EMME and EMMEnet - a Brief Introduction

The EMME Project consists of development, dissemination and coordination of a toolset (EMMEnet) that is used by the global health community for early detection of indicators that are significant for potentially pandemic-scale outbreaks of infectious diseases. The network of sensing and analytic resources provides a method of data fusion and human-machine intelligence for identifying, within flora and fauna variations linked with environmental and particularly climate changes, a broad set of pointers and signs of species migration and mutation, concentrating upon microorganisms and vectors that can lead to infectious transmissible zoonotic diseases.

EMME = Environmental Monitoring of Species Migration & Mutation to Minimize Zoonotic Disease Epidemics

The EMMEnet toolset consists of a diverse , open-ended and interoperable suite of technologies for physical sensing and monitoring of species transitions and notable changes in habitat and interaction. This toolset provides methods of data analysis, information extraction and knowledge generation. These incorporate satellite, robotic and manually-deployed sensors, including devices for multiple types of spectroscopic detection of changes in vegetation and atmospheric trace chemical contents. In addition EMMEnet processes include qualitative observations made by human and robotic observers; these are incorporated with specific quantitative data and a suite of algorithms for analysis and interpretation.

EMMEnet is designed so that it that can be employed by a very diverse group of operators in virtually any region of the planet, with an emphasis on particular regions that are most likely to provide predictively-useful information. The toolset enables creation and management of datasets which are openly distributable among investigators and analysts in different communities that are concerned with public health, epidemiology, microbiology, and clinical practices of medicine. The outcomes from deployment and use of EMMEnet consist of information that is used by epidemiological and public health forces worldwide to ascertain the emergence of new biothreats from known pathogens and potential new variants, enabling more expedient response with measures to eradicate or minimize the health threat and risk of epidemics outbreaks.

This toolset and the resulting datasets form the primary technological component of the EMME Network (“EMMEnet”). The network is used by people in a wide variety of institutions and initially within centres that are collaborating in the EMMEnet development and deployment. Initial geo-focus of deployment is in two region of the planet – East Africa and Southeastern North America.

EMMEnet may be metaphorically compared to a collection of applications that operate through the internet and result in distribution of observations that are shared and result in verifiable and predictable data. EMMEnet is employed in a controlled manner by persons trained in the requisite basic health sciences and it involves the use of a network computing and communications environment that supports the acquisition, evaluation, and interpretation of the datasets, leading to actions that follow logical epidemiological and public health methodologies for biothreat verification and consequent prophylactic response.

EMME Network

- Formal methodology and toolset that can be deployed and used widely by many groups and institutions, in a variety of ecosystems and human societies, for monitoring environmental parameters and interpreting observed changes in flora and fauna that can be employed in models to predict migration and mutation, due to ecosystem changes, of identified pathogenic agents and their vectors
- Sharing of data acquired, assessed and evaluated, and information produced through different models and simulations, that is pertinent to identifying the locations for precise attention to pathogenic agents and vectors, and risks of contagion and transmission among humans and domesticated animals, including activation of pre-epidemic epidemiological countermeasures
- Engagement of general population, especially youth and students, in the processes of education on relevant topics of environment, epidemiological observation, public health and personal hygiene, and in the processes of conducting EMME Network tasks that are pertinent to early detection, circumvention and prevention of epidemic-potential infectious diseases
- Engagement of social media networks in the implementation of this methodology and use of this toolset
- Creation and management of a population health informatics biomedical equity resource (PHIBER) that is pertinent to the tasks of avoiding and minimizing the consequences of epidemic outbreaks
- Sharing of the PHIBER information with researchers, educators, public health agencies, and medical institutions including pharmaceutical and other medical technology developers

3. Consortium organization to develop and operate EMMEnet

Among people and institutions involved and becoming involved within EMME are:

TETRAD Institute of Complex System Dynamics (lead organization)
Dr. Martin Dudziak PhD (principal project leader)
Dr. Kevin Ciresi MD
Dr. Larissa Brizhik PhD
Rachel Roman RN

Muhimbili Univ. of Health and Allied Sciences (MUHAS; Dar es Salaam)
School of Public Health and Social Sciences
Department of Environmental and Occupational Health
Dr. Luco Mwlange
Dr. Jovine Bachwenkizi
Dr. Deogratias Mzurikwao

Refer to project introductory slides (http://emme.tdyn.org/EMMEnet-10-simple-graphic-illustrations_mjd_12aug22.pdf) for additional information on other participants (USA, Europe, Tanzania).

Data Acquisition Focus

Physical (climate-focused)

- changes in rainfall, humidity – averages, peaks, and shifts in seasonal behaviors
- changes in vegetation types linked with classes of pathogen vectors and their predator-prey relations
- strong emphasis on the carrier species – plant and animal
- changes in deforestation, urbanization, agri-industrial use, waterway changes
- concentration on regions with highest probability of showing first signs of significant changes that can be linked, predictively, to type types of microorganisms of interest and concerned
- wetlands, marshes, riverbeds, lakes, watersheds, areas of known carrier-type animal habitation and transit

Bio-specific

- detection, verification and tracking of specific pathogens and vector species, and related flora/fauna known to have high probability for presence of target pathogen species

Social/demographic

- movement and relocation of human and domestic-animal habitations
- medical reports indicative of diseases and comorbidities which are closely associated with a set of the initial target diseases (which will be selected by the formative team - (these are TBD but we can expect that certain viral pathogens with higher risk for lethality and transmissibility will be the main targets).

SI (AI, ML), data modeling, pattern recognition, predictive analytics

We will be using the synthetic intelligence ("SI") in three principal ways:

Clean-up data, reducing noise and error in reported observations

Generation of simulated ecological conditions and observations, for use in modeling and projecting migration and mutation ("M&M") behaviors, and then applying this to the observed-world data, feeding the simulation into the system for training of algorithms for improved alertness and detection of potential actual-world disturbances that indicate M&M events that demand closer attention.

The actual formal logic algorithms for generating inferences regarding M&M events on the basis of observations that are principally indirect; i.e., changes in flora and fauna that are not specifically detections of pathogens or even of the associated vector species. A simple example that illustrates this part of the system is the detection of changes in grassland growth accompanied (or not) by changes in herd behavior of deer and other ungulate species which feed on those plants. This gives an indicator of the spread of various parasites such as deer ticks. Another example may be similar changes in vegetation and zoology that pertain to bats and other animals, especially small mammals, which are associated or linked with coronaviruses, ebola, yersina pestis, etc.

Initially these logics will be fairly concise rule-based inferential algorithms. The complexity is in the interpretation of raw field data from multiple sources that can lead to reliable assessments about the accuracy in observation of change in plant and animal movements. The main outcome from the inference engine will be weighted indicators for doing more detailed field observations that will then transition into actual sampling, in highly specific locations, to verify if there is presence of specific pathogens. After such determinations, then the existing and in-place epidemiological response processes need to be undertaken, which leads directly into the public health infrastructures for the regions involved and, if pertinent, on a national and international level of response.

Abduction, Inference and the use of Simulation

At the heart of EMME we are aiming to validate one or more hypotheses, in parallel, and to rule out a huge number of others. But from the outset, and even as we go forward, we do not and cannot know precisely all the hypotheses we need to be examining! This is because of the "nature of the Beast" - meaning, simply, that we cannot expect to have "advance knowledge" or "sufficient knowledge" to predict, "out of the box" (from existing bio-enviro-medical science) what may emerge and what may change among known species, etc.

We are engaged in using multiple and even competing logics to reduce our search spaces - our geographic, physical, environmental bio-social-sphere search space, and our analytical space of what to be looking for in the heaps of data that we have and those we will be continuing to add to and build higher and deeper.

Our primary logic work is essentially performing abductive inferencing that is based upon the "sparse network" observation data. In a way, we are aiming to perform, somewhat in parallel, competing forms of reasoning - formally, these forms are called logic-based abduction, and abductive validation, and set-cover abduction. (Don't worry about these terms or the implementation! Read C. S. Pierce for good foundations on all this.) The important point for here and now is that we have a large set of hypotheses (regarding outcomes to be sought within the pathogen/vector biosphere and then verified or discounted by empirical observations) and we aim to reduce this hypothesis-space with what we are observing from the first phase of analysis of the environmental datasets that we are working with.

Think of it this way, going back to the needles-in-haystacks metaphor: the more we can refine what are the possible shapes and geometries of the different kinds of "needles" (in this case, evidence of migrations and mutations that do or can lead to greater infectivity, transmissibility, lethality, and so on), that helps us to constrain our search space and everything we do (collecting data, especially new in-the-field data) to a smaller number of "haystacks" as it were.

Reduce the types of possible needles to search for, thus reduce the types and locations of haystacks wherein those types of needles might possibly be, and thus reduce the workload of searching and analyzing, and increase the speed and accuracy of finding what you are really, ultimately, seeking - the "poison needle" (the bad-ass pathogens and their vectors)...

One last point along these lines:

Our work with EMME is not just about collection, acquiring, and analyzing "data" from "out there" (environment, world). We also simulate the same - we generate "maybe-what-if environments" by employing our VAE/GAN neural net models to generate loads and loads of artificial "everything". We simulate "what ifs" for a myriad of changes in behaviors of certain species, in order to create scenarios which are for our abductive reasoning engine to "eat and chew on and digest", and to better tell us:

Look for these combinations of attributes in the environment, these dynamical patterns of weather, vegetation, measurable chemistry, animal movements, and inferences thereof about those kinds of behaviors - and if you find these, then you need to dig deeper, and also to examine in more detail, such as through sending out field teams (robots, satellites, people) to get more localized data, and also samples, which can then be put into instruments (mass spec, GC, PCR, immunoassay, SIMOA, gene sequencing) for analysis of the signifying details which can tell us what things are changing and in which manners.

Ultimately the outcome of EMME (one of them, a major one) is to produce sets of coordinates in a complex space that includes much more than the 3 spatial coordinates to which we are accustomed.

This outcome-set is something like this:

- Physical region defined by various geospatial coordinates
- Time coordinates, for some focus-narrowing on when we ought to expect some "X" to be taking place in the indicated region
- Pinpoint geospatial locations within that region that tell, "Go look here, especially, or first-and-foremost"
- What we are supposed to be looking at and for (e.g., sampling of some vegetation, animals or some material connected with them, microflora, viruses, etc.)
- What the system logic thinks may be going on (e.g., mutations in the RNA or DNA of some virus)
- Maybe also some pointers about the key factors of infectivity, lethality, transmissibility
- and the list does not stop here...

We attain these outcome-sets using the abductive reasoning logics, and some of that is downright "classical" but also we have several components, sub-systems as it were, that are again employed in parallel. One of the system paradigms is *Nomad Eyes*, which was a long multifaceted chem-bio-rad detection system that got its start in the early 2000s. One of the algorithm paradigms is something called HORUS - Hierarchically Organized Reasoning and Understanding System. It is very human+machine combined, and the three key elements involve formations of definitions (propositions, assertions, some axiomatic), associations (some deductive, most abductive), and relations ("trimming the hypothesis mega-shrubs").

But HORUS met up with SPSA in a big way about 5-6 years ago, and got smarter.

SPSA is a fairly well-known term, Simultaneous Perturbation Stochastic Approximation. It goes back to Kolmogorov, godfather of complexity theory and turbulence, and Arnold, another godfather in maths, especially non-linear dynamics and catastrophe theory.

With those outcome-sets, we perform the actions recommended and found to be sensible and then we achieve some of our goals which include catching a predator-pathogen before it gets out of control and is doing damage to humans and other species (including our food supply).

Among technology compoents in EMME are:

CHANT (aka Banyan) = Cooperative Heterogeneous Asynchronous Network Transprocess
ATHOS = Adaptive set-Theoretic Hierarchical Operating System (what evolved out of the HORUS)
ACTOR = acquisition, cognition, activation, operation, response
APIS = Anomaly and Pattern Interface Schema

In addition there is MADIT. This brings us straight into the challenges that are front-and-center in any viral outbreak, and we were involved in this back in the late 2000s with the H1N1 mini-pandemic and the subsequent almost-epidemics with various avian influenzas like H5N1 and H7N9. Those are among the diseases on the horizon and we need to be addressing them going forward with EMME. Even Lynga virus, a new one showing up now in China, is one of these new genre of zoonotic diseases, where there may have been very isolated cases of a human contracting some x from an animal vector, and now, more and more we will be seeing future cases where such a virus, heretofore not transmissible among humans, makes the adaptation and is easier to go from one human host to another. That spells pandemic-potential and it can spell Plague-potential if the virus is a fast mover and a lethal one.

MADIT (Mutation and Anomaly Detection, Identification, and Tracking)

It is not strictly a "neural network" computationally speaking, but more like a DAG (directed acyclic graph).

Just to round it out, these are algorithm-computation terms that are much more common in the history and literature, and they are the foundation-stones for all of the above:

CSP = Communicating Sequential Processes
MIMD (multiple instruction multiple data
PDP (parallel distributed processing)
CCC (C3) - Communication, Control and Collaboration

6. Notes on the STEM education and social networking components

There are several “thrust” areas in this and all have been done successfully in prior years, in variant activities with schools and youth groups, and in partnership/sponsorship with various institutes, universities and companies, in USA, EU and other countries (RU, Ukraine, Morocco).

The main points here:

[1] Direct engagement of youth in high-school levels in the fieldwork and in the data science tasks. Thus, some of this will involve students operating some of the drone-based and manually-deployed sensors, and some will involve their participation in the computer-based, remote-location work of processing data that is collected from various sources including web-accessible databases and especially satellite (e.g., Sentinel) real-time streaming sources.

[2] Public health education through existing and new social media - that means all the mainstream social media applications.

[3] Additional but similar health education for healthcare workers in all relevant countries beginning with USA and East African nations. This is about advancing their levels of awareness and watchfulness on a mass-scale.

[4] A unique and special CubeSat project activity involving students fro high schools and universities. While this may not be essential in the beginning, it really fits well for EMMEnet operations and for dissemination information and education among People!

This is principally a STEM education element but designed to provide novel data streams into EMMEnet. One CubeSat will be designed and constructed by a joint team of primarily students in Tanzania and USA working together.

The satellite specifics will be similar to those outlined in prior Tethys project documents from 2018-2020, with appropriate modifications in order to deliver information more pertinent to targeted flora/fauna change patterns of interest, and to accommodate and integrate with new work that has emerged during the past 3 years. The specific choice of sensors will be determined by the EMME team and may include spectroscopic measurements that will assist in estimates of vegetation changes underway and expected in the observable regions.

The orbital path of this satellite will be in keeping with the initial focus upon East Africa and Southeastern North America.

The satellite will optimally be launched aboard a SpaceX rocket, and we are in discussion with SpaceX regarding several aspects of support for EMME including use of their StarLink network.

Robotics and field-based monitoring tasks

We will employ geospatial intelligence (GSI) technologies such as have been studied and developed by our team members and others to optimize the task of pinpointing where and in what sequence field-based observations should be conducted, on the basis of the overall mapping that is generated from the combination (fusion) of satellite and earth-based (including airborne and maritime) observations.

The goal is to accomplish:

- fast response to where we should look next and with what technology (e.g., sensing apparatus), including human-engaged site visits
- elimination of potential targets that are either low-probability or redundant in focus
- elimination of false positives and false negatives!
- classification and prioritization of risks for significant epidemiological threats
- establishing a comprehensive and “universally acceptable” visualization including mapping, for use by different types of observers (in the medical and environmental fields, and including non-specialists in various responsible agencies and community centers), to understand in near-real-time the situation “on the ground” with respect to potential pathogenic threats.

Contact

Dr. Martin Joseph Dudziak
TETRAD Institute of Complex System Dynamics
+1 (231) 492-8301
martinjoseph@tdyn.org martin.dudziak@gmail.com www.tdyn.org www.intelrenaissance.com

www.emme.tdyn.org