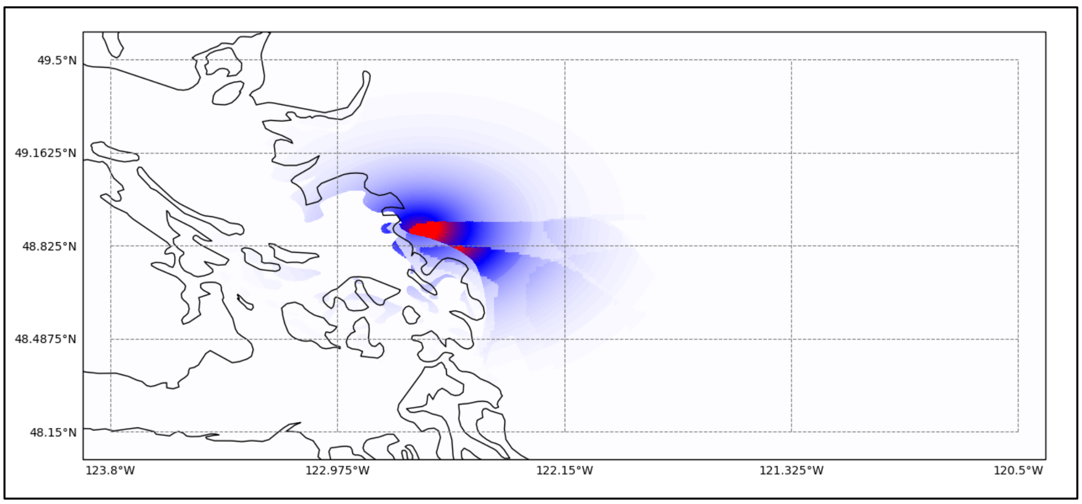
Problem Chosen	2021 MCM/ICM Summary Sheet	Team Control Number
C		2125707

A Buzz from Below: Tracking the Spread of the Invasive Asian Giant Hornet

Our paper seeks to facilitate the tracking and elimination of the Asian Giant Wasp in the State of Washington. By using traits characteristic of the species, such as nesting habits and environmental preferences, we detail a model that uses GPS coordinates and public sightings of the wasp to estimate the potential nest locations efficiently.

The process includes several steps: 1/ Locate positively identified reports and use them to create zones of frequent sightings; 2/ Estimate possible nesting locations and use exponential functions to track the expansion of V. Mandarinia; 3/ Crosscheck estimated nesting locations with geographic features and nesting habits to produce a heat map of noteworthy areas; 4/ Alert enforcement staff of new, unverified reports coming from high-risk areas to facilitate the prioritization of further investigation; 5/ Produce yearly heat maps to allow for a quick and easy visual identification of migration trends and population growth. The model is also robust enough to accept new sightings and recalculate estimations on the fly.

Testing revealed the developed model to be functional and capable of meeting the goals set out by the problem statement. However, several areas of improvement were discovered including an inherent dependence on large databanks of confirmed sightings and geographical features. Finally, we address several avenues of refinement to further expand our model's capabilities and ease of use, including social media or cell app integration, and AI image recognition.



An Estimated Projection of V.Mandarinia Expansion In the State of Washington Output by Our Model.

Table of Contents

A Buzz from Below: Tracking the Spread of the Invasive Asian Giant Hornet

Intro

In 2019, a highly aggressive Asiatic wasp known as Vespa Mandarinia, or Giant Asian Wasp, was discovered on the isle of Victoria, off mainland British Colombia. Soon further reports of the wasp were cropping up in the near-by state of Washington despite eradicating the initial nest. This was cause for concern as the species poses a significant risk to local pollinators such as the Honey Bee, while also being a potential health risk to humans. Furthermore, the wasps are difficult to track due to subterranean nesting and populations are hearty enough to survive the northern winters. The goal of this report is to examine the characteristics of V.Mandarinia and construct a predictive model capable of using reports and sightings from the public to estimate wasp population growth and probable nest locations. The model will also attempt to map potential high-risk areas in the State and use said data to prioritize incoming reports and flag problematic sightings for immediate investigation, while also recording and plotting easily readable heat maps for tracking yearly growth trends.

Background: Characteristics of Vespa Mandarinia

"Know your enemy, know his sword…"
Miyamoto Musashi

Despite having been catalogued for over a hundred and fifty years (first recorded by Fredrick Smith in 1852 (Skvarla, 2020)) there is surprisingly little detailed information to be found on Vespa Mandarinia in English, mainly due to it being an Asiatic insect. Additionally, many in-depth studies are very old and not readily available through electronic mediums. However some essentials such as their organizational structure and nesting cycles are known and available to us.

Appearance

V.Mandarinia, commonly known as the "Asian Giant Hornet" or "Sparrow Wasp" comes by its name honestly as it is the largest known hornet in the world. In general, adult hornets are approximately 1.38 to 2.17 inches long with a wing span of 1.38 to 2.99 inches wide with females being noticeably larger than males, and queens being larger still (Barth, Kearns, & Wason, 2013).

Koide, Yasunori; *Male of Japanese giant hornet, Vespa Mandarinia japonica*, Wikimedia *Commons, 2017. https://en.wikipedia.org/wiki/File:Male_Vespa_mandarinia.DSC_151*

The wasps have 6 legs, and two types of eyes – a large pair of forward facing compound eyes on either side of their head, and a clustered trio of simple eyes near the top of their head. Coloration is distinct and the same regardless of gender: a solid vibrant yellow/orange head, a solid black/dark brown thorax ("chest" section), and a banded abdomen of alternating yellow and black (Skvarla, 2020).

All members of V.Mandarinia have large, powerful mandibles used for hunting, masticating prey, digging subterranean nests, and for defence. Only females present with the powerful quarter-inch stinger, however, and are therefore the only members of the species who produce the extremely potent venom the hornet is known for (Barth, Kearns, & Wason, 2013).

Stings are exceptionally painful, with repeated stings from several hornets having been known to kill full grown adult humans. In fact, during a four month stretch in 2013, an estimate 40 people died in China alone (Gill, Jack, & Luck, 2020). Some victims die from allergic reactions, but many die from multiple organ failure due to a large number of repeated stings, usually caused by stumbling into a nest.

Organizational Structure

V.Mandarinia is a eusocial insect similar to bees, or ants (Barth, Kearns, & Wason, 2013). Eusociality is a societal structure characterized by cooperative brood care, and labor division along reproductive lines which has been referred to as a "caste" structure. In plainer terms, this is what most people typically think of when they envision a colony of bees: A fertile queen, whose job is to lay eggs, and a sterile caste of workers/soldiers who gather resources, tend to the young, expand the hive, and defend the nest.

Camazine, Scott; *Vespa Mandarinia processing its honey bee prey capture*. http://www.scottcamazine.com/

The Asian Giant Hornet is very similar in this regard: queens produce offspring, while the sterile members hunt, forage, and construct. Near the end of the year (around September), a 3rd caste starts to appear, that being the fertile males whose job it is to leave the nest and mate with near-by queens of other nests (Gill, Jack, & Luck, 2020).

Predation

Vespa Mandarinia are omnivorous hunters with no known predators other than humans, the occasional opportunistic Honey Buzzard, and rarely Mandarinia from other hives (Barth, Kearns, & Wason, 2013), effectively making them apex predators.

Common prey of V.Mandarinia are other hornet species, small insects such as beetles, and most distressingly for humans – Honey Bees of all varieties (Barth, Kearns, & Wason, 2013). During the majority of the year, the wasps feed on various insects, sap (particularly oak sap), and occasionally even fruit.

However, as the breeding season approaches and the Mandarinia hives are at full strength, what has been referred to as “Honey Bee Slaughters” begin to occur (Skvarla, 2020). One of these slaughters can see the eradication of an entire 30,000 member bee colony over the course of several hours (Gill, Jack, & Luck, 2020). Once the colony can no longer defend itself, V.Mandarinia will occupy the hive and retrieve all honey bee larva and pupae to feed its own young, while leaving many of the adult corpses untouched until later. Over the course of the next 2 weeks, the wasps consume the entire honey bee brood (Skvarla, 2020).

<u>Nesting Practices</u>

The Asian Giant Hornet is primarily a subterranean, forest dwelling species with a penchant for forests on the outskirts of mountain ranges. The vast majority of nests found were discovered in and amongst the roots of pine trees (Matsuura & Sakagami, 1973). Given their subterranean nature, soil surrounding a nest site must be well drained, and favorable for digging, meaning a copse of trees is preferred to a sandy beach or a rocky outcropping.

Kashiwagi; *Excavated Giant Hornet Nest*, Personal Blog, 2017. https://nittoboueki.com/blog/post-

Campbell, Andy; *Paper Wasp on August 8th 2006 in Umpqua National Forest in Western Oregon, USA*, Wikimedia Commons, 2006. https://en.wikipedia.org/wiki/File:Paper-Wasp-Nest.jpg

V.Mandarinia is also an opportunist, meaning queens will often inhabit the abandoned burrows of small animals, such as rodents or snakes (Skvarla, 2020). Nests may also present in tree hollows, but are rarely (if ever) found built higher than 6 feet above the ground (Skvarla, 2020). Urban areas have also been known to host nests, but this is comparatively rare and, when found, tend to be located in/near-to green spaces - both public and private (Azmy, Hosaka, & Numata, 2016). With this in mind, it makes sense then that V.Mandarinia is rarely found in areas devoid of trees, such as plains or the higher elevations of mountains.

The nests themselves are formed from paper-like combs created by the worker caste. Compared to other species of hornet, these nests are surprisingly rudimentary, lacking the more extensive outer structure of something like the Paper Wasp. It’s believed that this is due once again to the subterranean nature of V.Mandarinia where the surrounding soil and roots provide the structure and protection required (Matsuura & Sakagami, 1973).

<u>Nesting Cycle</u>

Of particular interest to us in our attempt to model the potential growth of V.Mandarinia is their yearly breeding cycle, as it greatly affects our forecasting efforts. The cycle begins with a fertilized queen awakening from its winter hibernation, called “overwintering”. The queen then sets out to find a suitable nesting site. Once a site has been found, the queen begins creating a rudimentary nest. This small nest is temporary, and will be expanded as the hive grows. With the temporary nest complete, the queen lays her initial brood of workers comprised of about 40 eggs (Matsuura & Sakagami, 1973).

During this time, the queen cares for and protects the brood while also foraging for food, mainly tree sap (Matsuura & Sakagami, 1973). This period lasts until the first brood has reached adulthood, about 40 days. Once the first workers mature, the queen essentially remains in the nest for the remainder of the season (Wikipedia, 2021), laying the eggs that will expand the hive. Given the short lifespan of sterile workers (approx. 3-5months (Barth, Kearns, & Wason, 2013)) in addition to losses from other sources, the queen has to lay an enormous amount of eggs, thousands over the course of the year. Hive growth continues throughout the season with peak hive capacity being reached around August.

Come September however, a change in the laying begins as the queen starts to produce fertile males, and the next generation of queens (Skvarla, 2020). Roughly 40 days later, usually near the end of October or the beginning of November, the fertile males and the new batch of queens begin their exodus from the nest. Males travel to near-by nests and wait for other queens to exit, males then attempt to mate with the said queens. Queens frequently fight this attempt however and flee if successful, resulting in a surprisingly low number of fertilizations. It's estimated that up to 65% of all queens remain unfertilized come the end of the breeding period (Skvarla, 2020).

Both fertilized and unfertilized queens then fly up to 30 kilometers away from their nest of origin in an attempt to find a safe location to overwinter. Once a site is found, the queens begin hibernation. The remainder of the nest, including the now year-old founding queen, begin to die off as winter arrives; the dissolution of the hive soon follows (Matsuura & Sakagami, 1973).

Furthermore, old nest sites are not re-inhabited, and so the hive goes extinct. During the course of the overwintering, fertilized queens begin to develop their reproductive organs, while unfertilized queens remain undeveloped. These undeveloped queens do not contribute to the next nesting cycle because of this, and disappear sometime around July (Matsuura & Sakagami, 1973).

The Problem

After having reviewed some of the characteristics of Vespa Mandarinia it becomes immediately apparent why an uncontrolled invasive expansion of this creature could potentially be devastating to both humans and native wildlife in North America.

From an ecological perspective, the wasps are highly predatory, extremely efficient hunters, with no known threats, that attack and kill pollinators such as honey bees, and wipe out entire colonies – A particular concern given the decline in honey bee populations seen in recent years and the importance of pollinators to the ecosystem.

Furthermore, from the human perspective, the venom carried in their stingers is extremely potent and more than capable of killing an adult human with sufficient attackers, let alone a child or infant. V.Mandarinia is also difficult to track, the nests are hidden and hard to locate, and population spread is resistant to frigid American/Canadian winters despite being a foreign tropical insect due to overwintering.

With that in mind, the best course is to eliminate the invasive species before it takes hold locally. Resources to manage this problem are not unlimited however, and so predictive modeling is our best chance to ensure the optimal use of available manpower and resources, while also highlighting and allowing for control of potential hot spots before they become problem areas.

Assumptions

Given that the introduction of a new invasive species is filled with unknowns, particularly with a species such as Vespa Mandarinia, we will have to make several assumptions:

1. We assume that distribution patterns are similar in range and scope to other related Vespa species (such as V.Velutina) with more detailed documentation and research.
2. We assume that the characteristics and habits of V.Mandarinia in North America are the same as those seen in populations elsewhere in the world.
3. We assume that growth and decay patterns are roughly exponential in nature within assumed migratory ranges.
4. We assume the location of nesting sites is directly correlated to population growth. i.e.: areas settled by the wasp continue to be inhabited by future generations – meaning nests in new previously uninhabited areas constitute species expansion and not simply migration.
5. We assume that the maximum recorded flight range of queens and workers represents the true limit of their range
6. We assume that all confirmed sightings are workers and not migrating queens.
7. We assume that when new queens migrate from their origin nest that the distance from this nest is a more powerful nesting factor than geographic location with the exception of obviously bad locations such as the middle of a lake.

Model

Criteria

Our model must satisfy several criteria as made evident by the problem description:

1. The model must accurately track known and estimated nest locations while providing an accurate forecast of future locations using known Vespa Mandarinia characteristics.
2. The model should output yearly and seasonal data in an easily readable form to allow for quick observation of population growth and spread
3. The model must be able to calculate an increase/decrease in the probability of finding a nest in a given area based on sightings by the public.
4. The model must allow investigators to accurately distinguish between high and low priority sightings to enable efficient allocation of enforcement resources.

Influencing Factors

1. V.Mandarinia tendencies and nesting preferences
2. Geographical features
3. Urban centers

Construction Considerations and Process:

Several factors pertaining to V.Mandarinia made creating an accurate propagation model difficult. Firstly, there is a distinct lack of migratory data to be found on the wasp. Estimated flight ranges and preferred geographical characteristics are known, however dispersion patterns, average distance between new nest, queen overwinter die off, average total fertilized queens per season, etc. are all virtually unknown making it difficult to ascribe a specific numerical model to predict growth and dispersion patterns. We therefore believe that an effective model will focus on predicting nest locations using public sightings. Given queens act in a predictable manner, estimated nest locations will then be used to forecast growth and species spread.

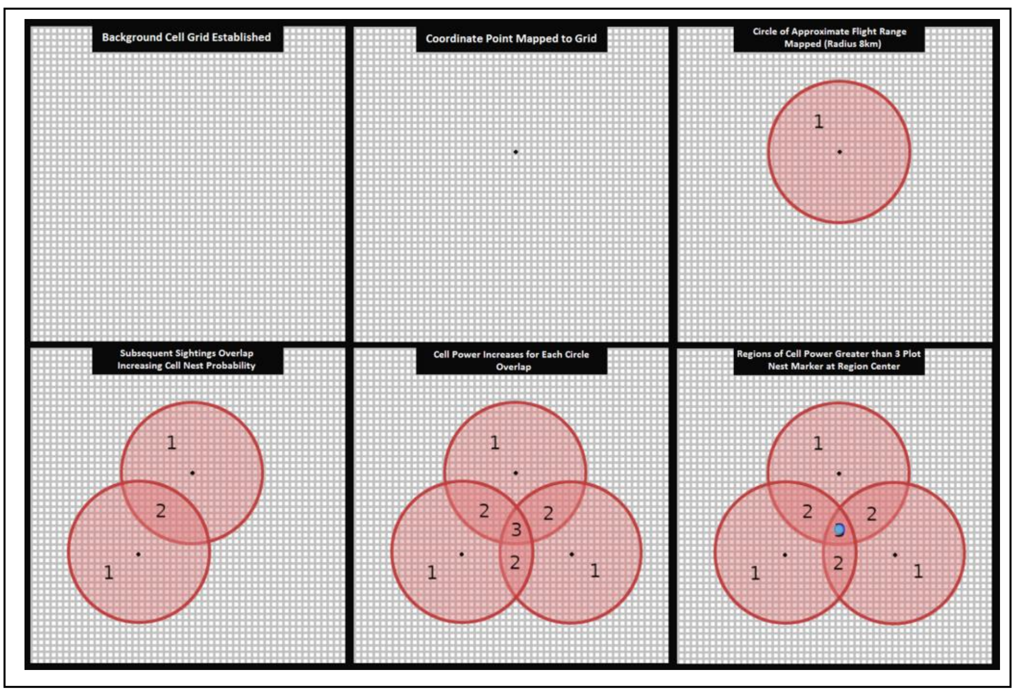
According to the information packets provided by MCM, the maximum range queens have been found to migrate from their nest of origin is approximately 30km (COMAP, 2021). Since each fertilized queen migrates to create a new nest only once in their year-long lives, we can assume the maximum yearly propagation of V.Mandarinia is roughly 30km/year.

To verify this assumption, we turn to a study of the distribution of a related wasp species, V.Velutina, through South Korea, which estimates yearly expansion to be approximately 10-20km/ year though urban centers (Choi, Martin, & Lee, 2012). The paper goes on to point out that though the home region of V.Velutina is tropical/subtropical, the surrounding South Korean land and nesting biology are similar to the wasp's home territory, meaning conditions for rapid spread are supported locally.

This is similar to the situation currently experienced in the state of Washington where, despite not being of tropical climate, the local geography is favorable for the rapid growth of V.Mandarinia (Zhu, Illan, Looney, & Crowder, 2020).This tells us then that it is reasonable to assume that the 30km max migration is possible in Washington, and that we can expect patterns to be similar to those found in Asiatic countries.

Drawing from previous forecasts of V.Mandarinia growth across Washington (Zhu, Illan, Looney, & Crowder, 2020) we chose to adopt a similar exponential potential distribution across the 30km migration radius, with probability being higher as one approaches the coordinates of origin. Mandarinia do not re-inhabit old nests however, meaning x=/=0.

Secondly, recorded sightings are typically workers and not queens, as queens tend to leave the nest only during a narrow window during October/November to mate before disappearing again to overwinter. Furthermore, worker sightings say very little about the precise location of a nest unless they're caught, tagged, and followed back home (a technique used to locate a nest in Blaine Washington (Segarra, 2020)). Workers also have a comparatively large roaming radius around their nest, typically 1-2km but they have been known to range as far as 8km (Skvarla, 2020). That means a sighting doesn't necessarily mean there is a nest in the immediate area.



Once again, nest sightings are relatively rare due to being underground. We therefore needed a method that allowed us to estimate the location of possible nest. Our idea was to center a circle the same radius as the maximum range of a typical Mandarinia worker on each confirmed sighting. Overlapping circles would then indicate areas common to all sightings, allowing for more accurate estimations of potential nests. In order to plot this overlap without the need for complex integral, we decided to begin by mapping a cell grid over the State of Washington. Each cell was then able to record how many circles were placed above it – subsequently increasing the cell value for each circle present. Cell size was small (approximately .1 kilometer on a side), meaning the approximation of the cell area deemed to be sufficiently accurate.

We decided that cells under a sighting circle should be assigned a value of 1 for each circle overlapping them. For instance, should a cell have two sighting circles overlap, the cell is assigned a value of 2 – should three sightings overlap – a value of 3 and so on. Once a common area is valued at 3 or higher, the model calculates the approximate center of the common area and a nest marker is placed. Further overlapping sightings refine the estimation, by either shrinking or enlarging the estimated area.

Sighting circles simply narrow down the possible area a nest could be found in, but in order to produce a probability heat map for tracking larger yearly trends, we settled on the idea of the Nest Markers. Nest Markers are estimations of migratory ranges based on the 30 kilometer flight range of a fertilized queen after mating. In order to apply a probability function effectively, we assumed that all fertilized queens coming from a given hive will establish a nest within the 30km radius circle, we did not take into consideration die off during overwintering as such information is not available. We also assumed that hive distribution was roughly exponential in nature and decided to implement an exponential probability function to be applied over the radius of the Nest Marker circle. Once again, overlapping Nest Markers resulted in increased nest probability, resulting in a heat map that estimated the location of nests in the coming year. Once again, isolated sightings don't directly contribute to the probability of a nest in the coming year as sighting circler are not directly linked to new nest probability, but rather current nest location. We then decided that due to the nesting cycle, yearly heat maps could be carried over to subsequent years in order to influence the initial probability of nests in the new year. This has the added benefit of providing immediate historic data to analyze potential growth trends.

To reduce the chance of having a large number of nest markers in one area should many sightings be reported in a small region, multiple high-value common areas existing within a circular area less than r = 8km, would be combined into a single hive marker centered in the highest value common area. To help prioritize investigation and enforcement, we decided that new sightings mapped into a region with high probability or high sighting frequency are to be immediately flagged as priority without having to analyze the location manually.

We then sought a way to refine the otherwise rudimentary exponential distribution. We settled on the idea of using geographic data to create a coefficient that modifies the initial Nest Marker probability field. Our research had revealed that V.Mandarinia queens are surprisingly choosey about where they locate their nests, and clear preferences were apparent. This also provided a way to eliminate regions that were impossible for a wasp to settle, such as lakes, or high mountains. We believe that is of particular use in a state such as Washington, where terrain type varies greatly depending on where you are. With high plains in the east, thick forest low lands in the west, and tall mountains running throughout, a blanket approach to probability was thought to be inaccurate. To this end, reports were analyzed and paired research from Asia to develop an initial Nesting Coefficient.

$$Nesting\ Coefficient = \frac{1 + (Trees) + (Drained\ Soil) + (Mountain\ Outskirts) + (2 * Pine\ Trees) + (2 * Oak\ Trees) + (2 * Urban\ Grn\ Spc)}{10 * (1 + 2 * Open\ Water)^4 * (1 + 2 * High\ Altitude)^4 * (1 + Plains)^2 * (1 + High\ Ground\ Water)^2 * (1 + Urban)}$$

Values in the numerator are desirable nesting factors based upon previously found nests in Asia and North America. For instance Matsuura & Sakagami state that of the “Among 31 (V.Mandarinia) nests examined, 25 were found in subterranean cavities formed around rotten pine roots or those made by small rodents, snakes, etc., and six in subterranean tree hollows”, other sources report similar findings (Skvarla, 2020). As such we can ascribe a positive value to regions in which desirable nesting factors are present. Multiple desirable factors increase the probability of nests in the area and therefore increase the likelihood of V.Mandarinia sightings.

Similarly, V.Mandarinia is rarely (if ever) found in areas with factors present in the denominator - indicating a strong dislike for regions with those traits. For instance, burrowing in regions with a high water table, or large areas of standing water are difficult to nest in and have a high probability of flooding their subterranean dwellings.

Factors are binary where 0 indicates the factor isn’t present while 1 indicates its present and of sufficient quality/quantity to impact the likelihood of finding a nest.

Given that V.Mandarinia is extremely rare in areas with negative factors, negative factors should therefore outweigh the positive; as such a greater weight is ascribed to them. Strongly negative regions result in a fractional coefficient which in turn decreases the probability of finding V.Mandarinia in the area.

After experimenting with the developed coefficient, it was decided to make it a scaling factor by dividing the function by the maximum positive Nesting Coefficient possible: 10. This gave us a function that output a number from (0,1]. Effectively we now had a number that served to reduces the exponential probability distribution of the Nest Markers. We assumed that the exponential distribution was the ideal. For example: an ideal location would have a nesting coefficient of 1, while an extremely unlikely location, such as the middle of a lake or the top of a mountain would have coefficient of 1/810, effectively zeroing out the probability. We also assumed that distance from the origin nest was a more powerful factor than geographic features; given queens don’t seem to migrate further than the 30km maximum. Highly desirable coefficients therefore do not modify the exponential probability at all, while undesirable qualities reduce it, in some cases bringing it near to zero.

We now needed a way draw the values required for the nesting factor. We did not want staff to have to input more data than just the location of a sighting. Therefore a map with a rough location of each individual term was created. This map was extremely rudimentary, drawn by

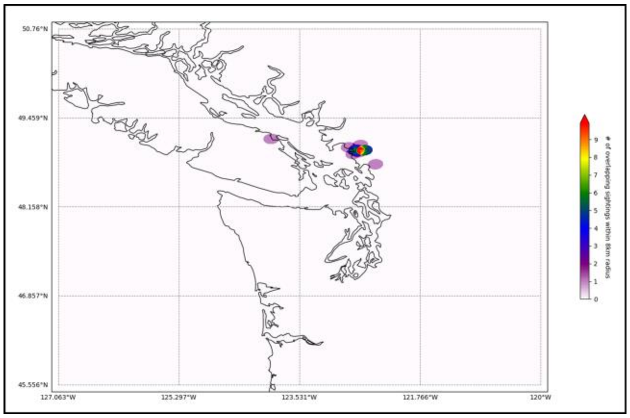
Example of “Forest” reference map.

hand while consulting Google Maps and State Forestry maps. It is simply meant to serve as a proof of concept.

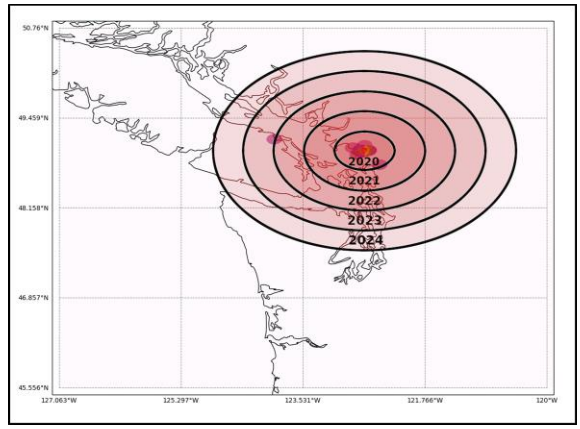
In order to make use of the map we created, a program was developed that used latitude and longitude to link a location to a vectorized version of the hand drawn map. If the area was shaded black on the vector map, the model would return a value of 1 for the associated term, indicating the factor existed at the location. If the location was in a white or grey area, the function returned 0, cancelling the term in the equation. Maps were created for every term in the nesting factor equation. Regions frequently had multiple factors present at a time, meaning the nesting factor acted as intended and provided a combination of desirable and undesirable factors to estimate the approximate desirability of a location.

Having the ability to track the progress towards eradication was also important. Therefore we decided to implement a time function that reduced the probability of extant nests if an area had no sightings several years in a row. After some experimentation, we settled on a function of exp(-t/2). This time functions reduces the sighting probability as the time between sightings increases. For instance, if a region has no sightings for 2 consecutive years, it is assumed that the probability of a new nest existing in that region will go down. As such, a region with yearly sightings would have no reduction in baseline probability, while a region that hasn’t had a confirmed sighting in several years would have its baseline probability returned to near zero.

Implementation, Testing, and Results

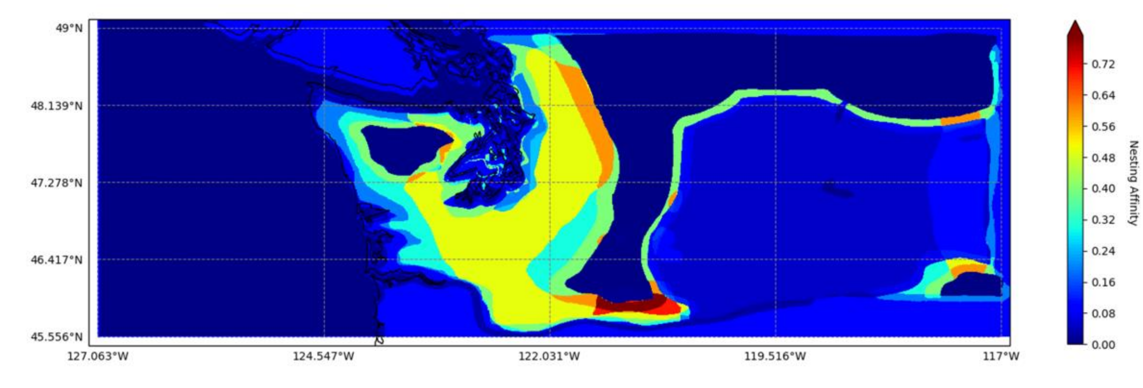


Confirmed Sightings Mapped along with 8k worker flight radius.



Basic Migration Model Based on Maximum Queen Migration

With the basic structure and desired features implemented we began to develop the model. So that we had something to compare our more complex model to, we generated a very simple progression map. Firstly tested our sighting circle program and used it to approximate the center of confirmed sightings provided in the MCM data. Circles ended up being plotted as ellipses due to the scaling factor present in the map produced by the Cartopy program. From there, we centered a series of expanding circles based on the 30 km/year migration of newly fertilized queens, and then plotted the next 5 years. The assumption of the simple model is that after a year has passed the region inside the radius 30km circle will be colonized by V.Mandarinia. This model does not take into account any factors beyond the maximum seasonal migration.



Nesting Coefficient Heat Map - State of Washington

With a basic estimation in place we could begin with the implementation of our more complex model, beginning with the Nesting Factor.

In order to reduce the amount of time it took the program to run, the Nesting Factor was plotted on its own heat map using our vector map program, Cartopy, and Python.

This map contains all the pre-calculated coefficients for a given region so that rather than calculating the coefficient every time the program is called, it simply needs to reference the coefficient map once per nesting site.

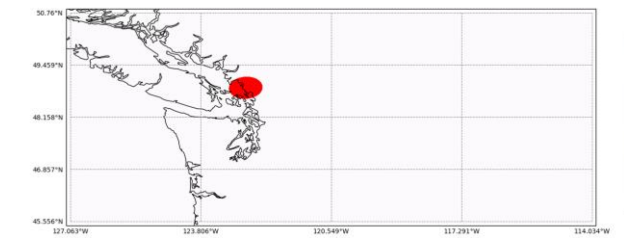
We now had a working program for sightings, and a reference map for the probability modifying Nesting Factor, it was now time to develop the Nest Marker program.

Using the sighting map produced for the basic model, we were able to locate the highest region of common sightings and used our program to plot a radius 30km circle centered at the center of the region. We then implemented the exponential decay function which when applied produced a probability heat map within the bounds of the circle.

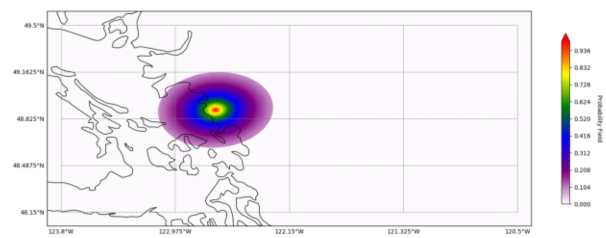
From there, we modified the basic heat map by consulting the Nesting Factor map. This reduced the probability in undesirable regions and allowed problem areas to be marked on the map output. Finally, plotted several years as a test of the time decay function and out initial model was complete.

By comparing the images we obtained from our model to the basic propagation model, several things become very clear. Firstly, we simply didn't have enough data. Our initial sightings plot only provided us with one estimated nest, so the probability field was weak compared to a field with multiple nests across a larger area.

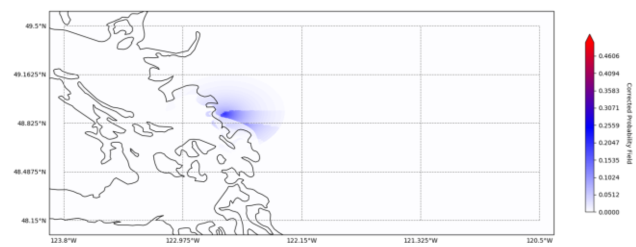
Secondly, the Nesting Factor needs to be refined - particularly through the use of



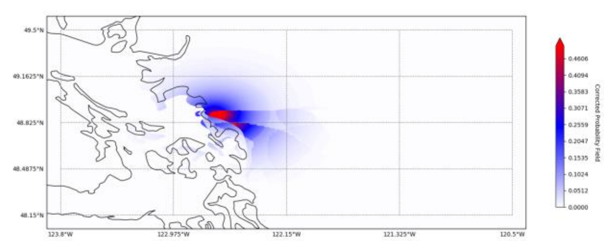
Nest Marker Plotted on Cartopy Map



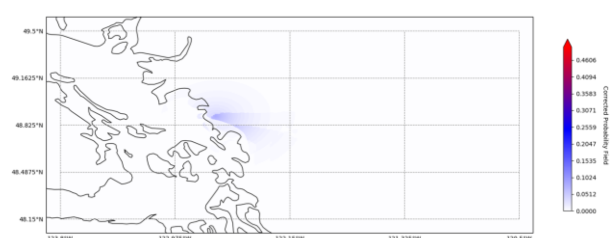
Zoomed in View of Probability Field



1 year Modified Probability Field – Probability Scale from 0 to .5



Year 4 with Time Decay Function Disabled



Year 4 with Time Decay Function Enabled

better geographic data. The rudimentary hand drawn plots were simply too coarse to allow for an accurate estimation of nesting desirability. Areas close to the shore were missed due to the nature of the editing program used and the associated vectorization of the image. However it was evident that the concept was workable and provided a viable option for improved tracking as the probability field was indeed modified inland toward forests and dry land, a move we would expect to see in real V.Mandarinia present in the State of Washington.

And finally, the time decay function is almost certainly too strong. We wouldn't expect the probability of a nest to decay that drastically over such a short period of time. However, with the data restrictions of the MCM problem, we simply don't have sufficient data over a large enough period of time to accurately refine the time dependence of the function.

We conclude then that our model is functional and capable of accomplishing the goals we set out to meet, however, more data and experimentation is required in order to reach its true potential. With sufficient refinement and application of the improvements listed in following sections, we believe that our model has the capability of providing a solid platform for future development.

Model Strengths

- Model doesn't require confirmed nest location or confirmed queen sightings to estimate species growth potential.
- Data input is relatively straight forward and only requires entering the approximate GPS coordinates of a given sighting
- Model can easily and efficiently track multi-year data due to yearly hive estimation heat maps.
- Model warns operator of high priority sighting at time of data entry based on prior sightings and historic data.

Model Weaknesses

- Geography types used in Nesting Coefficient are inaccurate as the regions were drawn by hand due to time constraints. Better geographical and topological data would greatly increase the accuracy of the nesting coefficient.
- Model is dependent on the public to submit sighting reports – high public involvement is required to ensure a successful model.
- While multiple sightings in an area are indicative of a near-by nest, it is not guaranteed and may actually result in a larger/smaller number of estimated nests than are actually extant.
- Highly dependent on large volumes of data to accurately predict future trends and population growth.

Refining the Model

Image Recognition

One of the first things attempted in our initial model was to implement image recognition to assist in processing submitted photographs of potential sightings. We turned to TensorFlow a Python API that seemed to meet our requirements. Using TensorFlow we had hoped to identify V.Mandarinia in user submitted images. This could allow banks of images to be rapidly processed by providing a sort of coarse pre-filter, reducing number of images needing to be verified by hand from thousands to hundreds. The AI could even be trained to classify priority of investigation based on the percent match output during the image analysis.

Further development could see the introduction of a web based identification system that would allow the public to upload pictures to a database without having to rely on email for communication. Enforcement staff could then follow up as needed if submissions were unclear or if more information was needed. This would drastically reduce the clerical and administrative time currently spent by staff and allow for faster response times and more accurate identification.

Unfortunately, training the AI with the bank of images available in the MCM folder was simply too unreliable. The variety of shot angles, image sizes and qualities, meant that a reliable baseline was extremely difficult to achieve. Additionally, images of confirmed V.Mandarinia were vastly outnumbered by images of unconfirmed or negative cases, making the training process all but impossible within our time-line.

This avenue of exploration is still extremely promising however and if sufficient quality images were available for AI training then we are confident a reasonably accurate image recognition program could be developed and linked with our model.

Public Buy-In

As for refining the dependency on the public, there is no easy solution to that particular problem. Due to the lack of manpower and resources available to enforcement staff, public buy-in is essential to report sightings of V.Mandarinia in virtually all cases. It simply is not possible to have staff everywhere at once. But perhaps a social media presence through Facebook, Twitter, etc. would increase awareness of the need for information while reminding the public to be on the look-out. In that vein, it might be worth exploring the possibility of Twitter integration or the creation of an app so that people on the streets can assist in documentation using their cellphones cameras and inbuilt GPS devices.

More Complete Geographical Data

The geographical data used in the calculation of our nesting coefficient is extremely rudimentary. Effectively being just estimation using Google Maps coupled with local forestry and water maps - nothing more than a proof of concept. This is one of the areas the model stands to see the greatest benefit from refining. Survey data is readily available for most areas in the country as are the locations of ground water, lakes, rivers, altitude, and even forest composition and coverage. With a sufficiently detailed database for the nesting coefficient to draw from, the probability estimation could be refined considerably, greatly improving the accuracy of the predictions.

Tracking V.Mandarinia Movements

Another area that could see great improvement from refining would be the model's nest marker itself and the associated probability circle. As previously stated, there is very little data on the movements of V.Mandarinia. We know general ranges and can make some educated guesses on how they might move within their home area. But a concerted effort to track precise patterns would be highly beneficial.

Should the model prove effective, it may be worth attempting to tag some members of a hive with trackers. This has already been done on a rudimentary level to find the nest in Blaine, Washington (a tracker and tooth floss) (Segarra, 2020). It stands to reason that a more refined tracking method would yield even more impressive results.

With a clearer picture of how V.Mandarinia moves it would be much easier to develop a clear estimation of migratory ranges and patterns, in addition to being better able to track the movements of queens and where they prefer to overwinter.

Perhaps rather than attempting to eradicate hives after they've been established it would be more efficient to locate overwintering queens and destroy them. Without a clearer picture of their movements however, this will remain unknown.

The Priority Model

A second 'priority' model was also discussed and partially implemented. In this model, the data given by MCM was first moved to a separate file containing only the GlobalID, Detection Date, Lab Status, and Latitude-Longitude of the sighting. Initially an Excel file, it was later converted to csv for easier compatibility with Python.

Once on Python, the file was read using Python's csv library. The data from the file was stored using an array of Location objects that included all the characteristics of each individual sighting. This allocation made it possible to create functions that used this data to create a list of all positive, negative and unconfirmed sightings.

Once the positive locations were extracted, every positive starting location was assigned a 'priority' of one. If the location was within 8km of another location (the maximum range of the workers), then the location's priority would increase by one. This was repeated for every location. Then, a map was made based on the Cartopy library, as well as Matplotlib. These libraries gave powerful tools to map both the locations of the sightings as well as accurate boundaries for the specific zone affected.

The locations on the map would also be modified based on the time of the reported sighting, so that the map could be changed based on the year and the season. The priority model would have given an estimate of how V.Mandarinia spreads over time in the area, as well as a metric to easily test new reporting against existing data.

This model was not possible to implement because of multiple errors using Cartopy, as the software is still on its development stage, as well as a lack of documentation. The program was also not possible to debug easily due to the time constraints, and even though a map was produced, the data on it was not accurately plotted.

A Word Cloud Analysis of Language used in Citizen Reports

An important component in determining the reliability of a reported sighting is the attached citizen comment. A word cloud analysis of the comments highlights a pattern: the submitter usually details the situation in which the hornet is detected. This helps us gain a general understanding of what submission merits further investigation

Word Cloud of Comments with Positive ID (left) and Negative ID (right) status

The negative IDs comment usually employs more uncertain and speculative language, such as "like," "huge," and "giant." The positive IDs are from more official confirmation, mentioning WDSA and citizen report. Due to the lack of positive IDs, not much information is drawn here but to refine the system of prioritizing sighting reports, it is worthwhile to look into training Machine Learning to figure out reliable reported sightings.

For some of the unverified IDs, the most telling information is from the Lab Comments where the entomology lab would comment on whether or not the attached image depicts the Asian Giant Hornet. As such, it is reasonable to immediately decrease the priority of investigation of such reports.

Conclusion

The goal of our paper was to facilitate the tracking and elimination of Vespa Mandarinia (the Asian Giant Hornet) in the State of Washington. Researching the invasive species and its characteristics, several exploitable weaknesses were found and a model based on public sightings was constructed.

After initial implementation, alterations to the model enabled it to consider geographic factors to better highlight potential problem areas bases on nesting desirability. Furthermore, the model used this information to alert enforcement staff of sightings in high risk/ high importance areas, while producing visual aids to help track migrations and population trends.

While the final model was fully functional, several areas of refinement were noted, including the application of better geographical data and larger data sets. Future improvements may also lay in social media integration or the development of an easy to use image recognition app to facilitate the gathering of quality data.

Memorandum

To: Washington State Department of Agriculture

Date: February 8, 2021

From: MCM Team #2125707

Subject: Tracking the Spread of the Invasive Asian Giant Hornet

Due to the natural characteristics of Vespa Mandarinia for thriving in the North American environment, where they have no natural predators and where they attack local bee populations, it is necessary to create a plan in order to eradicate the species from their non-native habitats. This plan has been partially implemented, as WSDA has already created a reporting website where anyone can report a sighting of V. Mandarinia. However, because of the amount of sighting reports submitted and the limited resources available, it is necessary to have a model that predicts the spread of V. Mandarinia and gives priority to those reports that have the highest chance of being positive.

Our Model

Taking into consideration the data provided, we created a model based on the existing positive sightings as well as the geography of the area, and made various heat maps that outline the possibilities of this pest expanding to other areas.

This model uses the existing research on V. Mandarinia's preferred environments, as this can be a very important factor in predicting where this species would nest and how they will spread over time. Based on this, we created various maps outlining basic geographic features that are relevant to this species, such as altitude, forestry, water bodies and soil type. With these geographic features in mind, a Nesting Preference formula was created, to calculate a heat map that prioritizes some specific features that are more important for this species, such as water bodies (since V. Mandarinia is not able to create nests under water) (Matsuura & Sakami, 1973).

Additional to the Nesting Preference map, we created another map based on the data given by the confirmed positive sightings. This new map generated an 8km circle centred at every confirmed sighting, based on the maximum distance travelled by a worker.

With this new map, we created a probability field that was centred at the highest number of sightings extracted from the positive sightings map. This probability field was based on the exponential decay function and the bounds of the maximum possible travel distance from queens.

Combining this information, we created a final map, based on the Nesting Preference map and the probability field that gives an idea of the most probable spots for V.Mandarinia to thrive. This map was also extended four years in the future, to make a rough prediction about possible migration patterns of the species inside the State of Washington as well as Canada. Without new confirmed sighting, an exponentially decaying time function is applied, thus reflecting the slowly eradication of the species from the area.

However, these maps are not definitive, and for it to be more precise, more data about positive sightings and migration patterns or timings is required.

Our Results

Referencing the Nesting Preference map created, the species would nest easily on the western part of the State of Washington, more easily on the southwest, and very easily on specific parts of the south west, such as areas nearby the Yakama Indian Reservation and the cities of Kennewick and Richland.

However, based on the current positive sightings and the probability map created, and assuming V. Mandarinia does not grow exponentially based on other factors such as travel and imports, this area is not very probable to be reached by the species because of the limited range of mobility of the queens.

Our Corrected Probability Field map predicts that in a 4-year period the species would continue expanding in the areas of Bellingham and in the surroundings of Mt. Baker, and in our worst-case scenario – including only the migration patterns of the species, and not the possible expansion due to human-based activity – the species would expand south to Seattle and its nearby areas, as well as the Olympic National Park.

Possible Improvements

There are various avenues for possible improvements to the model, including a more complete set of geographic data that could allow for more precise predictions of the species' possible nesting sites, as well as image recognition software that would allow for easy recognition of V. Mandarinia from the submitted pictures.

However, these avenues require more confirmed positive sightings, as the model falls short since it is not possible to exactly predict where the next nests of the species would be, and the image recognition requires more pictures of confirmed cases in order for the algorithm to be trained properly.

Conclusion

The spread of the Asian Giant Hornets can be predicted and investigated via citizen reports. Our model provides a reliable way to prioritize reported sightings and track species expansion and eradication.

Bibliography

Azmy, M. M., Hosaka, T., & Numata, S. (2016). Responses of Four Hornet Species to Levels of Urban Greenness in Nagoya city, Japan: Implications for ecosystem disservices of urban greenspaces. *Urban Forestry & Urban Greening*, 18,117-125.

Barth, Z., Kearns, T., & Wason, E. (2013). *Vespa Mandarinia*. Retrieved February 5, 2021, from Animal Diversity Web - University of Michigan: https://animaldiversity.org/accounts/Vespa_mandarinia/#62600107-3A4D-11E2-B90E-002500F14F28

Behnel, et al. (2011). Cython: The Best of Both Worlds. *Computing in Science and Engineering*, 13, 31-39.

Choi, M. B., Martin, S., & Lee, J. W. (2012). Distribution, spread, and impact of the invasive hornetVespa velutinain South Korea. *Journal of Asia-Pacific Entomology*, 15, 473-477.

COMAP. (2021). 2021 MCM Problem C: Confirming the Buzz about Hornets. COMAP.

Gill, C., Jack, C., & Luck, A. (2020, May). *Vespa Mandarinia*. Retrieved February 5, 2021, from Entomology Department University of Florida: http://entnemdept.ufl.edu/creatures/MISC/BEES/Vespa_mandarinia.html

Harris, e. a. (2020). Array Programming with NumPy. *Nature*, 858, 357-362.

Hunter, J. (2007). Matplotlib: A 2d Graphics Environment. *Computing in Science & Engineering*, 9, 90-95.

Matsuura, M., & Sakagami, S. (1973). *A Bionomic Sketch of the Giant Hornet, Vespa mandarinia, a Serious Pest for Japanese Apiculture (With 12 Text-figures and 5 Tables).* Sapporo: Hokkaido University.

McKinney, W. (2010). Data Sctructures for Statistical Computing in Python. *Proceedings of the 9th Pythin in Science Conference*, 51-56.

Met Office. (2015). *Cartopy: A cartographic Python library with a Matplotlib interface.* Retrieved February 5, 2021, from https://scitools.org.uk/cartopy/docs/latest/index.html

Meurer, e. a. (2017). SymPy: symbolic computing in Python. *PeerJ Computer Science*, 3e103.

Millman, K., & Aivazis, M. (2011). Python for Scientists and Engineers. *Computing in Science & Engineering*, 13, 9-12.

Nix, S. (2019, April 26). *Maps of Common United States Forest Cover Types*. Retrieved January 5, 2021, from Treehugger: https://www.treehugger.com/common-united-states-forest-cover-types-1343033

Oliphant, T. E. (2007). Python for Scientific Computing. *Computing in Science & Engineering*, 9, 10-20.

Pedregosa, et al. (2011). Scikit-learns: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.

Peterson, P. (2009). F2PY:a tool for connecting Fortran and Python programs. *International Journal of Computational Science and Enineering*, 4, 296-305.

Segarra, C. (2020, October 1). *A new map shows where Asian giant hornets could thrive in the U.S.* Retrieved February 5, 2021, from Science News: https://www.sciencenews.org/article/asian-giant-murder-hornets-new-map-habitat-united-states

Skvarla, M. (2020). *Asian Giant Hornet.* University Park: Pennsylvania State University.

U.S. Geological Survey (USGS). (2021, February 5). *Ground Water Levels for Washington*. Retrieved February 5, 2021, from National Water Information System: https://nwis.waterdata.usgs.gov/wa/nwis/gwlevels

Van der Valt, et al. (2014). scikit-image: Image Processing in Python. *PeerJ*, e253.

Virtanen, et al. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17(3), 261-272.

Wikipedia. (2021, February 3). *Asian Giant Hornet*. Retrieved February 5, 2021, from Wikipedia: https://en.wikipedia.org/wiki/Asian_giant_hornet

Wikipedia. (2021, January 14). *Quercus Garryana*. Retrieved February 5, 2021, from Wikipedia: https://en.wikipedia.org/wiki/Quercus_garryana

Zhu, G., Illan, J. G., Looney, C., & Crowder, D. W. (2020). Assessing the ecological niche and invasion potential of the Asian giant hornet. *Proceedings of the National Academy of Sciences of the United States of America*, 117, 24646-24648.