Weather-driven insect dispersal simulations using workflows that combine HTC and HPC resources

Matthew Garcia, Ph.D.

Postdoctoral Research Associate

Dept. of Forest & Wildlife Ecology

University of Wisconsin–Madison

matt.e.garcia@gmail.com

Principal Collaborators:

Brian R. Sturtevant, U.S. Forest Service

Jacques Régnière, Canadian Forest Service

Rémi Saint-Amant, Canadian Forest Service

Yan Boulanger, Canadian Forest Service

Philip A. Townsend, UW–Madison F&WE

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

CENTER FOR HIGH THROUGHPUT COMPUTING

Open Science Grid

FOREST SERVICE U S DEPARTMENT OF AGRICULTURE

Canada
Natural Resources Canada
Canadian Forest Service

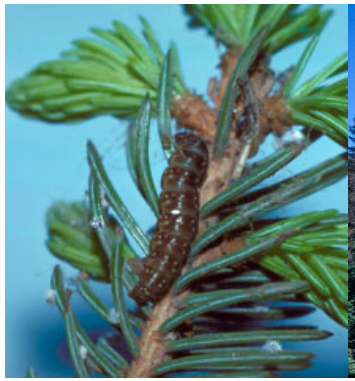
US-Canadian Forest Health Initiative

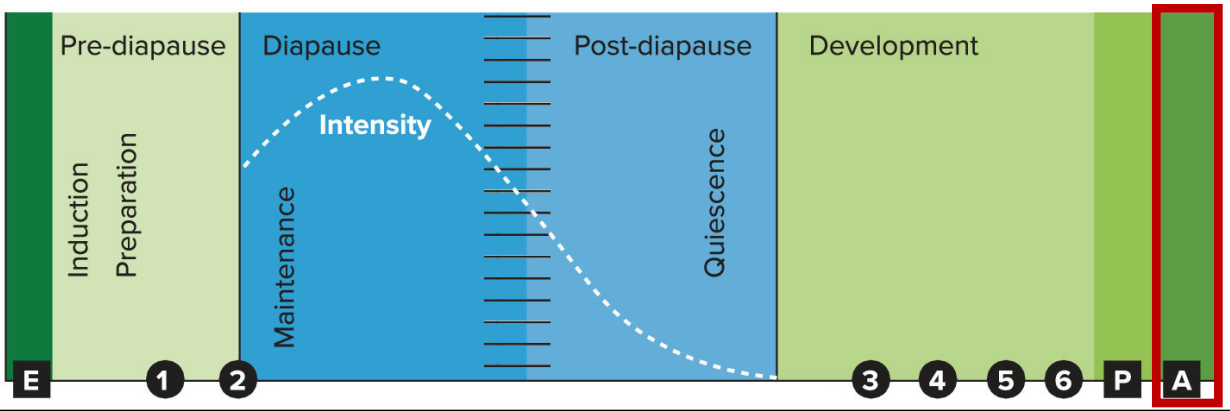
"One Continent, One Forest, One Threat"

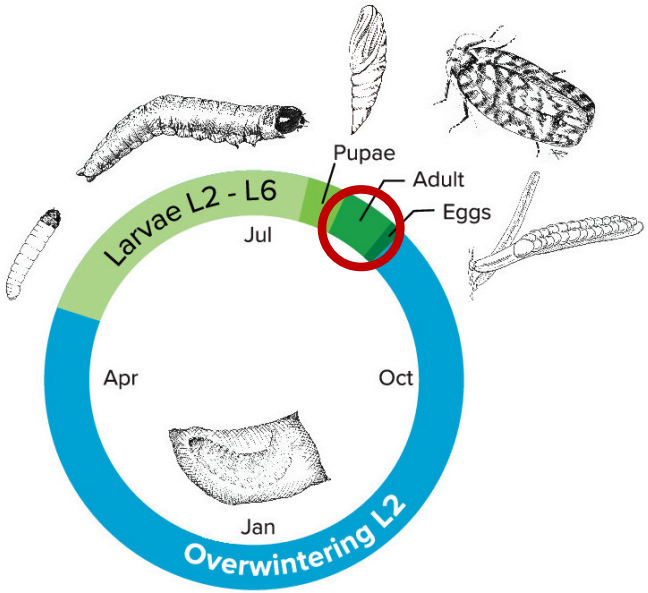
U.S. Endowment for Forestry and Communities

Multidisciplinary project: modeling dispersal of SBW moths

Eastern spruce budworm (*Choristoneura fumiferana* [Clem.])
life cycle, defoliation, aerial dispersal, oviposition
Current focus on Quebec, New Brunswick, Maine







Development cycle graphics adapted from Marshall and Roe (2021, *Physiology*)

Part 1: Model description (Jan 2022)

Agricultural and Forest Meteorology 315 (2022) 108815

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

journal homepage: www.elsevier.com/locate/agrformet

Modeling weather-driven long-distance dispersal of spruce budworm moths (*Choristoneura fumiferana*). Part 1: Model description

Matthew Garcia [a,*], Brian R. Sturtevant [b], Rémi Saint-Amant [c], Joseph J. Charney [d], Johanne Delisle [c], Yan Boulanger [c], Philip A. Townsend [a], Jacques Régnière [c]

[a] Department of Forest and Wildlife Ecology, University of Wisconsin–Madison, Madison, WI 53706, USA
[b] USDA–Forest Service, Northern Research Station, Rhinelander, WI 54501, USA
[c] Natural Resources Canada, Canadian Forest Service, Quebec G1V 4C7, Canada
[d] USDA–Forest Service, Northern Research Station, Lansing, MI 48910, USA

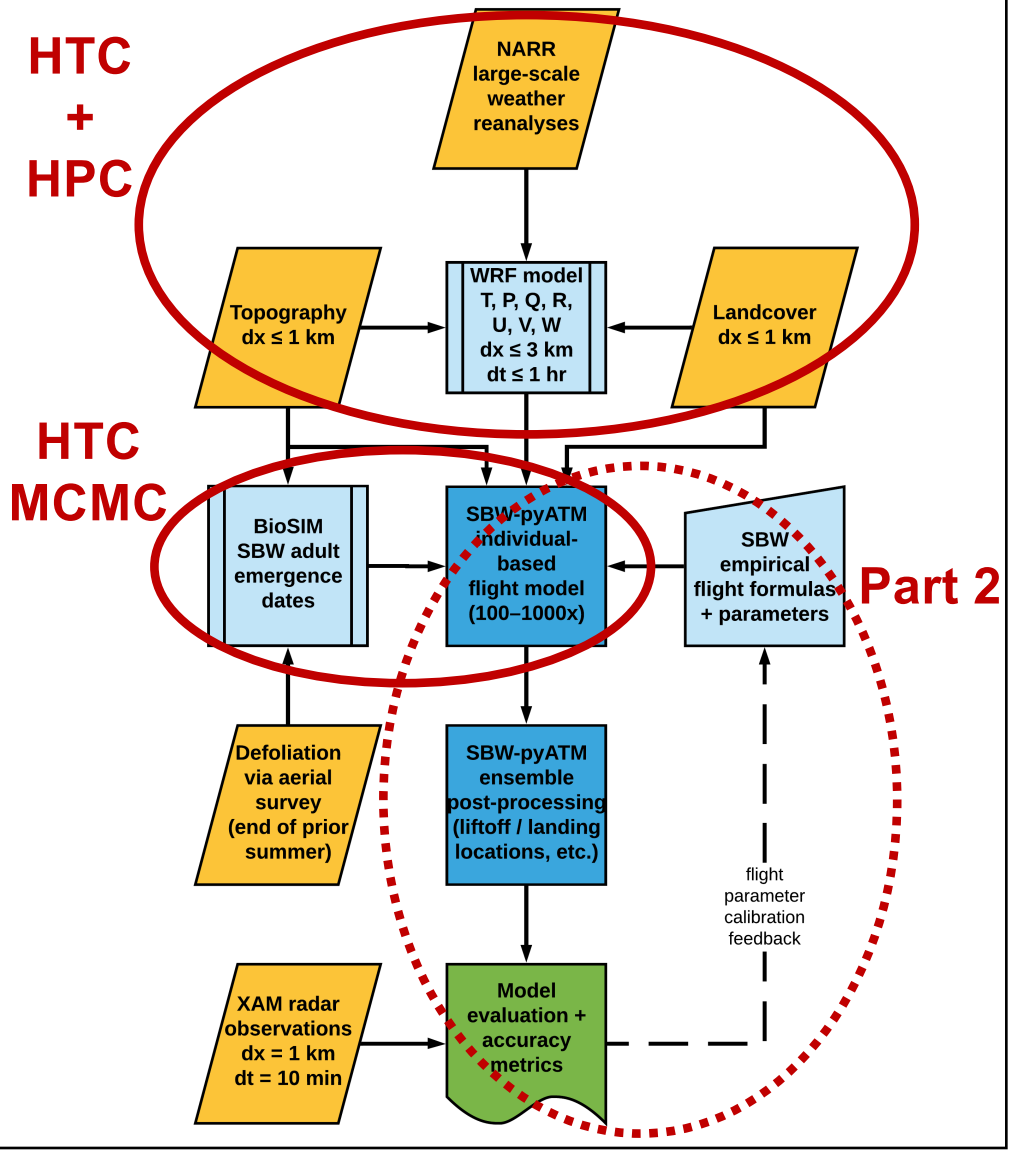
ARTICLE INFO

Keywords:
Aerobiology
Individual-based modeling
Insect outbreaks
Lepidoptera
Numerical modeling
Weather radar

ABSTRACT

Long-term studies of insect populations in the North American boreal forest have shown the vital importance of long-distance dispersal to the maintenance and expansion of insect outbreaks. In this work, we extend several concepts established previously in an empirically-based dispersal flight model with recent work on the physiology and behavior of the adult eastern spruce budworm (SBW) moth, *Choristoneura fumiferana* (Clemens). An outbreak of defoliating SBW in Quebec, ongoing since the mid-2000s, already covers millions of hectares of forests in eastern Canada and threatens to spread into neighboring areas through annual summertime episodes of long-distance dispersal. Such flight events in favorable conditions frequently include billions of SBW moths dispersing in the warm atmospheric boundary layer, typically starting around sunset and often lasting through several hours of wind-driven transport over hundreds of kilometers. Successful SBW dispersal to possibly distant host forest areas depends acutely on the weather. Here we describe the components and results of SBW–pyATM, an open-source individual-based modeling framework developed in Python for the simulation of these weather-driven SBW dispersal events. Using seasonal SBW phenology results from BioSIM at known outbreak locations and high-resolution Weather Research and Forecasting (WRF) model output, we focus on modeling dispersal flights over two successive nights in July 2013 in southern Quebec. Our flight model closely reproduces the SBW spatial patterns and motions observed by weather surveillance radar over the St. Lawrence estuary. With SBW–pyATM we can estimate landing locations for both male and female SBW and the resulting spatial patterns of egg distribution, allowing us eventually to forecast future larval defoliation activity in new locations where immigration could help overcome local limitations on SBW populations. This information could then support forest management decisions where SBW outbreaks threaten valuable resources.

doi: 10.1016/j.agrformet.2022.108815

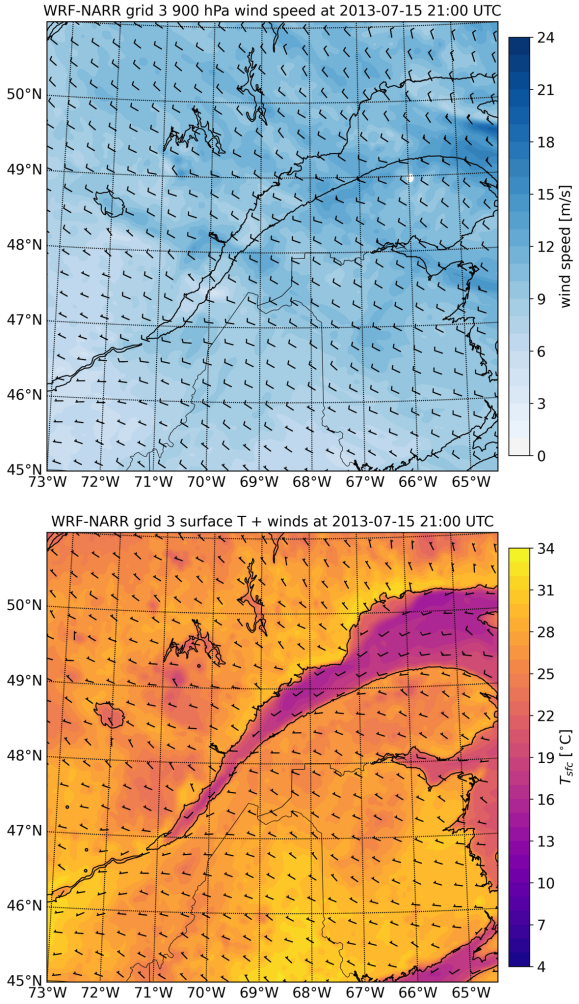


Weather Research & Forecasting (WRF) Model v4.1

NARR input: Δx = 32 km @ Δt = 3 hr
+ static fields (topography, LC, etc.)
→ WRF initial and boundary conditions

WRF model configuration: 3 nested grids
d01: Δx = 27 km, output @ Δt = 3 hr
d02: Δx = 9 km, output @ Δt = 1 hr
d03: Δx = 3 km, output @ Δt = 10 min

Post-processing data reduction
d02 @ Δx = 9 km → synoptic analyses
d03 @ Δx = 3 km → SBW flight simulations



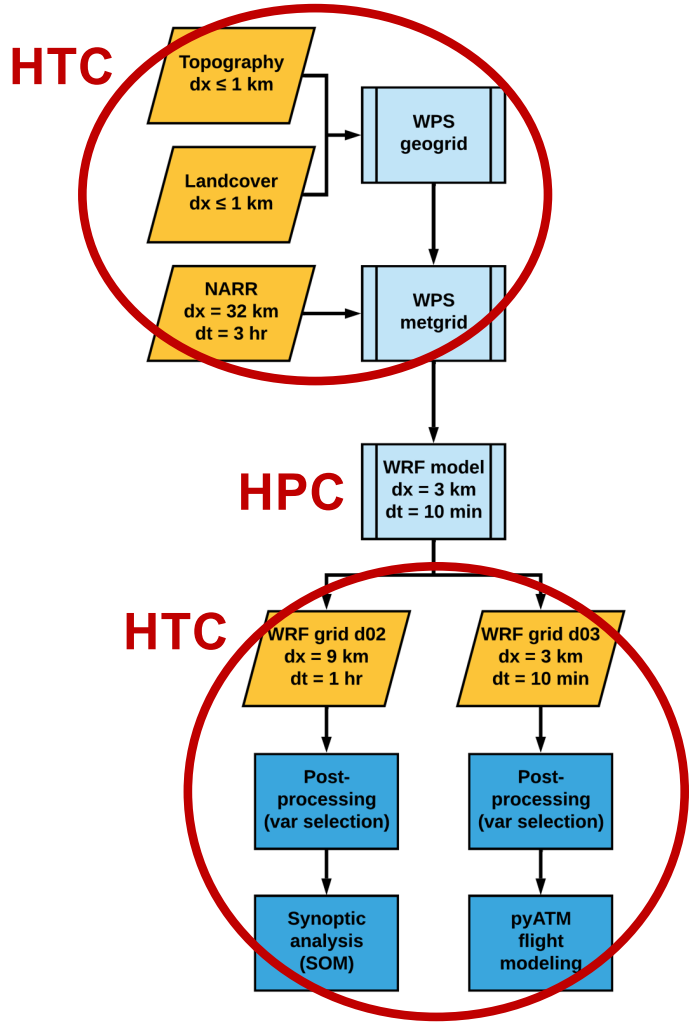
Weather Research & Forecasting (WRF) Model v4.1

Pre-processing: 1 month at a time
    single-processor, high-memory

Main model: 1 day (30 h) at a time
    HPC using MPI (60+ processors)
    separate days can run concurrently

Post-processing: 1 output file at a time
    single-processor, distributed

HTCondor DAGMan script(s)
  with mixed submission protocols
    HTC → HPC → HTC



Weather Research & Forecasting (WRF) Model v4.1 Pre-processing

```
# WPS_WRF_201307_dag.sub using HTC with NARR data from external source and /staging and HPC

JOB WPS_fixed WPS-NARR_fixed.sub

JOB WPS_201307_geogrid WPS-NARR_geogrid.sub
VARS WPS_201307_geogrid yyyy="2013" mm="07" dd1="01" dd2="31"
PARENT WPS_fixed CHILD WPS_201307_geogrid

JOB WPS_201307_metgrid WPS-NARR_metgrid.sub
VARS WPS_201307_metgrid yyyy="2013" mm="07" dd1="01" dd2="31"
SCRIPT PRE WPS_201307_metgrid WPS-NARR_pre.sh 2013 07 01 31
SCRIPT POST WPS_201307_metgrid WPS-NARR_post.sh 2013 07 01 31
PARENT WPS_201307_geogrid CHILD WPS_201307_metgrid

JOB WPS_201308_geogrid WPS-NARR_geogrid.sub
VARS WPS_201308_geogrid yyyy="2013" mm="08" dd1="01" dd2="31"
PARENT WPS_fixed CHILD WPS_201308_geogrid

JOB WPS_201308_metgrid WPS-NARR_metgrid.sub
VARS WPS_201308_metgrid yyyy="2013" mm="08" dd1="01" dd2="31"
SCRIPT PRE WPS_201308_metgrid WPS-NARR_pre.sh 2013 08 01 31
SCRIPT POST WPS_201308_metgrid WPS-NARR_post.sh 2013 08 01 31
PARENT WPS_201308_geogrid CHILD WPS_201308_metgrid

SPLICE WRF_201307 WRF_201307_dag.sub
PARENT WPS_201307 WPS_201308 CHILD WRF_201307
```

WPS (WRF Pre-processing System)

WRF model + post-processing

Weather Research & Forecasting (WRF) Model v4.1 + Post-processing

```
# WRF_201307_dag.sub using HPC with pre-model configuration + executable generators + HTC

...

JOB WRF_20130715 WRF_60_tiles.sub
VARS WRF_20130715 yyyy1="2013" mm1="07" dd1=”15" yyyy2="2013" mm2="07" dd2=”16"
SCRIPT PRE WRF_20130715 WRF_pre.sh 2013 07 15 2013 07 16 60
SCRIPT POST WRF_20130715 WRF_post.sh 2013 07 15 2013 07 16

JOB WRF_20130715_d02_reduce WRF_out_d02_reduce.sub
VARS WRF_20130715_d02_reduce yyyy1="2013" mm1="07" dd1=”15" yyyy2="2013" mm2="07" dd2=”16"
SCRIPT POST WRF_20130715_d02_reduce WRF_out_reduce_post.sh 2 2013 07 15 2013 07 16
PARENT WRF_20130715 CHILD WRF_20130715_d02_reduce

JOB WRF_20130715_d03_reduce WRF_out_d03_reduce.sub
VARS WRF_20130715_d03_reduce yyyy1="2013" mm1="07" dd1=”15" yyyy2="2013" mm2="07" dd2=”16"
SCRIPT POST WRF_20130715_d03_reduce WRF_out_reduce_post.sh 3 2013 07 15 2013 07 16
PARENT WRF_20130715 CHILD WRF_20130715_d03_reduce

...
```

WRF model

post-processing (varies seasonally)

Weather Research & Forecasting (WRF) Model v4.1 on HPC

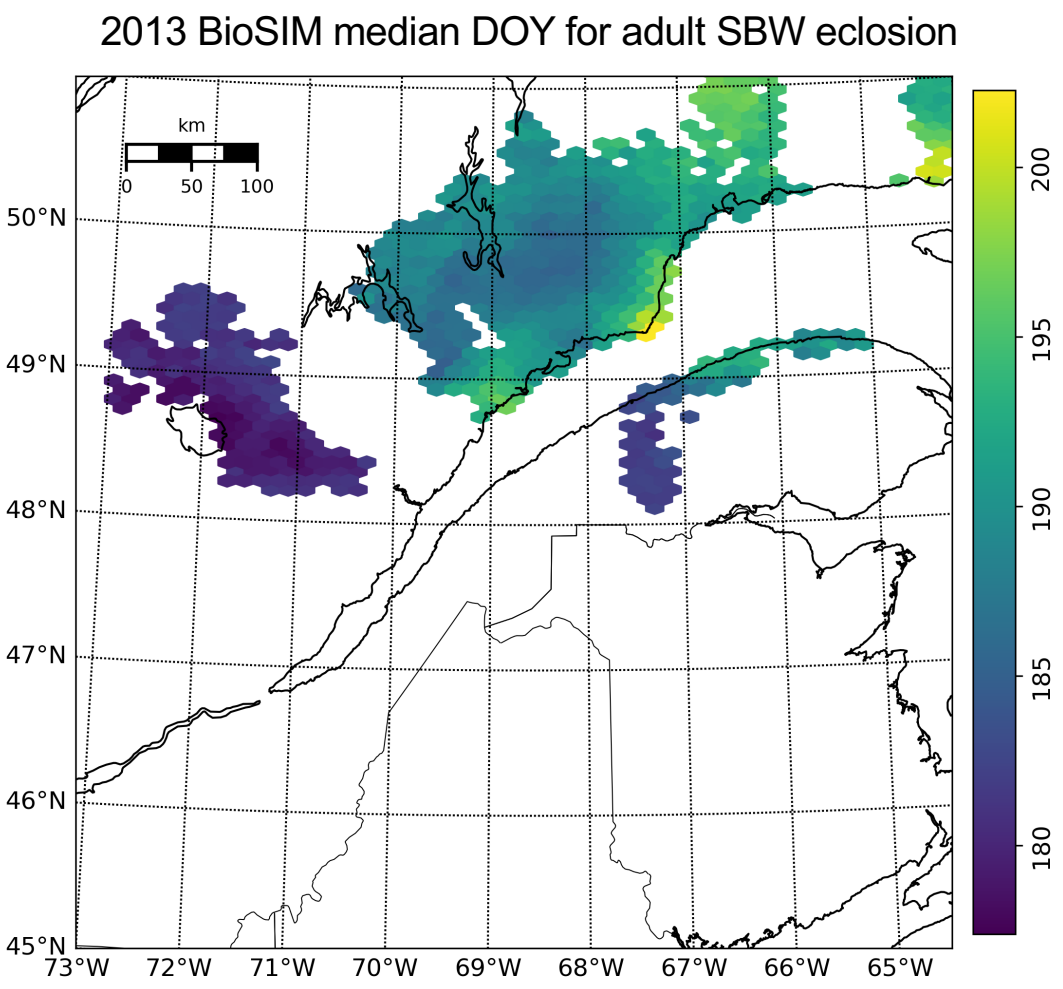
```
# WRF_60_tiles.sub using HPC with input from HTC and pre-processing script

universe = grid
grid_resource = batch slurm hpclogin1.chtc.wisc.edu
batch_queue = univ2
batch_runtime = 86400
batch_extra_submit_args = --nodes=5 --ntasks-per-node=12 --cpus-per-task=1 --ntasks=60 --mem-per-cpu=10280
                          --error=WRF_$(yyyy1)$(mm1)$(dd1)06-$(yyyy2)$(mm2)$(dd2)12_60_tiles.slurm_err
                          --output=WRF_$(yyyy1)$(mm1)$(dd1)06-$(yyyy2)$(mm2)$(dd2)12_60_tiles.slurm_out
log = WRF_$(yyyy1)$(mm1)$(dd1)06-$(yyyy2)$(mm2)$(dd2)12.log
error = WRF_$(yyyy1)$(mm1)$(dd1)06-$(yyyy2)$(mm2)$(dd2)12.err
output = WRF_$(yyyy1)$(mm1)$(dd1)06-$(yyyy2)$(mm2)$(dd2)12.out
requirements = (OpSys == "LINUX")
+RequestCpus = Undefined
+RequestMemory = Undefined
request_disk = 64GB
should_transfer_files = YES
transfer_input_files =
          WRF_input/$(yyyy1)/$(mm1)/WRF_$(yyyy1)$(mm1)$(dd1)06-$(yyyy2)$(mm2)$(dd2)12_60_tiles_namelist.wrf,
          WRF_input/$(yyyy1)/$(mm1)/met_em.d0x.$(yyyy1)-$(mm1)-$(dd1).tar.gz,
          WRF_input/$(yyyy2)/$(mm2)/met_em.d0x.$(yyyy2)-$(mm2)-$(dd2).tar.gz
executable = WRF_input/$(yyyy1)/$(mm1)/WRF_$(yyyy1)$(mm1)$(dd1)06-$(yyyy2)$(mm2)$(dd2)12_60_tiles.sh
when_to_transfer_output = ON_EXIT
transfer_output_files = WRF_output
queue 1
```

→ starting to use Stampede2 @ TACC via NSF XSEDE

Python-based Atmospheric Transport Model (pyATM) for SBW moths

Adult dispersal is weather-driven, occurs almost nightly in summer

Each night's cohort includes "new" moths *and* those that survived the previous night(s)

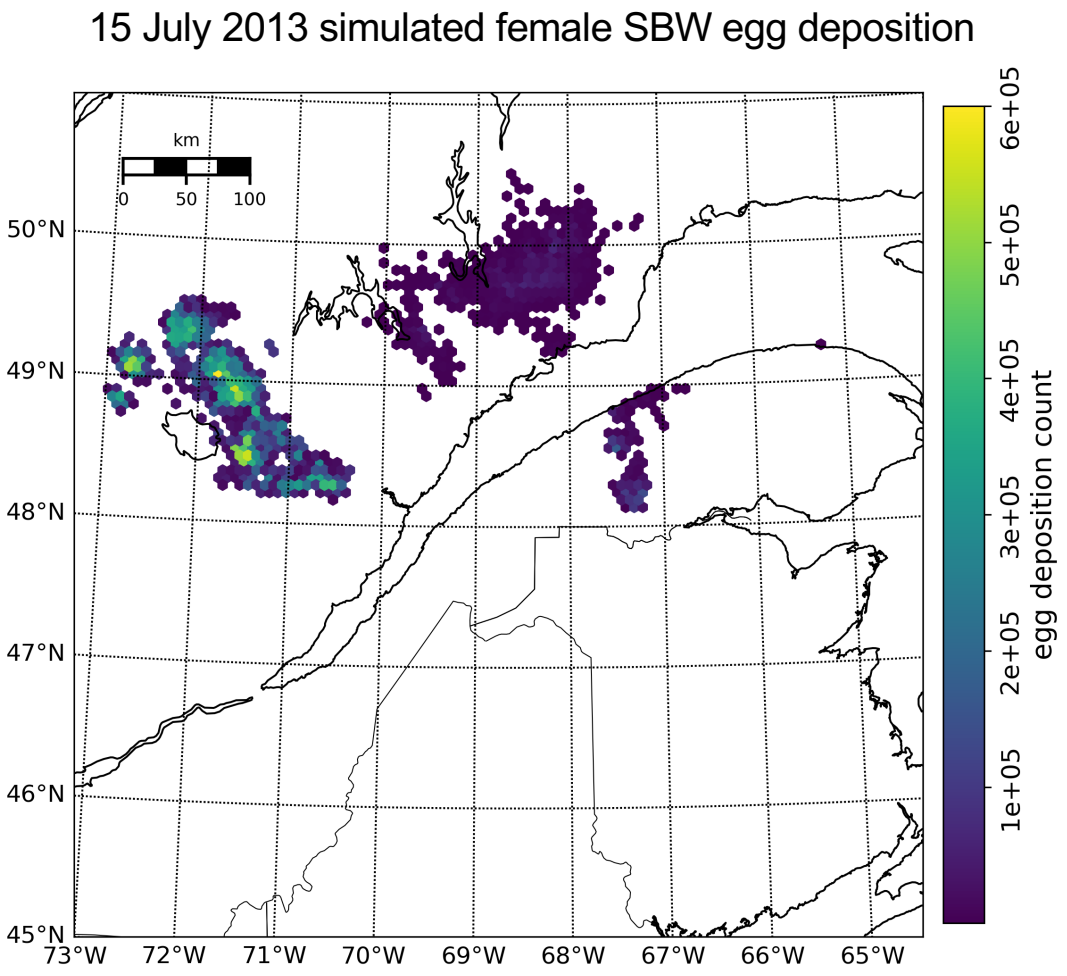


Python-based Atmospheric Transport Model (pyATM) for SBW moths

Adult dispersal is weather-driven,
occurs almost nightly in summer

Each night's cohort includes "new"
moths *and* those that survived
the previous night(s)

F lay eggs on host trees, potentially
at each new landing location
→ daily changes in weight
and flight capability



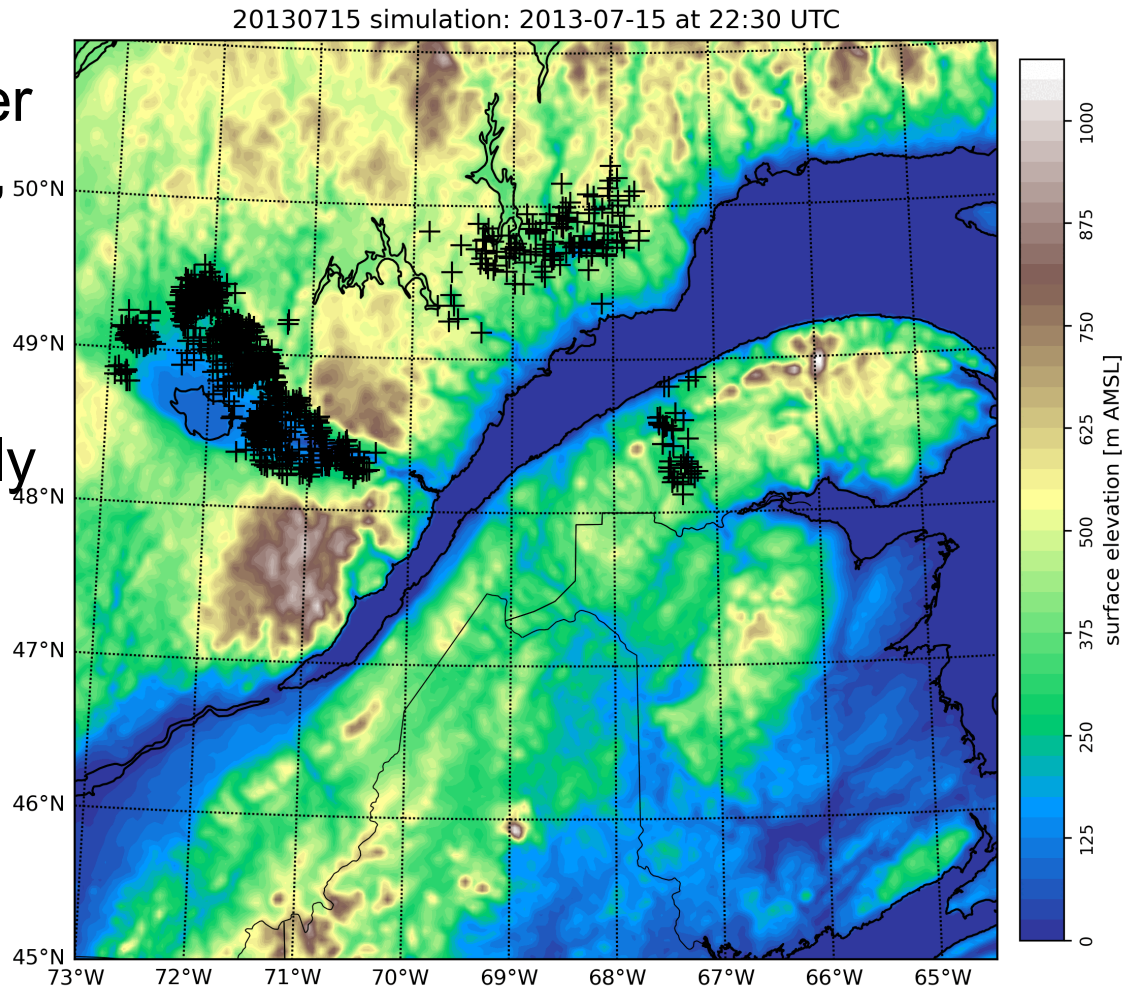
Python-based Atmospheric Transport Model (pyATM) for SBW moths

Adult dispersal is weather-driven,
occurs almost nightly in summer

Each night's cohort includes "new"
moths *and* those that survived
the previous night(s)

F lay eggs on host trees, potentially
at each new landing location
→ daily changes in weight
and flight capability

M + F have diff. flight capabilities
→ flight altitudes + distances



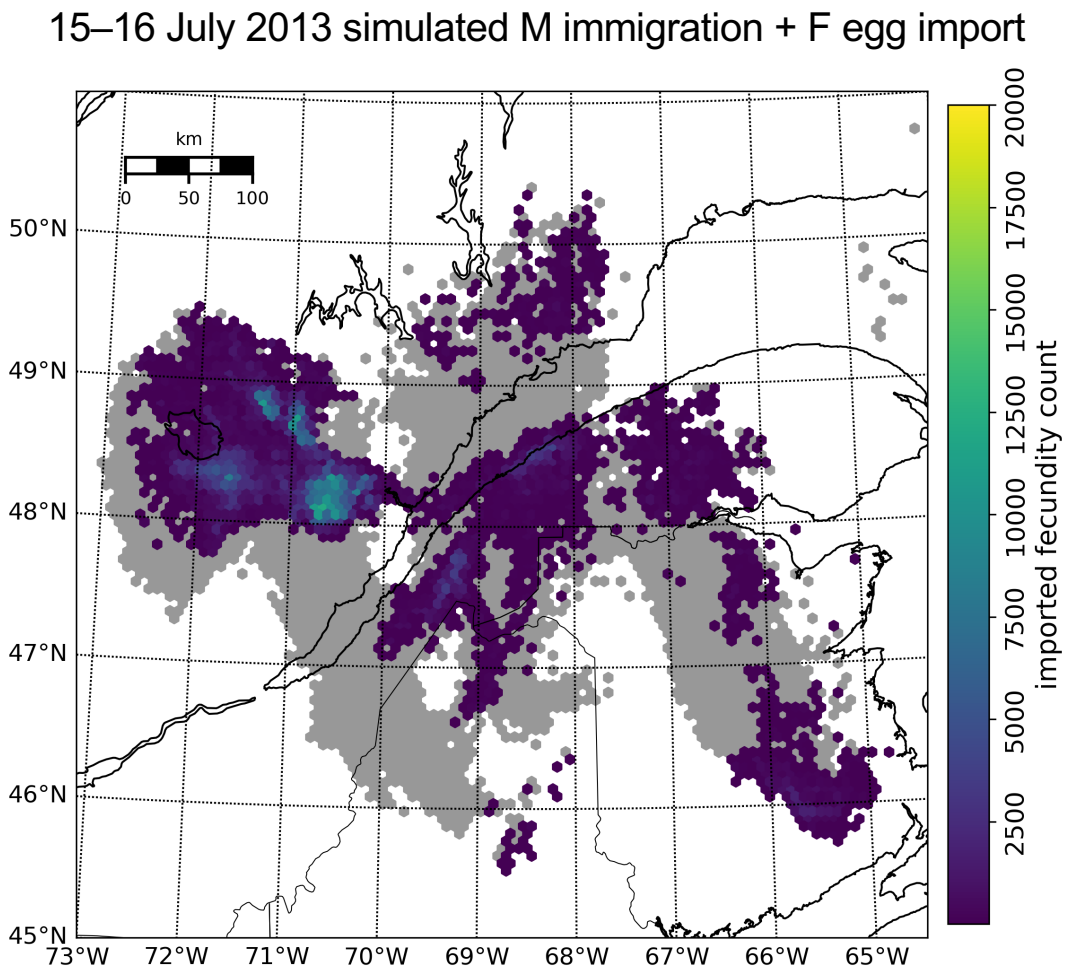
Python-based Atmospheric Transport Model (pyATM) for SBW moths

Adult dispersal is weather-driven,
occurs almost nightly in summer

Each night's cohort includes "new"
moths *and* those that survived
the previous night(s)

F lay eggs on host trees, potentially
at each new landing location
→ daily changes in weight
and flight capability

M + F have diff. flight capabilities
→ flight altitudes + distances
→ spatial outcomes + roles



Python-based Atmospheric Transport Model (pyATM) for SBW moths

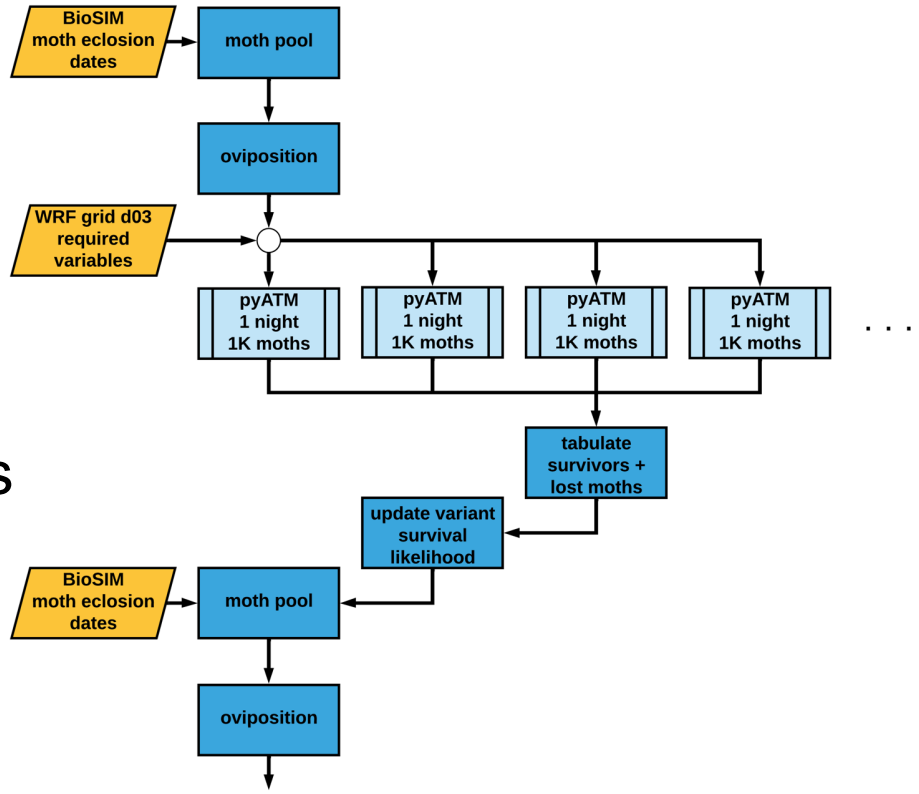
Stochastic simulations
- random subsets of available moths
- liftoff times vary with local conditions
- weather, esp. near the surface (BL)

→ ensembles of simulations

Sequential nightly simulations
- accounting the variety of prior results
- compounded timelines + variants

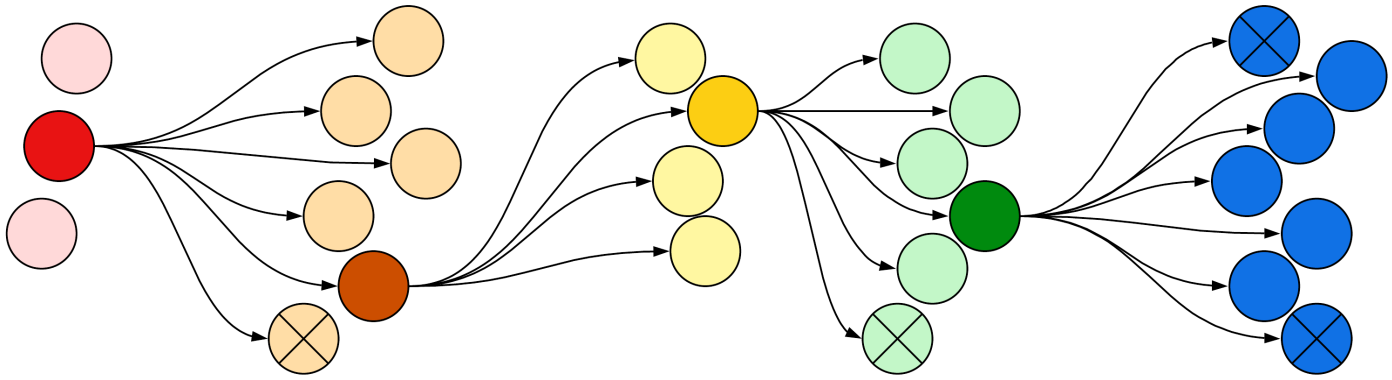
→ Markov chain Monte Carlo process
with Bayesian likelihood updating

→ Probabilistic outcomes over the
mating + dispersal + oviposition period



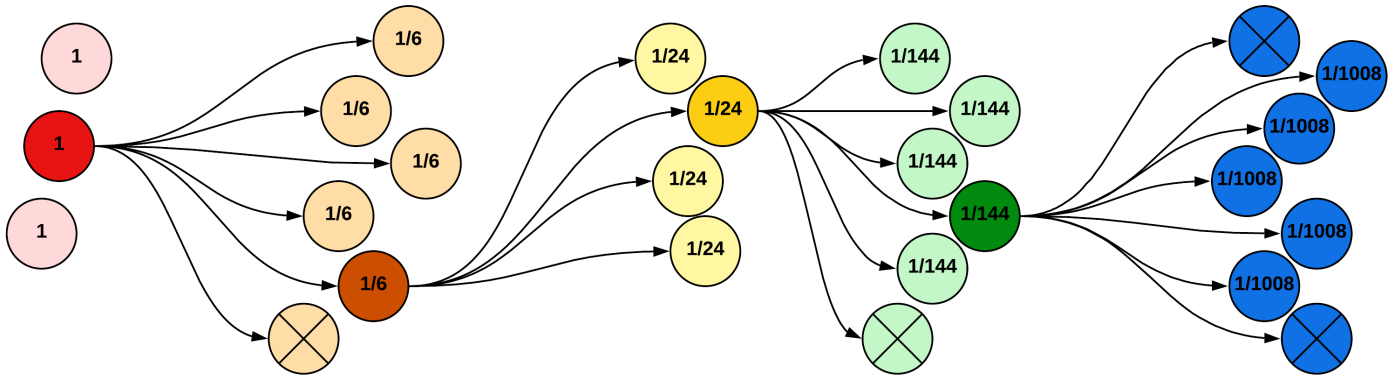
Python-based Atmospheric Transport Model (pyATM) for SBW moths

```
# 2013_sequential_dag.sub using Flocking + Glide-In (OSG) with WRF-based input and input moth data processing

...

JOB ATM_20130714 ATM_WRF-NARR_default.sub
VARS ATM_20130714 grid="d03" begin_date="20130714" end_date="20130715" nreps="100" start="0"
SCRIPT PRE ATM_20130714 ATM_WRF-NARR_default_pre.sh d03 2013 07 14 2013 07 15
SCRIPT POST ATM_20130714 ATM_WRF-NARR_default_post.sh d03 2013 07 14 2013 07 15
PARENT ATM_20130713 CHILD ATM_20130714

JOB ATM_20130715 ATM_WRF-NARR_default.sub
VARS ATM_20130715 grid="d03" begin_date="20130715" end_date="20130716" nreps="100" start="0"
SCRIPT PRE ATM_20130715 ATM_WRF-NARR_default_pre.sh d03 2013 07 15 2013 07 16
SCRIPT POST ATM_20130715 ATM_WRF-NARR_default_post.sh d03 2013 07 15 2013 07 16
PARENT ATM_20130714 CHILD ATM_20130715

...
```



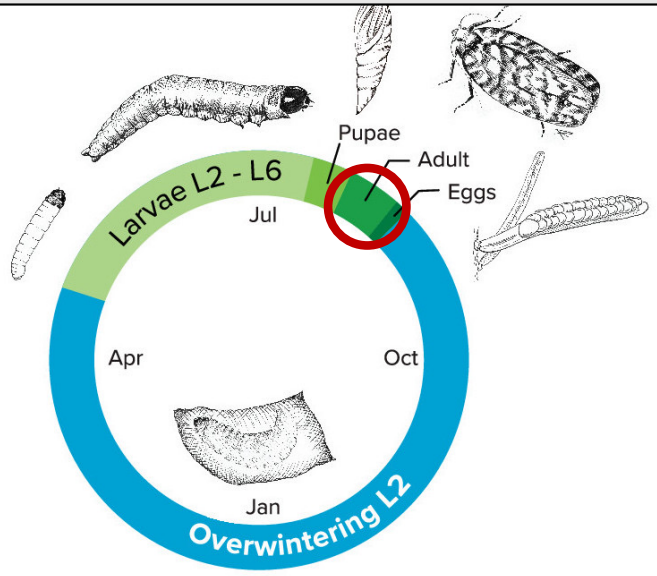
Python-based Atmospheric Transport Model (pyATM) for SBW moths

```
#!/bin/bash
# ATM_WRF-NARR_default_pre.sh generates pool of candidate moths and pyATM configuration options

grid="$1"
begin_year="$2" ; begin_month="$3" ; begin_day="$4" ; begin_date="$2""$3""$4"
end_year="$5" ; end_month="$6" ; end_day="$7" ; end_date="$5""$6""$7"

python generate_Simulation_flier_pool.py "$begin_date" "$begin_year"_BioSIM_SBW_output.csv >
     ATM_WRF-NARR_"$grid"_"$begin_date"_ready_moths_processing.out

python generate_Simulation_specifications.py "$begin_date" "$end_date"

[various file management commands]
```



Python-based Atmospheric Transport Model (pyATM) for SBW moths

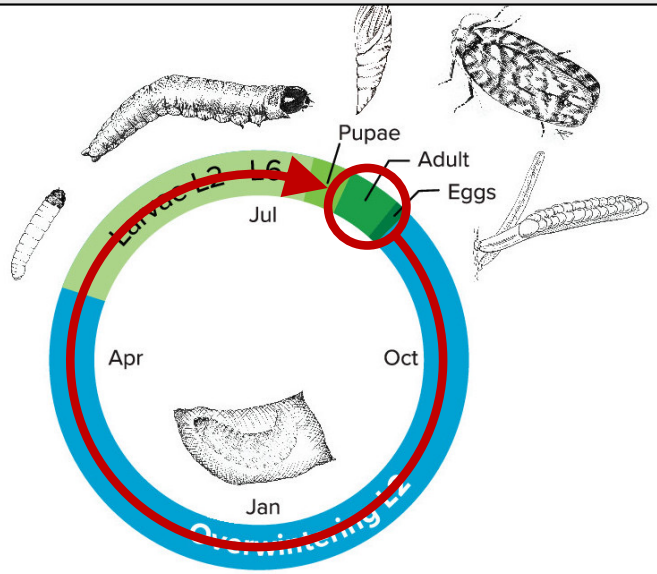
```
#!/bin/bash
# ATM_WRF-NARR_default_pre.sh generates pool of candidate moths and pyATM configuration options

grid="$1"
begin_year="$2" ; begin_month="$3" ; begin_day="$4" ; begin_date="$2""$3""$4"
end_year="$5" ; end_month="$6" ; end_day="$7" ; end_date="$5""$6""$7"

python generate_Simulation_flier_pool.py "$begin_date" "$begin_year"_BioSIM_SBW_output.csv >
     ATM_WRF-NARR_"$grid"_"$begin_date"_ready_moths_processing.out

python generate_Simulation_specifications.py "$begin_date" "$end_date"

[various file management commands]
```



→ can easily abstract to new DAG node for daytime biological process model (aging, mating, sub-canopy dispersal)

Python-based Atmospheric Transport Model (pyATM) for SBW moths

```
#!/bin/bash
# ATM_WRF-NARR_default_pre.sh generates pool of candidate moths and pyATM configuration options

grid="$1"
begin_year="$2" ; begin_month="$3" ; begin_day="$4" ; begin_date="$2""$3""$4"
end_year="$5" ; end_month="$6" ; end_day="$7" ; end_date="$5""$6""$7"

python generate_Simulation_flier_pool.py "$begin_date" "$begin_year"_BioSIM_SBW_output.csv >
     ATM_WRF-NARR_"$grid"_"$begin_date"_ready_moths_processing.out

python generate_Simulation_specifications.py "$begin_date" "$end_date"

[various file management commands]
```



→ can easily abstract to new DAG node for daytime biological process modeling (aging, mating, oviposition)

→ can easily (-ish?) expand DAG scope for seasonal/annual biological modeling (from gen-0 eggs to gen-1 eggs)