Integrating genetic and environmental data to model and forecast movement and habitat use in the major insect vector of sleeping sickness in Uganda (Glossina fuscipes fuscipes)

Abstract: Tsetse flies (genus Glossina) are the obligate vectors of the trypanosome parasite that cause animal nagana and human sleeping sickness. One of the most effective strategies in controlling these dangerous and costly diseases is through vector control. Establishing feasible programs that reduce on-the-ground disease risk require knowledge of vector movement and habitat use. We use a novel machine learning strategy based on genetic data, field records, and remotely-sensed environmental data to model and predict both vector movement and habitat use across the landscape. The final output can identify areas (i) with the highest disease risk and greatest need for medical infrastructure, (ii) with marginal habitat that can be controlled at low cost but also require dedicated monitoring to prevent re-colonization, and (iii) likely to harbor isolated populations that can be effectively eradicated and/or use in the development of novel vector control strategies. I will also discuss the possibility of using output to parameterize mathematical models of disease transmission, and ultimately optimize best placement of tsetse fly traps and cattle treatment stations for eradication of nagana and sleeping sickness. To our knowledge, this is the first application of machine learning to integrate genetic data, field records, and environmental data to predict these important disease transmission parameters, and represents an important step towards improved strategic planning of vector control efforts to reduce human and animal African trypanosomiasis.