 (0.48)	11 (0.52)	0 (0.00)
Camera Interaction - Physical Inspection	12 (0.30)	14 (0.35)	14 (0.35)
Camera Interaction - Observation	7 (0.50)	4 (0.29)	3 (0.21)
Error - Video Error	12 (0.09)	7 (0.05)	120 (0.86)
Error - Camera Issue	12 (0.09)	47 (0.34)	78 (0.57)
Parent Behavior - On Nest	409 (0.09)	168 (0.04)	3745 (0.87)
Parent Behavior - Off Nest	253 (0.25)	29 (0.03)	750 (0.73)

We were able to directly compare user observations from the new interface to the expert observations.

Given a buffer time (events matched if the start and end times were within X seconds of each other), we were able to significantly increase user accuracy.

On nest - 51% to 85-87%

Off nest - 69% to 68-73%

Absence - 61% to 74-79%

A Tale of Two Interfaces

5 second buffer

Event	Misses	Type Mismatch	Matches
Parent Behavior - Not In Video	221 (0.23)	23 (0.02)	708 (0.74)
Chick Behavior - In Video	13 (0.93)	0 (0.00)	1 (0.07)
Territorial - Predator	8 (0.53)	1 (0.07)	6 (0.40)
Territorial - Non-Predator Animal	14 (0.93)	0 (0.00)	1 (0.07)
Camera Interaction - Attack	12 (0.57)	9 (0.43)	0 (0.00)
Camera Interaction - Physical Inspection	22 (0.55)	7 (0.18)	11 (0.28)
Camera Interaction - Observation	9 (0.64)	3 (0.21)	2 (0.14)
Error - Video Error	12 (0.09)	7 (0.05)	120 (0.86)
Error - Camera Issue	12 (0.09)	47 (0.34)	78 (0.57)
Parent Behavior - On Nest	484 (0.11)	152 (0.04)	3686 (0.85)
Parent Behavior - Off Nest	315 (0.31)	16 (0.02)	701 (0.68)

10 second buffer

Event	Misses	Type Mismatch	Matches
Parent Behavior - Not In Video	177 (0.19)	26 (0.03)	749 (0.79)
Chick Behavior - In Video	13 (0.93)	0 (0.00)	1 (0.07)
Territorial - Predator	8 (0.53)	1 (0.07)	6 (0.40)
Territorial - Non-Predator Animal	13 (0.87)	1 (0.07)	1 (0.07)
Camera Interaction - Attack	10 (0.48)	11 (0.52)	0 (0.00)
Camera Interaction - Physical Inspection	12 (0.30)	14 (0.35)	14 (0.35)
Camera Interaction - Observation	7 (0.50)	4 (0.29)	3 (0.21)
Error - Video Error	12 (0.09)	7 (0.05)	120 (0.86)
Error - Camera Issue	12 (0.09)	47 (0.34)	78 (0.57)
Parent Behavior - On Nest	409 (0.09)	168 (0.04)	3745 (0.87)
Parent Behavior - Off Nest	253 (0.25)	29 (0.03)	750 (0.73)

Also, we feel that the numbers would be even more accurate as a recent survey of users found that 38% do not consider themselves fluent in English - which could hamper their understanding of use instructions for the more complicated new interface.

A Tale of Two Interfaces

	Easy	Medium	Hard
Misses	2529 (0.15)	145 (0.14)	90 (0.20)
Type Mismatch	1056 (0.06)	57 (0.05)	24 (0.05)
Matches	13774 (0.79)	863 (0.81)	330 (0.74)

We also provided a way for users to specify how challenging it was to mark events in a video.

Interestingly, those with the highest accuracy had medium difficulty (as opposed to easy).

Travis Desell, Kyle Goehner, Alicia Andes, Rebecca Eckroad, and Susan Ellis-Felege. **On the Effectiveness of Crowd Sourcing Avian Nesting Video Analysis at Wildlife@Home**. *In the 2015 International Conference on Computational Science*. Reykjavík, Iceland. 1-3 June, 2015. **Under Review.**

Computer Vision Methods:

Motion Detection

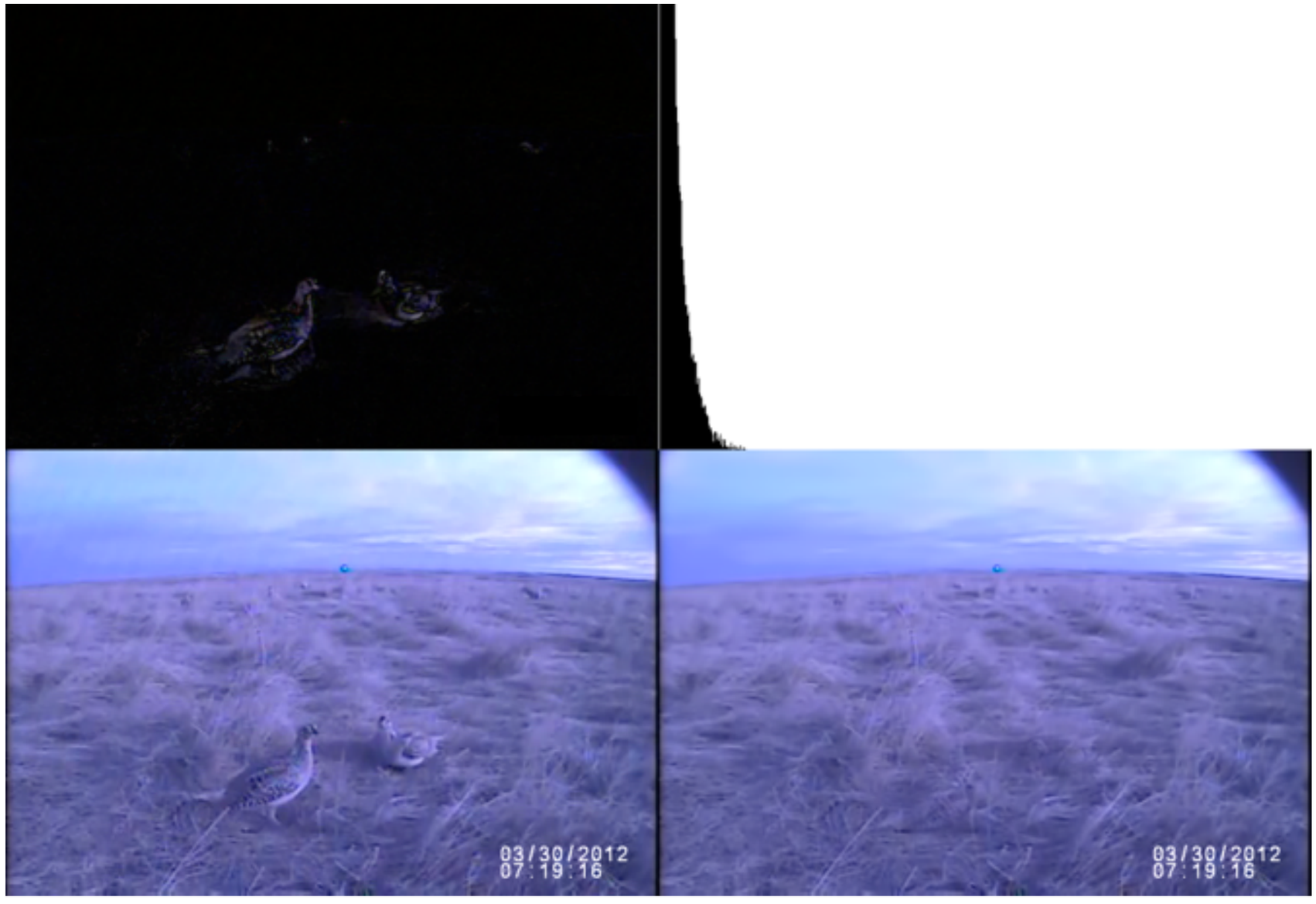
Feature Detection

Background Subtraction

Motion Detection

Initial results gathered using a method called *average window differencing*.

Each frame (lower left) was subtracted from the average of +/- 5 seconds of frames surrounding it (lower right), resulting in a measure of motion (upper left).



Using this, a likelihood of non-noisy motion was for every segment of video.

This was calculated as the average sum of the RGB pixel values in each difference frame divided by the maximum possible difference (3 x width x height x 255).

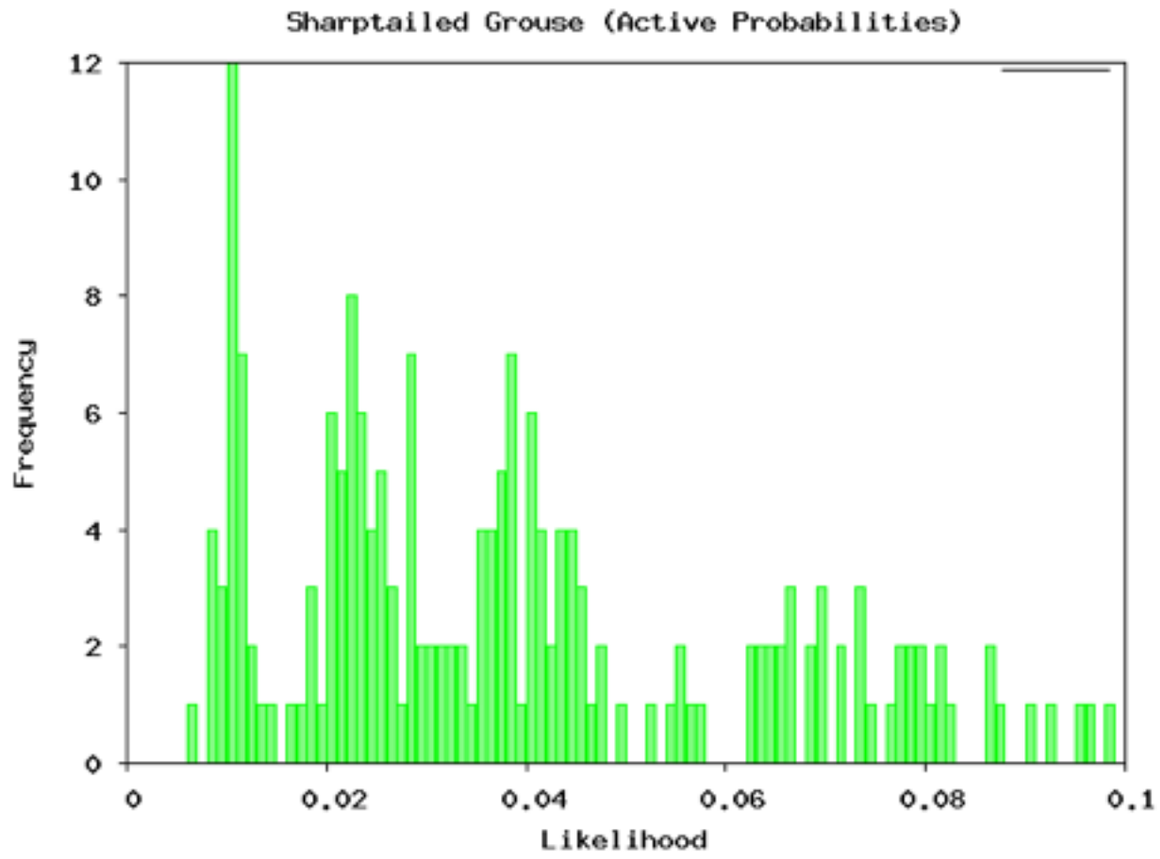
Motion Detection Results

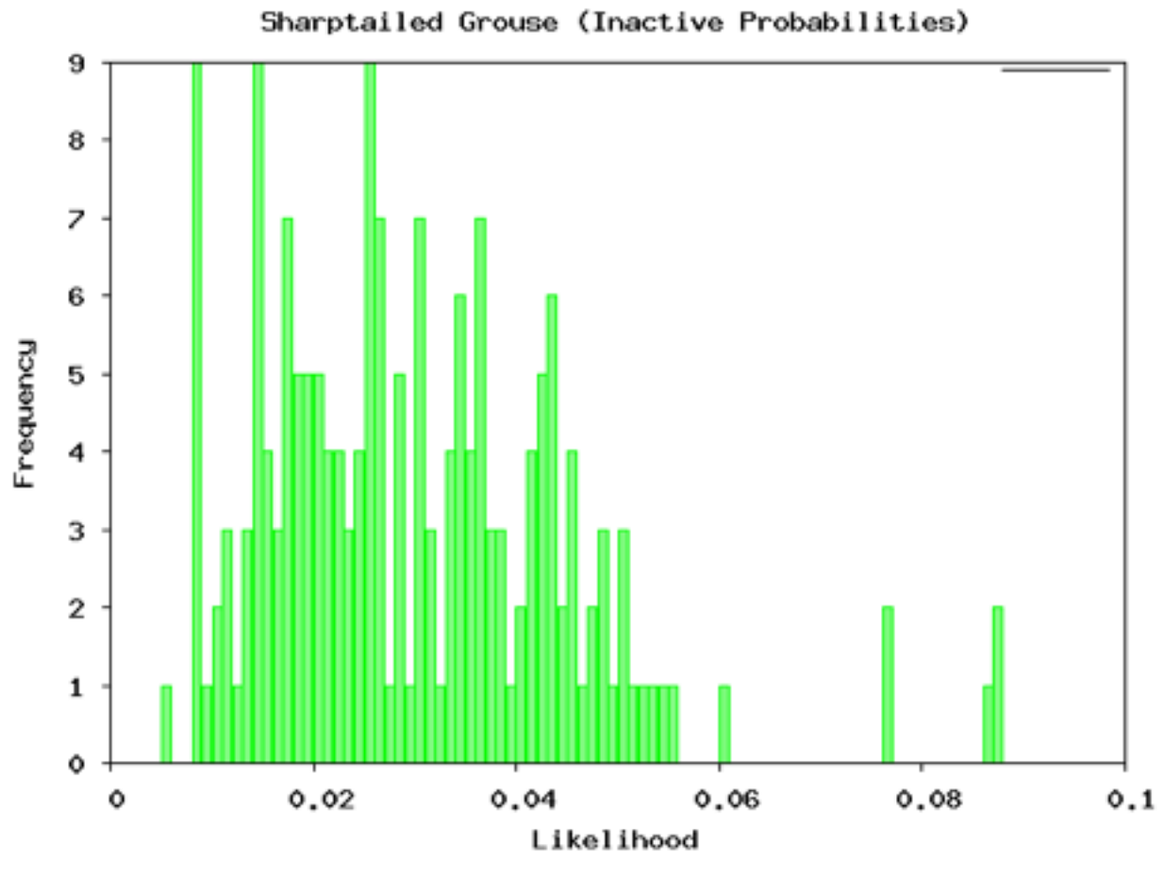
Results for sharp-tailed grouse.

At time of publication:
- 188 videos contained active events (bird return, bird leave, interesting, predator, nest defense)
- 179 contained no active events (bird incubating nest, no bird presence)

Detecting events of interest difficult due to weather, wind and vegetation.

Average and median likelihoods:
- active: 0.039, 0.035
- inactive: 0.030, 0.028





Feature Detection

A feature file was generated by extracting cropped images of birds at their nests in different positions.

Features were extracted using SURF for each image, and then these were merged, by removing any features within a threshold of each other.

This combined feature file was used to calculate a likelihood of a bird being in any segment of video using a bounding rectangle approach.

A rectangle was drawn around all matched features, and the larger the rectangle the less likely there was a strong match to a bird.

Where R_a is the average size of each feature bounding rectangle in each frame of the video segment, and R_f is the size of the frame:

$$likelihood = 1 - R_a / R_f$$

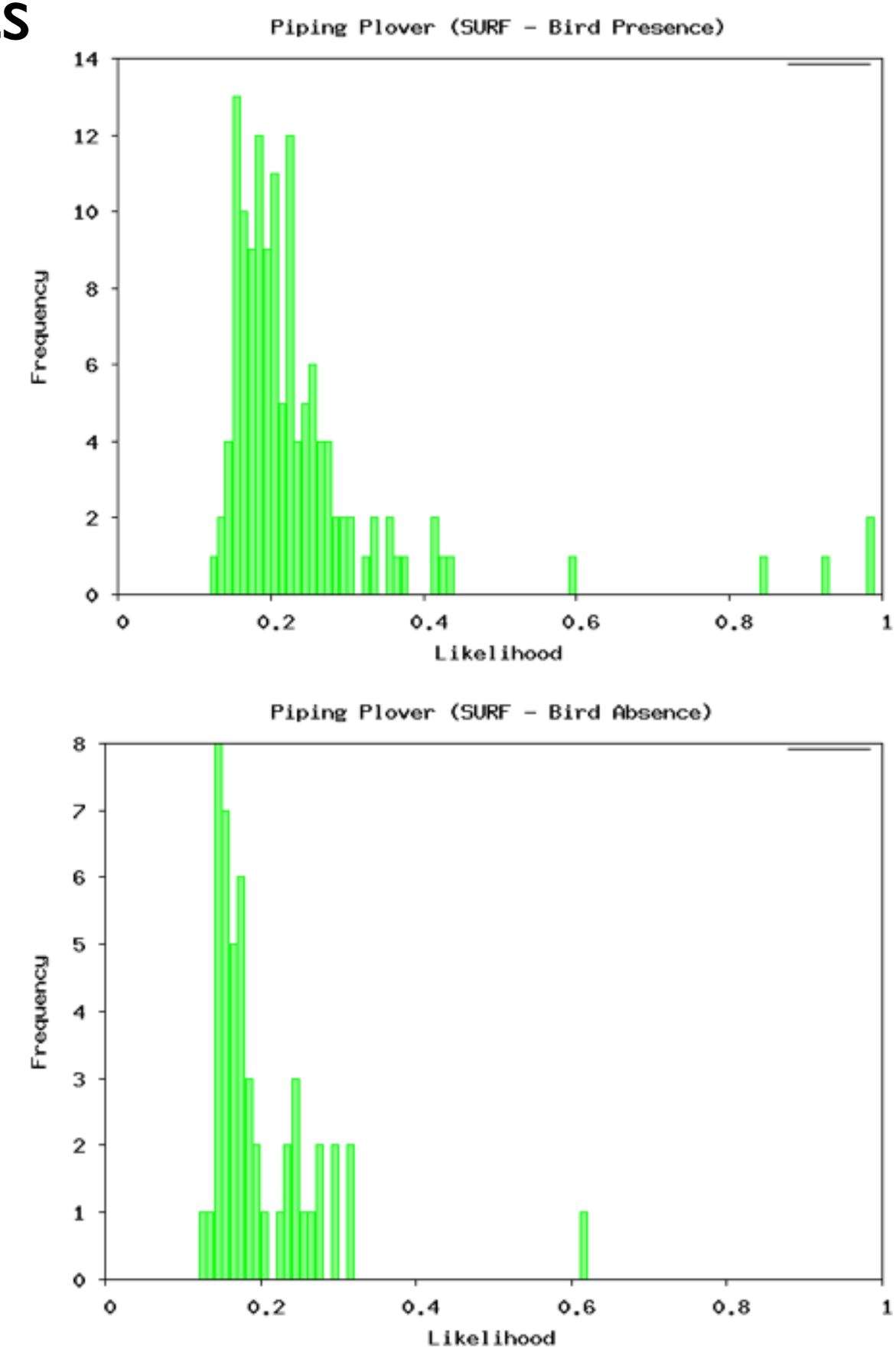
Feature Detection Results

Results for piping plover.

At time of publication:
- 133 videos contained bird presence
- 50 contained bird absence

Note: bi-parental investment means not as many videos without a bird at nest.

Average and median likelihoods:
- presence: 0.24, 0.21
- absence: 0.20, 0.17



Performance Results

At the time of publication, ~70 users had watched over 8400 three minute video segments.
This resulted in ~120 hours of validated observations.

Motion detection was run across the entire video set (~20,000 hours at publication time) and the application processed video at approximately 120 frames per second. At 10 frames per second, this was ~1700 compute hours.

The volunteered hosts processed all videos and returned validated results (meaning each video was analyzed by a volunteer at least twice) in 4-5 days.

Performance Results

SURF feature detection runs much slower (1.7 frames per second).

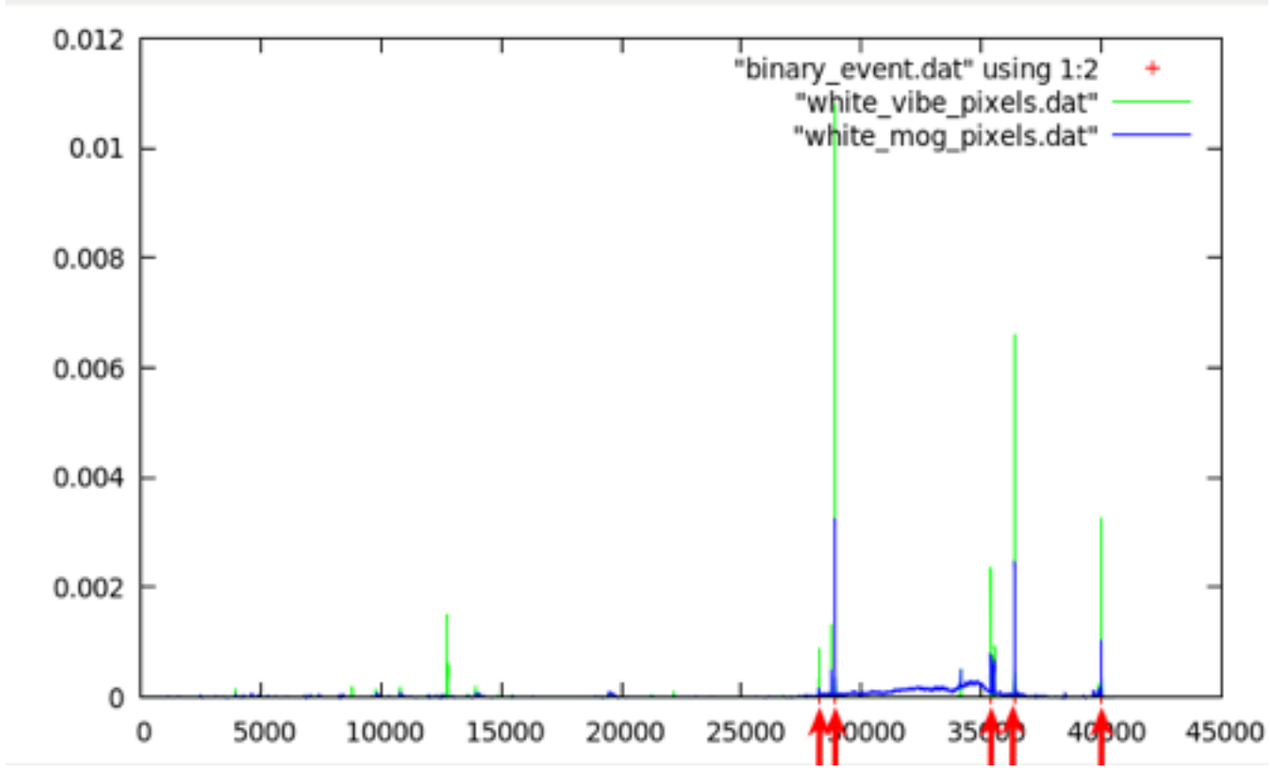
To run this over the piping plover video (682 hours at time of publication), at 10 frames per second or 4000 compute hours results were gathered in under a week.

Travis Desell, Robert Bergman, Kyle Goehner, Ronald Marsh, Rebecca VanderClute, and Susan Ellis-Felege. **Wildlife@Home: Combining Crowd Sourcing and Volunteer Computing to Analyze Avian Nesting Video**. *In the 2013 IEEE 9th International Conference on e-Science*. Beijing, China. October 23-25, 2013.

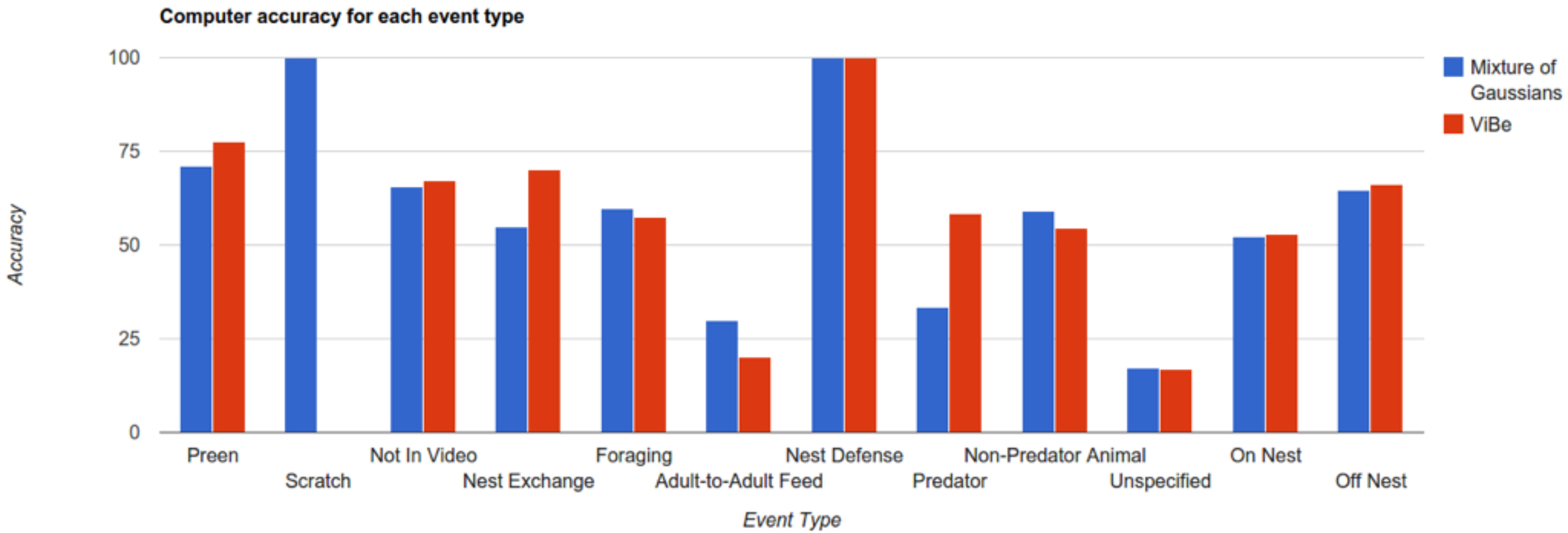
Background Subtraction

Foreground pixels are extracted from an input video file using both the Mixture of Gaussians (MOG) and ViBe algorithms.
Foreground pixels are counted as a percentage of total pixels.
Spikes are classified as an "interesting" event.

- Red arrows indicate scientist classified events (clusters of events).
- Green line indicates pixels marked as foreground with ViBe.
- Blue line indicates pixels marked as foreground with MOG.



Background Subtraction



- Accuracy is determined by the number of expert classified events that have a corresponding algorithm spike.
  - 10 seconds in either direction

- Algorithm accuracy for this video
  - ViBe: 96%
  - MOG: 54%

- Quick lighting changes remain an issue
  - Camera brightness adjustment
  - Overhead shadows created by clouds

What's Next?

What's Next?

Convolutional Neural Networks for animal and event detection on Wildlife@Home.

Analysis of the Hudson Bay imagery.

Aviation@Home - data mining the National General Aviation Flight Database to improve general aviation safety. (Jim Higgins & Brandon Wild, Aviation)

ClimateTweets - crowd sourcing the analysis of tweets involving climate change (Andrei Kirilenko, Earth System Science and Policy).

And I'm always open to new collaborations!

Acknowledgements

Wildlife@Home is currently being supported by NSF award no. 1319700 through the Division of Intelligent Information Systems's Information Integration and Informatics (III) program.

Wildlife@Home has been generously supported by a collaborative research award and new faculty SEED grant from UND's Office of Research Development and Compliance. The project's video streaming server is hosted by UND's Computational Research Center and the volunteer computing server is hosted by UND's Scientific Computing Center. DNA@Home is under partial support from a Basic Sciences SEED Grant.

North Dakota Game and Fish has provided financial support for field logistics to collect sharp-tailed grouse videos.

The US Geological Survey has provided financial support for camera equipment, video storage, and field assistance to collect data for the piping plover and interior least tern.

And of course all our volunteers.

Thanks!

Questions?

http://people.cs.und.edu/~tdesell/

http://volunteer.cs.und.edu

tdesell@cs.und.edu