Which is hosted by the Yale Center on Continent Health.

So today we have a hybrid seminar due to the COVID pandemic,
so we have the students joining us in person,
but also for the students who could not join us,
they can also join us online (indistinct).

But before we move on,
I just want to have two quick kind of housekeeping rules.
So, you guys have submitted questions to our speakers.
So at the end, we will have a Q&A session,
so you guys feel free to ask your question,
raise your hand so the speaker
and for the folks online, if you have any questions,
also please don’t hesitate put them in the chat box.
And we will also go through those questions
on behalf of the attendants.
So, it’s my great pleasure today to introduce
our first speaker of the seminar series,
Dr. Benjamin Zaitchik.
Zaitchik is a Professor in the Department of Earth
and Planetary Sciences at the Johns Hopkins University.
His research addresses hydro-climatic variety,
including fundamental work on atmospheric science
and hydrological processes,
and application to program on water resources,
agriculture and human health.
Dr. Zaitchik is actually also the President of the Two House Session of the American Geophysical Union, in short AGU.
So another thing he want to mention is Ben actually got his PhD from here in 2006, from the Department of Geology and Geophysics.
So we are very pleased to welcome back Ben at Yale, although virtually.
So without further ado, let’s welcome Dr. Benjamin Zaitchik.
Great. Thanks so much Kai.
And thank you for the opportunity to speak.
I've somewhat enjoyed this remote world and our ability to talk and interact at a distance,
but I was a little disappointed when I'm not able to be up there in Newhaven right now,
because it would’ve been fun to come back.
As Kai mentioned, I did do my PhD there, but not in public health, kind of cross on Science Hill.
in geology and geophysics.
But while I was there,
I was not yet working in the geo health area,
particularly between Durland Fish.
and some of his students in public health, my geology department, which really was my first exposure to this idea that you could really make use of some of our environmental information analyses to inform infectious disease analysis. So the talk today, I'm going to be focusing on malaria in the Western Amazon.

I've got a long list of names here, that's only a partial list. I want to particularly acknowledge Bill Pan at Duke University, who has led most of the work I'm going to present on today from the epidemiological side, as well as Mark Janko, Cristina Recalde, and Francisco Pizzitutti, whose results I will be showing. So, might start with some deep background and perhaps an apology in that the idea that malaria somehow in an environmentally mediated disease is not particularly new, right? It shouldn't come as a surprise to anybody. In ancient times, malaria was associated with the rise of Sirius, the dog star, which would come in the days of mid to late summer, around the Mediterranean, where the Greeks and others were studying this and aware of its impact.
That is why we call them the dog days of summer, because that’s when Sirius became visible. And you can see writings about this across Mediterranean at the time. Hippocrates, who was famously very interested in the relationship between environment and meteorology wrote specifically about how these cyclical fevers that we now understand to be malaria were associated with the season, and clearly understood quite clearly that this was not an astrological phenomenon, but that this was a phenomenon tied to the seasonality. Now, this is millennia before the mosquito-mediated pathway of malaria transmission was confirmed, as well as before the plasmodium was identified, certainly as the parasite. And yet this understanding that malaria was sensitive to these changes was clear. I mean, the very fact we call it malaria, right? Bad air. It’s the disease that is most associated inherently in our naming system with this idea of environmental sensitivity. And so, you might think that we had this kind of figured out, right? So why in the year 2021,
I am here to talk to you about our attempts and our struggles to continue to understand in a predictive fashion, the way in which malaria responds to environmental variability? And I think the answer is that it’s a bit complicated. And so what I’m going to talk about here is something where we really need to understand the environmental influence, and the climatic influence as well as other environmental influences, in the full context of a coupled natural human system that evolves with time. And so, simply understand that malaria has the potential to be sensitive to environmental factors is not in and of itself a useful or actionable predictive system. So the talk today, I’m going to start off with some background on malaria in the Western Amazon, and apology in advance if doing so is insulting to folks in public health who have a deep understanding of malaria in this region, but I’m not sure of everyone’s background. So we’ll go through a little bit of that history and current dynamics. Then, I’m going to spend a little bit more time.
than you probably want me to on physical geography and hydrometeorology because that’s really what I bring to these set of analyses. Then I’ll move on and just give three of the cases in which you’ve tried to integrate these kinds of environmental information systems to our understanding and forecast of malaria in this region. And I want to emphasize something that Kai said, was certainly type into the chat if you would like to say anything. Also feel free just to unmute and interrupt if I say something that is unclear. So again, based back on the malaria in the Amazon, this is from the malaria Atlas. And what we see here is that the dominant type of malaria will be vivax that is present throughout the Amazon basin, but you also see falciparum in some concentration, and the Western Amazon part of Peru and Western Brazil, and then we focusing on, you will see both in significant amounts. I should note that I’m zoomed in here on the Amazon basin. The Amazon is home to over 90% of malaria in the Western hemisphere.
And so it’s really in terms of studying the Americas, it’s the place that one would want to be focusing a lot of effort on malaria reduction. And in this region, malaria is classically associated with deforestation, encroachment into the natural forest. So, it’s just a satellite time-lapse, showing over about 30 years what we all know to be true, this massive deforestation. This particular lapse rate is from Brazil. You see similar things throughout the Amazon basin. Classic pattern here is a road gets built, you surmise from that flash across the screen at the beginning of the time series. Once the road is built, you get this herringbone pattern of deforestation as land is cleared for logging, but then also for agriculture and ranching. And this dynamic was associated with a massive burst of malaria in the Amazon region, particularly in the ’80s and 1990s. And so that was really the time of the most rapid deforestation going on over much of the Amazon. Continues to be a major issue today, but that’s when the rate was the highest. And what you had there was a situation
where epidemiologically naive populations were entering into a region where the anopheline mosquitoes, the dominant vector of malaria, were present in large numbers, and the kinds of livelihoods we were seeing in particular, this kind of entering the wilderness for logging and such, and then a lot of mobility going on all led to this really strong epidemic peak. And from observing the dynamics, this what now we would consider to be a classic hypothesis emerged called the malaria frontier. And so you have frontier malaria in situations where you have populations that do not have immunity, and who do not have behavioral patterns associated with trying to avoid malaria, because they’re new to the area, enter into the wilderness frontier, and you get this burst of epidemic peaks, followed by a gradual adjustment as you get some resistance building up, as you get populations’ behavior changing, and as you get livelihood changes that maybe are a little less mobile and include less interface with wildlands, and you settle into an endemic pattern, this endemic malaria.
And so, you know, this has flashed through enough times that maybe you’ve noticed by now that you can kind of see the timing, right?

While things change throughout the time series I’m showing, after about the year 2000 or so, the change isn’t as rampant. You don’t see as much clear cutting, right? That mostly happened in the ’80s and ’90s. Again, this is a time series from Brazil. You’d see similar things in the parts of Peru and Ecuador that we’re focusing on.

So, when I talk about malaria today, I’m going to be focusing on the last 20 years, which is really coast frontier malaria.

Okay, so this is the time where we say, okay, we’ve kind of been through that initial burst of malaria that happens when you enter the frontier. And now, we’re in the situation where we are looking at transmission patterns in populations that I wouldn’t say that it’s a stable population. There’s always movement going on. But you’re no longer talking about this encroachment.

You’re talking about interfaces within what is more or less a settled area. Okay, and so what does that look like
232 00:10:45.760 --> 00:10:48.390 if you just look at case numbers in the Amazon?
233 00:10:48.390 --> 00:10:51.390 So here, I’m showing a time series from 2000 on.
234 00:10:51.390 --> 00:10:53.200 And so what you’re listing over to the left here
235 00:10:53.200 --> 00:10:55.420 are there really high numbers that preceded this?
236 00:10:55.420 --> 00:10:57.660 So the numbers on this curve, you can kind of see Brazil,
237 00:10:57.660 --> 00:10:59.710 that red curve coming down, right,
238 00:10:59.710 --> 00:11:02.127 from what was a really big peak in the 1990s.
239 00:11:03.150 --> 00:11:05.300 And if you ignore Venezuela,
240 00:11:05.300 --> 00:11:08.010 which as we all know has had its own challenges,
241 00:11:08.010 --> 00:11:08.843 you would generally say,
242 00:11:08.843 --> 00:11:12.220 "Oh, this is kind of a story of cases falling, okay,
243 00:11:12.220 --> 00:11:15.569 from that frontier malaria peak."
244 00:11:15.569 --> 00:11:18.540 But if you look a little more closely,
245 00:11:18.540 --> 00:11:21.130 over the last 20 years, you’ll see that progress
246 00:11:21.130 --> 00:11:23.180 has stalled and even reversed.
247 00:11:23.180 --> 00:11:25.520 And so expanding the Y axes a little bit here
248 00:11:25.520 --> 00:11:28.380 to look at Columbia, Ecuador and Peru,
249 00:11:28.380 --> 00:11:32.030 just over the past 15 years or so,
250 00:11:32.030 --> 00:11:35.110 what you see is a rather significant peak in Ecuador.
251 00:11:35.110 --> 00:11:36.177 It came down a little bit after that
252 00:11:36.177 --> 00:11:37.670 but it’s come back up.
253 00:11:37.670 --> 00:11:40.270 Peru, quite a significant percent wise increase,
254 00:11:40.270 --> 00:11:44.793 because the case has got so low in the the early 2010s.
255 00:11:47.710 --> 00:11:48.710 Sorry, that was Ecuador.
256 00:11:48.710 --> 00:11:51.050 Big, significant increase in Ecuador.
I missed my labels here. Then bottom one is Peru showing the significant increase, again. And so you see these large percent wise increase in these Western Amazonian countries. Focusing on Peru specifically for a moment, because that’s what a bunch of our data are going to come from, that I’m going to show in the next section. What you see here is a phenomenon where, again, cases were quite high in the 1990s, but there seemed to be a period where you were at a kind of a stable level in the 2000s, and then a rapid decline to the point where it was really getting close to elimination around 2010, before it burst back up. And so now what’s been happening? So that period, as I’ll get to it towards the end of the talk, was a period of a significant intervention. So the PAMAFRO program, which ran for about five years involved a number of malaria control activities. Again, details come later, and it really did seem to work. Then in 2011, you had this historical flood.
There was a flood of record over much of the Amazon, the biggest one in the observed record. And it had tremendous impacts across the region. But one thing that happened was what we saw an increase in malaria cases, this reversal, okay? Now this flood coincided with the end of the PAMAFRO program. And so we have some disentangling to do, about what’s going on when it increased. And when this first happened, there was a sense of like, "Okay, a flood happened, there’s going to be a bunch of malaria, and it’ll come back down,” But didn’t. Just kept going up and up and up. In the time since that flood, you’ve had several other destabilizing events. as you might be aware, was this mega El Nino, with global effects. You also had dengue and Zika, and particularly with the Zika scare coming through this region at that time, which really stressed health systems. And so, one thing that we’re trying to do now is say, "Okay, in this context of intermingled climatic effects,
308 00:13:49.310 --> 00:13:51.590 social effects, epidemiological effects,
309 00:13:51.590 --> 00:13:54.140 what exactly is going on here?”
310 00:13:54.140 --> 00:13:56.510 And this is critical, because, you know, 10
years ago,
311 00:13:56.510 --> 00:13:57.960 if I were giving this talk 10 years ago,
312 00:13:57.960 --> 00:14:00.900 we’d be talking about elimination of malaria
in the Amazon.
313 00:14:00.900 --> 00:14:02.150 We are not talking about that right now.
314 00:14:02.150 --> 00:14:03.390 We’re talking about trying to control
315 00:14:03.390 --> 00:14:05.480 what seems to be an increase...
316 00:14:05.480 --> 00:14:06.798 Though you don’t see it on this graph,
317 00:14:06.798 --> 00:14:09.430 because Peru seems to settle down a bit,
318 00:14:09.430 --> 00:14:10.571 not just an increase, but really,
319 00:14:10.571 --> 00:14:14.740 maybe a significant continuing increase of
malaria
320 00:14:14.740 --> 00:14:15.573 in the region.
321 00:14:17.800 --> 00:14:21.830 Okay, so let me jump into the physical geog-
raphy
322 00:14:21.830 --> 00:14:23.973 and hydrometeorology of the problem.
323 00:14:27.140 --> 00:14:28.970 So, let me start off with a little bit about the
vectors.
324 00:14:28.970 --> 00:14:33.270 So as I will attempt to stress throughout this
talk,
325 00:14:33.270 --> 00:14:34.640 when we talk about the influence
326 00:14:34.640 --> 00:14:36.190 of environment and hydrometeorology,
327 00:14:36.190 --> 00:14:39.940 we’re not just talking about mosquitoes, okay?
328 00:14:39.940 --> 00:14:41.550 Mosquitoes are a big part of it.
329 00:14:41.550 --> 00:14:42.880 So, that’s why I start off with them,
330 00:14:42.880 --> 00:14:45.233 but we always want to be thinking about
mosquitoes.
331 00:14:45.233 --> 00:14:46.899 You want to talk about the pathogen,
332 00:14:46.899 --> 00:14:49.373 and we also want to talk about human behav-
ior.
Nevertheless, the influence of land cover in hydrometeorology in particular on anophleles mosquitoes is going to be a big part of our story, so I want to make sure you’re familiar with what’s going on in the Amazon. So, the red here is showing anopheles darlingi. That is the dominant malaria competent vector in the Amazon. There are a whole bunch of others, a great diversity of anopheles mosquitoes here, but the darlingi is going to be the number one.

If we zoom in a little bit, just a little box there, around this portion of the Western Amazon, centered on the Laredo district of Peru, which is kind of the Northern Amazonian district in Peru, you can go and study this there, because a lot of really good work has been done by some of the members of the team that were on my title slide, and people who preceded them or partnered with them in this area doing really strong work on mosquito surveys, or collecting or doing species typing. And this happened along various areas in the region. And I don’t know how well this is showing up on your screen,
but that red inset there is a Landsat satellite snapshot of the area. And you might see red dots, yellow dots, green dots. Those are all showing collection sites where breeding habitats and mosquito species types were collected at larval and adult stages. And they were organized along transportation corridors, these surveys. And so the red dots are along a highway that connects Iquitos to Nauta, a town to the south. The yellow dots connect Iquitos to Mozan up in the north. And then the green dots are going along various rivers that are used as transportation corridors. Let me just zoom in on that a little bit, so you get a sense of this region. So here, this is just kind of a true color satellite image of what I showed in the previous slides. You see the Amazon river flowing south to north here through the region. That urbanized area that you see along the banks of this meander is Iquitos. Iquitos is famously the largest city in the world that you can not reach by road. You either have to come in on the river.
or you have to fly in.

The rivers are the dominant transportation networks,

but we have these roads I showed before.

There’s one to the north that kind of cuts off here, going to Mozan,

but this highway here, the Iquitos to Nauta highway

is kind of the biggest road in the area.

And you see that herringbone deforestation coming along that road.

And so, what we have here are mosquito collections

in an area of land use contrasts,

including the pristine forest and breeding into areas of significant agricultural activity

and urban activity.

And so, we can then use our satellite images to classify the different types of cover we see here,

and these range from different water types.

We always want distinguish between clear water and silky water in the Amazon.

They’re very different ecologies.

And then different kinds of Amazon basin land cover type,

including the anthropic types,

such as disturbed vegetation and bare ground,

and roads and buildings,

and the natural vegetation types,

including different types of forest.
And so when we analyze these together, the land cover information with the mosquito information, you find some interesting patterns. And what I have here are all anopheles species. Okay, I didn’t bother spelling out all of the species names, because they’re long and it doesn’t matter too much. But what this box plot is intended to demonstrate is that, as your forest area decreases, okay, as you go down on the Y axis into the negative area here, you will see decrease. You will see different relationships with different species.

The Y axis here is the association. Okay? And so you see negative associations between forest area and some species, and positive associations between forest area and other species. Okay.

And so, what’s interesting about this is that you say, "Okay, there’s going to be changing species assemblages, as land cover shifts from natural forest to more cleared area." But it’s somewhat systematic,
are the malaria competent species.

You’ll see anopheles darlingi here on the far left.

And so, that’s a dominant vector and all of these others

are vectors, also.

These are not, okay?

So it so happens that as you clear forest,

you might not actually see an increase

in the total number of anopheles mosquitoes.

You often will see a decrease in the total number

of mosquitoes of all species,

but you’ll see an increase in the prevalence

and absolute number of darlingi, of your vector species.

And in fact, it’s even quantified.

Here’s some data we had.

We found that for every 1% increase in clear land area,

you have close to a 4% increase in the odds

of finding anopheles darlingi at a collection site.

So we have here is human wildlife interface

causing more mosquito human interactions.

And also, the anthropic disturbances of the landscape

increasing the proportion of your competent vectors.

So this is a recipe for increased malaria transmission.

So this is a fairly detailed study

that we could only do in places where we had

really detailed entomological collections.
We don’t have that everywhere, but at least from the satellite perspective, we can take this kind of last and done at high resolution and zoom out of it. And so as we try to look across all of the Laredo states, this shows Laredo state of Peru, and this analysis has now been extended to include the Amazonian portions of Ecuador, as well as parts of Colombia and Brazil. We can make use of satellite data. And here I’m showing the MODIS satellite data. If you’re not familiar with MODIS, it’s a NASA-supported mission has been up for about 20 years now. And unlike the previous images that I showed, which is a Landsat higher resolution, 30 meter resolution, but you only get snapshots every once in awhile, MODIS is giving you 250 to 500 meter resolution, but it’s giving you daily images. And these really cloudy areas that’s important, right? So you need to catch when you can a view through the clouds. And that allows us to use phenology. That is the seasonality of the vegetation to do a more detailed classification of land cover types.
And it says on the left, just a classification using MODIS. We can then, because the satellite’s been up for 20 years, look at change in these forest types over time. All of that can go into our malaria risk analyses. And on the right, what I’m showing you is a card that I did not develop, that NatureServe developed, which used a combination of satellite data and measurements on the ground to come up with ecological systems, that we view as potentially relevant to malaria. In particular, the red areas on this map are areas that are forested, that are flooded by what they called black water. So those tannic waters of the Amazon. And then in the light green, you’ll see other areas that are flooded by what they’re calling white or clear water. Might have sediment in it, but it’s not tannic, okay? So again, different water quality, different ecology.

And so, what I’ve taken here is land use, look at really high resolution land use, to understand the scale of distribution. Used a different satellite assets in order to zoom out and say, "What can we say at scale about land use
and vegetation types?"

And also, thanks to the NatureServe analysis, link that somehow to hydrology, right?

Because now we’re talking about ecological zones that are defined, in part, by their flooding regime,

which is a key consideration in the Amazon, right?

There’s a lot of forest that’s different from other forests,

and much of that has to do with these flooding regimes.

So this brings hydrometeorology into the picture, right?

And so, how does hydrometeorology matter?

As I mentioned, it’s going to affect the vector, right?

We’re concerned about breeding sites,

survivability of different life stages,

the life cycle, speed of the life cycle of the mosquito,

dispersion of mosquitoes,

influenced by winds and temperature.

And so, wind, temperature and certainly precipitation

and moisture conditions in the soil and surface puddles

are going to be a big deal.

We also know the plasmodium has temperature sensitivities,

and that the vector’s competence transmit the plasmodium

is a function of temperature.
On top of that, you’ve got human behavior. And so migratory labor in particular, logging in this area is very sensitive to the river height, because you need the rivers to be a certain height in order to float the logs downstream. And so will have an influence. And then of course, agricultural activities will be sensitive to the seasonality of hydrometeorology, as well as the inter-annual variability. When you get interventions, you also have an issue that anyone who’s worked in malaria knows, which is, "Will people use bed nets?" And when it gets really hot, very often, it gets harder to comfortably use a bed net. So, how are we going to do hydrometeorology? So there are a lot of different ways you can do this. The system that my group uses, and kind of one of our major contributions to this malaria problem is called the land data assimilation system. So that probably doesn’t get discussed too much at schools of public health, which is appropriate. Let me give you a little background, because this is an area where any of you potentially working on various climate environment
influence on disease,
but really any host of public health issues
might be able to make use of such a system,
collaboratively or on your own,
to really bring environmental data in, in a
powerful way.
So what an LDAS does is it merges observa-
tions
with numerical models,
in order to get your best possible estimates
of what’s going on with the land surface
and the lower atmosphere than your surface meteorology.
Why do you do this?
You do this because satellite observations
are amazingly powerful tools, but they’re snapshots
of single variables.
And so, if you want a comprehensive view
of what’s happening with all the potential
variables of interest, you kind of want a model, right?
You want something to give you spatially
and temporally complete and consistent rep-
resentation.
But those models don’t necessarily represent reality,
particularly in data limited environments, like the Amazon.
And so what you do with an LDAS is you basically
pick at the best of both worlds to the extent possible.
You have an advanced, physically based model
that is trying to simulate what’s going on with your weather and with your hydrology. And then you’ve got satellite observations that inform that model and kind of keep it realistic.

And so, in schematic form, what you have is a bunch of landscape information, such as the land cover analyses I’ve just shown you, often satellite-derived. You have meteorological data, which is also often from satellites, or from other weather analysis systems. Those all drive a numerical model, which is then going to produce estimates of energy balance and hydrology, okay? So that’ll get you, you know, the temperature, radiation, wind, moisture conditions you care about.

As you run this model forward, you assimilate observations. And so you can update observations. So for example, information about soil moisture variability.

Graded estimates come from satellite can be brought into the numerical model to update the model’s estimate of soil moisture. And so, you end up with a system. This should be obvious, because we’re using updated observations.
This isn’t like a future projection model, right?
The model itself might be able to, but the LDAS system is retrospective, up through real-time monitoring, where you’re bringing in these update observations, because the observations you can only have after we’ve taken the observation.
Okay?
And so these LDS systems are in a lot of places, you know?
It’s related, first of all, to weather forecast.
Weather forecasts use LDAS, as well as assimilation of atmospheric variables.
So those are used all the time.
We also use these LDAS in a lot of the work we do, for example, on agricultural monitoring in the United States.
Climate assessment reports are very often include LDAS, like the National Climate Assessment of the United States.
Work we do with the Famine Early Warning System in Africa.
These LDAS are known to be pretty useful ways to get information.
And so some of them have outputs that are available, that you can just get, because there’s already someone running it.
If you’re interested in that,
please contact me and I’ll try to put you in touch.

And then sometimes we run them ourselves to optimize them for a region we have here. There’s a couple more minutes on this, just so you understand the basic principles here.

One of the most important starting points is satellite-derived rainfall. We’re using a couple of products here. I’m not going to bother with the acronyms. They don’t matter. They are, in case anyone attending today is from the satellite world and is interested in what we’re using, okay? So CHIRPS and GPM-IMERG.

We then use that MODIS satellite that I already described, get our land cover and vegetation characteristics.

And this cartoon here is showing you our model. It’s called the Noah MultiParameterization Land Surface Model. And what it’s doing is it’s simulating multiple layers of the soil, different vegetation types, shallow groundwater.

We also work into it a downscaling routine to get better surface meteorological estimates. It doesn’t simulate the atmosphere, but it can help to downscale atmospheric conditions.
And it also does snow, which actually does matter to us because we want to get the runoff coming out of the Andes, but it doesn’t matter locally in the Amazon, obviously.

So, that’s all kind of for the local energy and water balance solution. We use Noah MP. We then couple it to a river routing model called HyMAP.

And HyMAP, the hydrological modeling and analysis program that’s what that stands for, allows us to model things like the flood plain, and that’s, as you can imagine, really critical when you’re talking about mosquito habitats.

So we get the river heights. We also get the river width, and the area of flooded river boundary at any given time.

We run this at five kilometer, gritty resolution. Five kilometers by five kilometers, or 25 square kilometer. And then around Iquitos, that city that has the largest population center.

We nest into one kilometer for some higher resolution analysis. As we run the model forward, we can take advantage of these assimilation capabilities.
and we run multiple simulations for different purposes. Sometimes we might be assimilating satellite-derived estimates of soil moisture, or leaf area index, or water storage, terrestrial water sources, meaning all the water stored in the soil column and groundwater. These are all observables at different resolutions from space using different civilian space missions. And those will all help to improve the performance of our model. And then you can get an output like what I'm showing on the right-hand side of the screen here, which is just a standardized anomaly in soil moisture, showing a period where, in our area of interest, for example, there were some drought going on in the Northwestern Amazon, as shown by a negative standardized anomaly in soil moisture, as captured by our system. I'll come back to this in a moment, but this particular snapshot is an interesting example, showing what might be considered the classic El Nino pattern, okay? So it’s an old snapshot. This one’s from 1998. I’ve accidentally cut the date off of it.
There's the monthly anomaly from a month in 1998. And what you're seeing here is the 1997, '98 El Nino bringing catastrophic flooding to the coast of Peru and Ecuador, and drought to the Amazon basin. Okay, I'll return to that in a moment, but that's kind of a classic El Nino pattern in the region. And so, here's just a quick animation to show what you're getting through time. I'm showing monthly up what's here. In fact, we get, you know, hourly outputs from the system that we can then extract for different geographies to perform our malaria analysis. Information on things like your air temperature anomaly, your rainfall, your soil moisture anomaly, your runoff, your river height, et cetera. Okay, and so this is all the information that we're going to be bringing in, combining with the land cover and ecological information, to try to get this environmentally informed malaria analysis and early warning systems set up. So, one thing that you might be wondering is, "Okay, I just mentioned this was a data scarce area, right?" And these are outputs of some system.
that’s combining satellite data with its uncertainties, and a model with its own uncertainties. How good is it, right? And can you trust it? And the answer is that in any study you do, where you want to make use of this kind of environmental data, you want to make sure that either you or someone else has done a good, clean analysis of how well that system performs in your region and season of interest, okay? You don’t want to just take this off the shelf and say, “Oh, I know this going, going to be fine where I am.” And so we’ve done some analysis. I’m not going to make you sit through our whole analysis kind of thing that we spend our days, nights and weekends doing, right? Make sure the systems work well and trying to fine tune them. But we have some data here that Cristina Recalde, a PhD student working with me has from Ecuador, and some data from Peru, looking at things like, "Okay, how well do we do in observations in blue, versus our model simulation on rainfall?" And there are good and bad things if you stare long enough at this chart,
like, yeah, we’re in about the magnitude of rainfall is not bad. The seasonality is pretty good most places, but then you’ll find there’s some wet and dry bias in different places that we’re investigating. Similarly, you can then look at the soil moisture. Soil moisture is harder, because rainfall, there actually are rainfall observations. Not many, but there are some, right? Soil moisture, there’s like basically no in-situ observations in a consistent way in the study area, and so we have to use satellite data to compare it to. So here, we’re comparing this observation in gray, which is really a satellite observation, with our model simulation. And again, seeing some good, some bad. Here, we really do have to question the fidelity of both the observation and the model, since the observation is satellite-derived. At least it gives us a sense. Do we have a consensus across our different estimates, as to what’s going on here? And this is tricky, right? Because getting soil moisture right in a complex hydrology like the Amazon is no trivial task.
So this is a scenario where we spend a lot of our effort.

Last point I want to make on this physical hydrology hydrometeorology before finally getting just the natural malaria results:

It’s really important, whenever you’re doing a study like this, right, to distinguish between, when I say that there’s hydrometeorological variability, am I talking about geographic variability? You know, wet versus dry places. Am I talking about seasonal variability, right? A wet season versus the dry season, for example.

Or am I talking about something like inter-annual variability?

Like, "Oh, we had a drought year, or we had a year with more flooding.” And that’s really important, you know, first and foremost, to understand process, right?

You want to know that you get a statistical result that comes out of throwing some environmental variables into your model.

They’re significant. What is it that you’re seeing? Right?

And also, is what you’re seeing a proxy for something else?

Right?
If you classically see like, "Oh, there’s a wet versus dry season response," or a warm versus cold season response, and when I look at my cases of malaria, is that because temperature’s affecting malaria, or is it because there’s a seasonal cycle in temperature, and seasonality for some other reason is affecting the malaria, and I’m calling it temperature? Okay.

And so, you want to be clear on whether you’re looking at the geography, the season, or the inter-annual variability. And this is on my mind a lot these days, both because I do a lot of this work. And as I know Kai appreciates and probably others in the audience as well, there’s a lot of conflation of these things in the COVID-19 literature, with different claims or attempts to claim environmental sensitivities. Some really good work, right? But also a lot of these kind of naive, I would say, studies that came out showing correlations or associations that were simply showing a seasonality, or you know, a coincidence of two patterns.
The whole correlation versus causation problem, that I think part of the problem there was a misunderstanding or there’s a mis-framing of what kind of climatic variability we’re talking about. Okay, got off that soap box. And simply say for that third thing, all I’ve shown you here is seasonality and spatial variability. I haven’t shown you inter-annual variability. I want to comment a little bit on that in this region, because anyone who’s worked on malaria in the Amazon or other malaria zones probably are aware of a lot of studies, good studies, right? That have associated malaria with various large scale climate modes. Certainly these drivers of variability, okay? And so the big one is El Nino. The El Nino Southern oscillation, okay? But there are many others. It’s an alphabet soup that I won’t get into. El Nino, in this part of the region. One might well expect an El Nino effect here, right? It’s called El Nino because of the effects it had, you know, was first characterized in the coast of Peru, and what it does to the sardine fisheries.
off the coast of Peru.

And so, this is kind of like the home of El Nino, right?

And so, we certainly expect an El Nino effect.

And as I showed a few slides ago, a classic pattern would be high rainfall on the coast,

drought in the Amazon,

for dynamical reasons that I won’t get into.

It’s not that simple or that predictive as a simple univariate association in this part of the Amazon, at least.

There’s some other parts of the Amazon that respond a little bit more reliably,

but I’ll tell you, it’s always a little complicated.

But here, just taking it again from Cristina’s work here,

looking at El Ninos and La Ninas the past 20 years.

And if it’s red, it means you’ve got drought,

or drier conditions.

If it’s blue, it means you have wet anomalies.

And again, during El Nino, we should be seeing red

in the Amazon.

And here, you know, we got our Laredo state.

Sorry, it was just Ecuador and Peru I’m showing you.

So we’ve got this kind of, here’s your Northern Amazon portion of our study region.

And what you’re seeing is that, yeah, during some El Ninos,
you do see that drought pattern, okay?

But you also see it in this La Nina,

and then there are some El Ninos

where you don’t see it at all,

and in fact, that big monster El Nino that hit in 2015

and had effects globally, it was wet

in our part of the world,

when you might’ve thought it was supposed to be dry.

And so, there are some complications here, okay?

All I can say that one could use, and so El Nino,

oscillations effectively, statistically,

in a forecast in here,

if you accounted for enough other variables.

I’m highlighting the fact that it’s not enough

of a predictor of rainfall in its own right, okay?

But combined with other factors,

you can probably get some scale.

But we decided to take a different approach,

which is, rather than using these kinds of teleconnections,

these like remote connections to El Nino directly

in our model,

we run a dynamically based forecast.

And so what we’re doing there is, again,

this one’s a little detail for those who might be

working at this interface of climate and health.
We run what we call subseasonal to seasonal forecast. You know, a few weeks out to...
Well, you can go to nine months. We're really only going up to three months right now, for this application.
And what you do is you take what I already showed you in the LDAS, the satellite landscape analysis, run it through a land data simulation system. That provides initial conditions, from which you generate an ensemble.
So your seasonal forecasts are these probabilistic ensembles, just like weather forecasts.
And these are these global atmospheric models that we run forward.
We run them forward using initial conditions of the hydrology locally, and the ecology locally.
We downscale the meteorology from those global forecast systems using some algorithms that, again, I won’t get into, but happy to follow up with anyone doing this kind of work.
And then, we put that into hydrologic work. As we run it through the same modeling system,
it’s no longer data simulation because we don’t have observations.
We run that system forward.
Okay. So why do all of this? Because it pushes your forecast time horizon out.

If I, as the climate guy in the team, give Bill and Mark, the epidemiology guys on the team, a monitoring system that is operationally saying what the moisture is right now, they can forecast malaria because it’s a time lag, right?

So they’ll get a pretty good forecast, because it takes time for the signal I’m sending them to propagate through the ecology, and the human systems.

But if I can give them a forecast of what it’s going to be like in two months, that gives them, you know, eight weeks more lead time, and you can make a different set of decisions, given an extra two months, right?

So it’s all about this uncertainty time horizon trade-off year. The more we push out for a greater time horizon, the greater our certainty, but also potentially the greater power that kind of system can empower.

So, how did these forecasts look? I’m not going to make you sit through a whole forecast scale analysis,
but just want to make one point here. If you just focus, let’s say, on correlation here, for the sake of time, if there’s hashing, it means a statistically significant scale. And what you see here is that looking at something like soil moisture, we get really good forecasts for one month, and then it begins to degrade, particularly degrading these wet areas. You’ve maintained some forecast scale out in the dry areas, because there’s so much memory, right? If it’s not raining much, most of the initial conditions that matter. But as you go out, the result here we might say is that we can really do a nice job of getting you an extra four weeks, right, on the system. If you want eight weeks or 12 weeks, we’re not going to be contributing that much stuff in the forecast. And so it’s important both to have the capability, and to understand the limitations of the capability. All right. So we do all those analyses. And then, this is not my work. This is work that Bill led. He took all of this ecological and hydrological analysis,
and did an objective regionalization,
did principal components analysis on the variability.
End up with these three different factors that are loaded by different properties of the system,
and counting for about, you know, human systems,
as well as land use and hydrometeorological conditions.
And from that, derived these seven socioenvironmental regions.
And the principle here is that these regions are reasonably homogeneous and regionally distinct from each other,
with respect to human and environmental factors.
And also, as it happens,
this was not necessarily integrated to that,
but because you’ve included the human systems in the analysis, most of the travel stays within the region.
And you typically have similar vector species within a region.
Okay, and similar livelihoods.
So, what we then say we’re not going to develop one malaria risk model.
And again, this is now, we’re seeing Laredo regions,
so this part of Peru.
We’re going to develop a system that has customized models, based on socioenvironmental regions. So, in the remaining time that I have, which isn’t much, I know, so I’ll touch on these lightly, but these are just examples of how we can pull this all together, all right? And so the first thing, this whole project is forecast, right? And so, using these socioenvironmental regions, then aggregate malaria data, which we have about 300 health posts contributing data, passive surveillance. They get aggregated to a socioenvironmental region. And then we try to predict whether there’s an outbreak, based on the Ministry of Health’s definition of what an outbreak is, which is, you know, exceeding a certain threshold, in terms of case number per population. Again, this work led out of Duke by Bill, and he uses observed components model as a statistical method, and was seeking to get a time horizon of four to 12 weeks. And again, because it’s customized by region, what you’ll find is that the model
has different variable importance and is structured differently for the different models.

So region one, which includes Iquitos, so it’s kind of like our most urban area,

we can describe that in terms of the characteristics of the socioecological region.

And then we can say, "Okay, what explanatory variables from our environmental suite end up being significant?" It turns out to be soil moisture.

We can then look at a region like region three,

kind of really out in the forest,

very low population density.

It has a different description.

It’s going to have different statistical characteristics to this unobserved components model.

And in this case, minimum temperature came out of the more significant variable.

Both of these variables, of course, if you look at the literature, are using malaria prediction.

So they’re both plausible, they’re possible pathways.

different ones came out as more predictive in these different regions.

Okay? So then we run the system.

We have to run the system starting four weeks.
before the present.

Why?

Because it takes about four weeks for surveillance to come in.

Here’s the percent of health post reporting of malaria data.

As you can see, this is time, this is the present.

At the present, you have fewer than 20% reporting.

If you go back four weeks, you have close to 100% reporting, which means that you have a good...

You know, previous cases are important predictor of future cases.

So the forecast includes a four week forecast of the past.

And then, we want to go out to eight or 12 weeks.

We have this moving outbreak threshold, because it varies seasonally and by location, what MINSA, the health ministry decides is the right threshold to declare an outbreak.

And then we might have an observation, and a competence interval around that observation.

Just to give you an example of performance, 2016 was the first year we really tried this.

So this isn’t just a systematic analysis, just showing you the kinds of things you look at.
True positives, false negatives, false positives, true negative.

For an outbreak in any of these eco regions, looking at eco region one and three here, over the different forecast time horizons, our sensitivity and our specificity.

In a nutshell, we do really well in eco region one.

Fades a little in specificity as we get out to 12 week time horizon, still pretty good. Eco region three, we do not do that well, okay?

And so again, small sample one year, but just our first test was showing us that we’re going to get different performance in different eco regions.

Okay.

And so, that’s all at the eco region level. I’m not going to get to too many more results at that level right now, but rather say that to be decision relevant, we have to go down to the district level.

So, the lines here on this map are separating the districts.

Okay.

And so the colors of the eco regions aligns with the district.

We really want to be at a district level.

And so for this, again, won’t get to the details right now, but Mark Janko implemented this hierarchical
Bayesian spatio-temporal logistic model, where you basically have your district outbreak probability being a function of the probability of an outbreak in the eco region that contains the district, and some district-specific properties. When Mark downscaling and looked at some of these analyses and then did an evaluation over a retrospective period, these are the kinds of sensitivities and specificity we’re getting for different districts within each eco region. Again, just showing you eco region one and three here. And you’ll see that again, pretty high variability. So we were doing well in eco region one at eco region level, but you’ll see that, for example, in the district of Fernando Loris, there were some pretty significant errors in this retrospective period, and different kinds of errors in different places. So also for us to look at, kind of uniformly doing worse in general, than eco region one. So why is that? Why are we doing poorly in region three?
Multiple reasons.

One thing I want to emphasize is that eco region three was very much located kind of up in this area.

So first of all, malaria cases are generally low there in total, because it’s such a sparsely populated area.

But it’s also a border area.

It’s a border area that is transected by trans boundary rivers.

The trans boundary rivers are the transportation.

The model fits most poorly here in eco region three and another eco region dominated by trans boundary river.

Doesn’t do well in places along the rivers. Okay?

And so that’s one big weakness in the model that we’re working on.

And oops, the slides got reversed.

And I just want to point out that we are looking at, and we had a paper recently, led by students.

And so this is students from Duke, Johns Hopkins, Ecuador and Peru, who took the initiative to really lead an analysis of this cross-border spillover.

And that’s something we’re looking at now.

Okay.
So, that’s where the forecast system is. We brought it in 2019. We did some operational forecasts for the Health Ministry. We brought it in 2019. Was all looking good. Then there’s political change and COVID, so we’re a little bit on hold right now, but we’ve got a system that we’ve proved we can use operationally. We continue to try to improve the performance.

Policy evaluation. Okay. So I’m going to give one example of policy analysis we’ve done. That was PAMAFRO, which was this project for malaria control on the Andean border areas, active 2006 to 2010 or 11, depending on how you counted. They did four kinds of things. Long-lasting insecticidal nets, better rapid diagnostic tests, and other monitoring tools, case management, with antimalarial drugs and training, and environmental management for vector control. So doing these four kinds of things. And it kind of worked, right? So this is by vivax and falciparum in Laredo. And it sure looks like over the PAMAFRO period, the case counts were going down, down, down,
approaching eradication, which was the goal of the program.

Then stops suddenly in 2011, cases start coming back up.

And what we can do is we can leverage that district model that Mark Janko developed, right?

Not only using it for forecasts, but then saying,

"Well, let’s include in that model structure the different interventions, especially with PAMAFRO."

Because we know at district level and with monthly timing,

what kind of interventions were done where.

Let’s integrate that to a model and then do an interrupted time series analysis, and see what those interventions actually accomplished on the background of climate variability, and all the other variables in our model.

So kind of an environmentally controlled analysis of the effectiveness of the intervention.

Mark’s found is that, well, you can kind of quantify this.

So the blue line here in the top left, top is vivax, bottom is falciparum.

Blue lines are the model, dots are the observation.

On the left, we have the PAMAFRO period.

And we see that our model, if you don’t tell it about the intervention,
systematically overestimates the cases in this period, for both vivax and falciparum.

In the post PAMAFRO period, starting in 2011, quite the opposite. Our model has cases down here. The observed cases were much higher. And so, take those together and come up with estimates that about 150,000 cases were averted by PAMAFRO. That was the amount of malaria averted thanks to PAMAFRO, and had you continued it for another five years, you would’ve averted another 150,000, not to mention the long-lasting impact of driving cases that low, right? And so here we have an analysis of both the effectiveness and the cost of removing a program without a good continuity plan. And then you can zoom in, because again, we have this district level information on each kind of intervention. I see I’m running out of time, so I won’t spend too much time walking through these maps, but green shows incidence ratio less than one. And so we can look district by district, Okay, for falciparum and vivax,
environmental management, bed nets, etcetera,
in which districts do we see the most effect when we add or remove this from our interpretive time series analysis?”
And there’s some interesting patterns that appear that we’re in conversation with some of our partners about to figure out what might be effective in the future.
One of the cool thing just mentioned that you can do with this is try to figure out how much malaria and dengue there is right now in this area, because we have no idea.
If you look at what happened in 2020 with surveillance, I mean the health system basically shut down. And so, it looks like it was a great year for malaria control, but of course it wasn’t. So we can then use this same modeling approach to try to estimate how many cases there really were in the year, 2020 and 2021. And as you can see, we estimate that there were at least three times as many cases. Okay. Last point I want to make here is that I’ve showed you some malaria modeling cases
that are process-informed, but at their heart, statistical, right? These are empirical analyses. And looking at intervention scenarios, we are also looking at explicit simulation of behavior, okay, to get these coupled natural human systems right. And the way that we are doing that, led by Francisco Pizzitutti, is with agent-based modeling. And this is a kind of Coolidge based model Francisco built, in that it has agents that are mosquitoes, humans, and plasmodium, okay? So. you have all of these are agents interacting. And here is just an example of one of the villages where he’s applied this, where you can have different households, and all these agents are interacting and influenced by the environment. In that here, we see different kinds of breeding habitats influenced by seasonal flooding, with information from our environmental analysis system, changing the hydrology. And then you’ve got the cases happening in this household, each of which is also experiencing its own environmental conditions, okay?
You can then run scenarios of control. For example, vector control strategies, one thing we like to look at. And so we’re looking at here at one of these environmental control applications, and saying, "Well, what if you do larval habitat control around a certain buffer radius, around each household, right?" How well do you do at 50 meters, 100 meters, 150 meters, 200 meters, when you talk about malaria incidents? Total vivax falciparum. And the idea here is that, by understanding this agent based model movement patterns, right? And the sensitivities of the different agent types, we can get a sense, say, "Well, really you want to probably get out while you take your pick, but I would say at least 150 meters might be considered very effective. Anything beyond 200 is unnecessary." All right. And so this is parametrized for one set of villages. It’s very data intensive, but nevertheless, I think it indicates a powerful way to, you know, use your environmental information.
in a different manner, not as an empirical predictor,
but as a variable within a model in which different agents are responding
decision rules to this variability.
You can also use the same tool, and Francisco has,
to look at the importance of mobility, right?
So that’s something people talk a lot about in the past couple of years, right?
How much mobility influences disease transmission.
It’s an old story from malaria.
What you’ll see here is if you look at your observed black line here
of the average monthly malaria incidents
along the Napo river,
first thing you know, is that,
"Well, okay, if I run this model with no asymptomatic cases
considered in travel,"
you assume that no asymptomatic people are traveling,
you way underestimate the incidence rate.
So we know there’s a lot of asymptomatic activity going on.
And then we can say,
"Okay, as the percent of traveling workers increase,
we would expect the incidence rate to increase."
And we’re right about the right order of magnitude.
And it looks like some of this movement
really does need to be accounted for,
to understand the incidence rates
with significant implications, again,
or how you would do monitoring and control in the region.
So, ran a little longer than I wanted to. Sorry.
That’s what happens when you let professors talk.
But just a few of the next steps here.
I break them into four categories.
We’re really working on the application here.
As I noted, there’s been a lot of political turnover
in Peru for those who know the region,
which has hampered our ability to operationalize a forecast.
So now, we’re starting to train and transfer to some universities and research institutions
in the region, rather than straight to the government,
to be able to spare stability.
We’re just having our first meeting this week
on an effort to expand to include Columbia and Brazil.
So it’s a big up-scaling of the effort.
And we’re also seeing,
can we transfer this to an area in central America,
working with the Clinton Health Access Initiative, sorry.
Flipped the letters.
On Central America, where the case counts are low
and therefore the ecology and the environmental sensitivity of the system shifts. It seems to cross a threshold. So we want to see how the approach works there.

And last, but certainly not least, through these combined methods, but again, all trying to leverage the power of the different fields to understand malaria sensitivities. How can we continue to explain these coupled natural human mechanisms, which, despite the fact that we've known about these relationships since ancient times, we continue to struggle to understand in a predictive manner today.

I realize I didn't leave too much time for questions, but maybe we have time for a couple. So, thank you again for the opportunity to talk.

I think we only have time for one question, and the students have already read the papers that you mentioned published in your page. So, any of you want to ask a question directly?

Okay, so let me ask you this question.
So Ben, you gave a wonderful talk on the importance of value, time and migrating, the importance of having the data, and then from the very state of the art subseasonal to seasonal forecast. The students when they read the paper, they have a question regarding (indistinct) also COVID-19 related.

So, did you see how to apply this malaria focus system? The application to COVID-19 control focus system?

Yeah. Interesting point.

So, I'm going to answer in a very general way. They're obviously very different diseases, right?

We're talking about a vector-based tropical disease versus a pandemic virus with a lot of airborne transmission. But I would say that the general challenge of bringing these different data sets together is really critical. And we can do cross-learning across diseases, because one thing we've really struggled with in COVID is to bring all the information together in systematic databases for responsible analysis. And we were able to leverage some of the things we've done with malaria and other tropical diseases,
to build COVID information databases, to support research. And I know that Kai did his own work to pull his own database together. So moving forward, how can we use all of these diseases to inform those kinds of data structures, I think would be... And cross-learning approaches will be the way to go. I wouldn’t necessarily endorse any single thing that I did here on malaria as the answer for COVID-19 model. They’re too different. But if you can really focus on that kind of informed integration, I think there’s a lot to be learned. <v Kai>Thank you so much, Ben.</v> And thank you, guys, for coming today, and thank you for our online audience. And just kind of reminder that today’s lecture is recorded and will be available online, on our (indistinct) websites, so you can check that. Want to sincerely thank you, Ben, for giving this incredible talk. <v Benjamin>Great, thank you.</v>