

December 18, 2012

# Framework for Reconstructing Epidemic Dynamics

## Models of Infectious Disease Agent Study

# XSEDE

Extreme Science and Engineering  
Discovery Environment



# MIDAS Project

- Established May 2004
- Collaborative network of research scientists
- Currently receives around \$13 million a year from the National Institute of General Medical Sciences (part of NIH)
- Develops computational, statistical and mathematical models used to understand the dynamics of infectious disease

# MIDAS Project

- University of Pittsburgh National Center of Excellence studies behavioral, environmental and evolutionary factors underlying infectious disease epidemics
- Key component is the creation of outreach and training programs aimed at educating scientists in epidemic modeling techniques

<http://www.nigms.nih.gov/Research/FeaturedPrograms/MIDAS/>

<https://midas.psc.edu/>



# Framework for Reconstructing Epidemiological Dynamics (FRED)

- Open source agent based modeling system
- Developed by the University of Pittsburgh Public Health Dynamics Lab, Pittsburgh Supercomputing Center, and the School of Computer Science at Carnegie Mellon
- Uses synthetic populations derived from census data to capture geographic and demographic distributions



# Framework for Reconstructing Epidemiological Dynamics (FRED)

- Incorporates households, schools, and workplace social networks
- Support for:
  - Multiple circulating and evolving strains
  - Mitigation strategies such as vaccinations and school closure policies
  - Behavioral changes such as vaccine acceptance, personal hygiene, and spontaneous social distancing

# ECSS Project Team

- Shawn Brown, Principal Investigator (Pitt)
- John Grefenstette, Principal Investigator (Pitt) for the FRED development team
- Jay DePasse, Programmer (Pitt), initiated investigations of new software components and addressed all implementation details
- David O'Neal, Consultant (PSC), provided baseline, intermediate, and final metrics; configured and performed Blacklight tests

# FRED – Test Cases

- Benchmark test cases were based on population and location databases for:
  - Allegheny County, PA (1.2M agents)
  - State of Pennsylvania (12M agents)
  - United States (274M agents)
- Single influenza strain; attack rate of 50%
- 100-day simulations ignored births, deaths, aging of the population, school closures, etc.

# FRED – From the developers

- US model input files
  - 17 GB population data (274M records)
  - 6 GB location data (114M records)
- Memory usage
  - 540 GB
- Run time
  - 96 hours

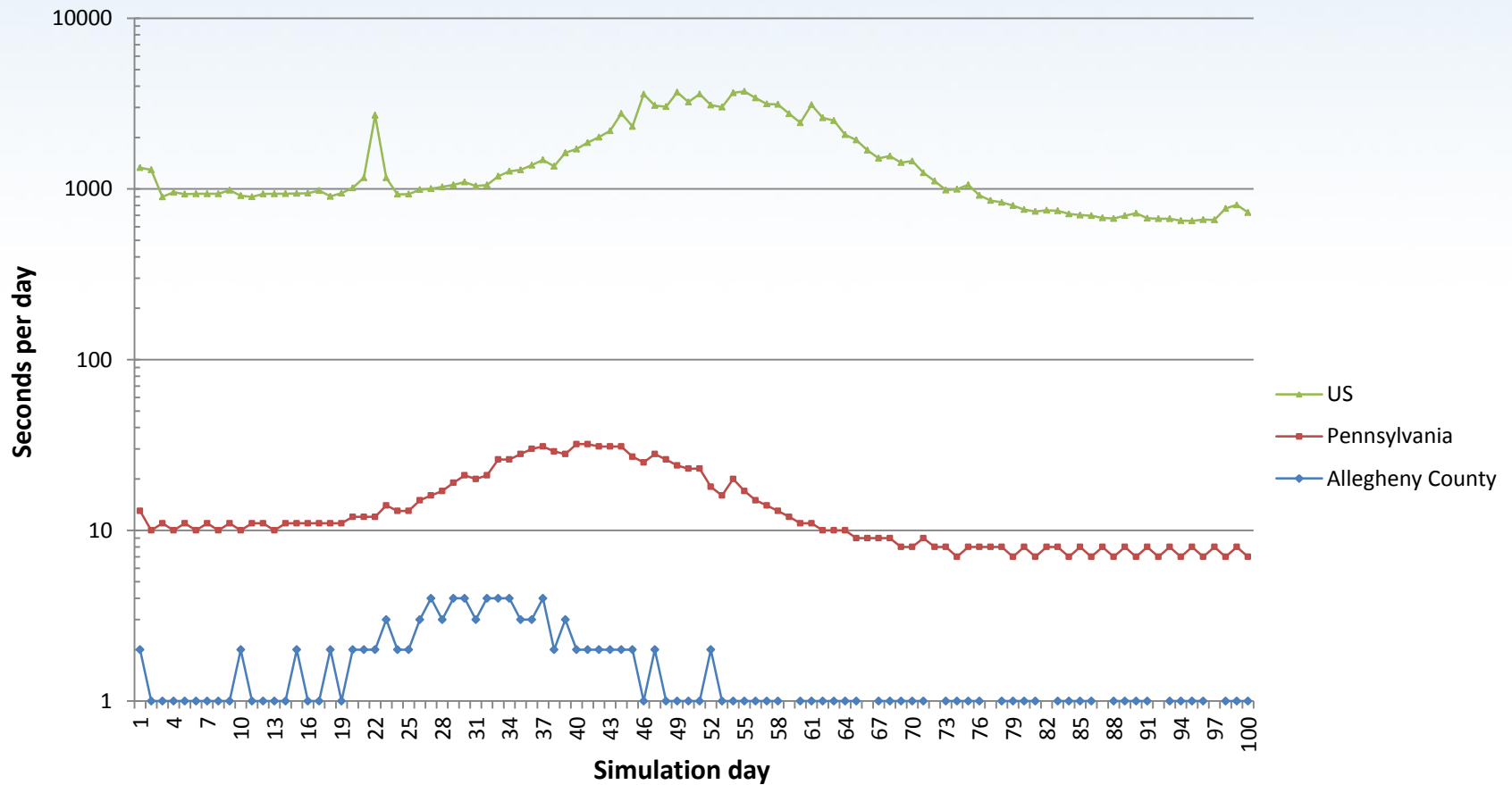


# FRED r0 – Baseline

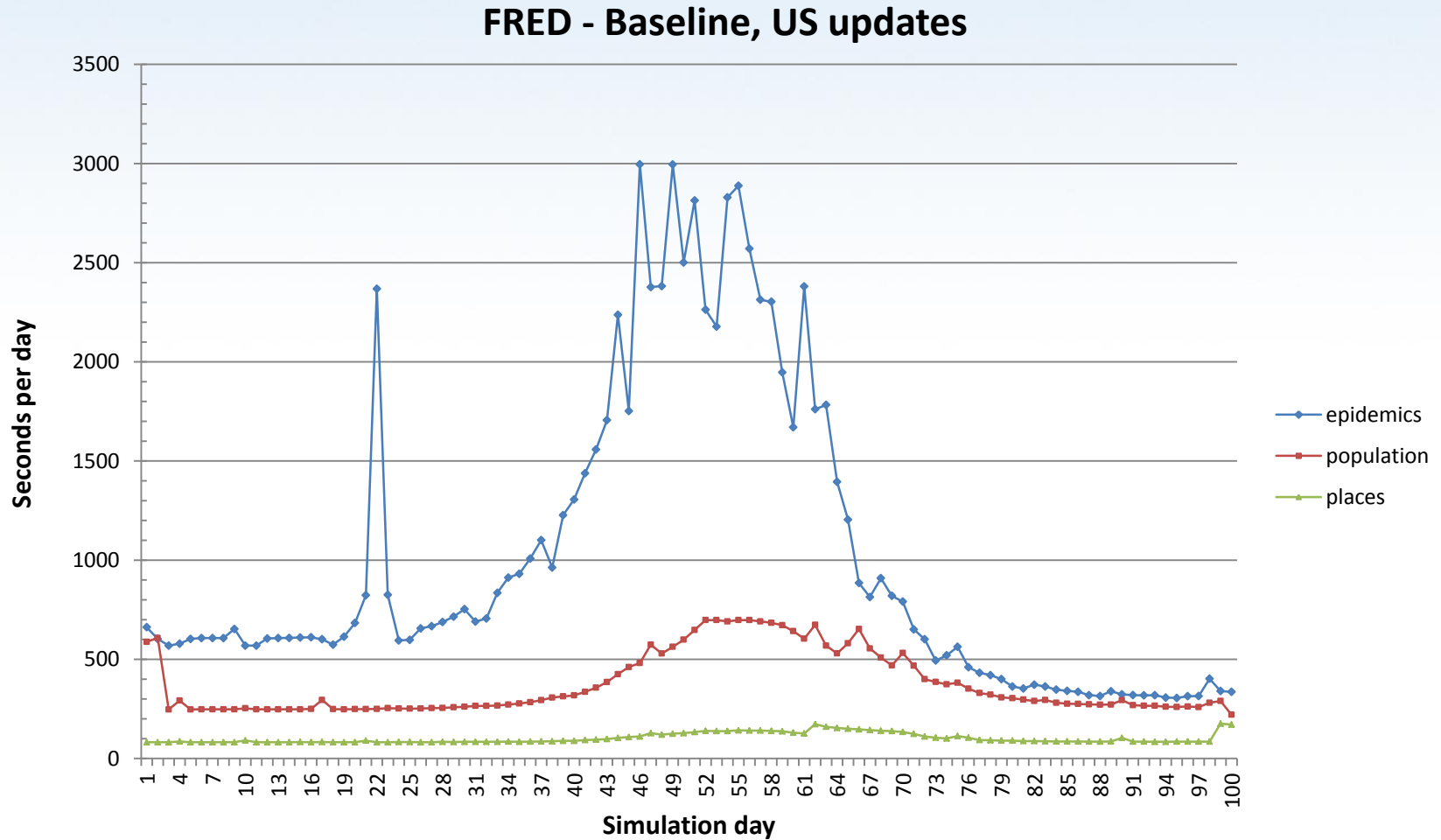
- US model input files
  - 23 GB
- Memory usage
  - 585 GB (*SCRATCH\_RAMDISK I/O*)
- *Numerous bug fixes*
- *Aggressive compiler optimizations*
- Run time
  - 51.5 hours

# FRED – Baseline Chart, Multiple Domains

FRED - Baseline, All updates



# FRED – Baseline Chart, US Updates



# FRED r1 – Hoard

- US model input files
  - 23 GB
- Memory usage
  - 540 GB (*back out SCRATCH\_RAMDISK usage*)
- *Hoard 3.8 preload*
- Run time
  - 31.4 hours

# FRED r2 – Data structure mods, dSFMT

- US model input files
  - 23 GB
- Memory usage
  - **403 GB** (*data structure mods, block allocations*)
- Hoard
- ***dp SIMD Fast Mersenne Twister 2.1***
- Run time
  - **20.5 hours**



# FRED r3 – Snappy, OpenMP

- US model input files
  - 8 GB (*Snappy*)
- Memory usage
  - 230 GB (*data structure mods, block allocations*)
- Hoard and dSFMT
- *Snappy 1.0.5 file compression/decompression*
- Run time
  - 5.1 hours at 4 threads

# FRED r4 – TCMalloc, dSFMT update

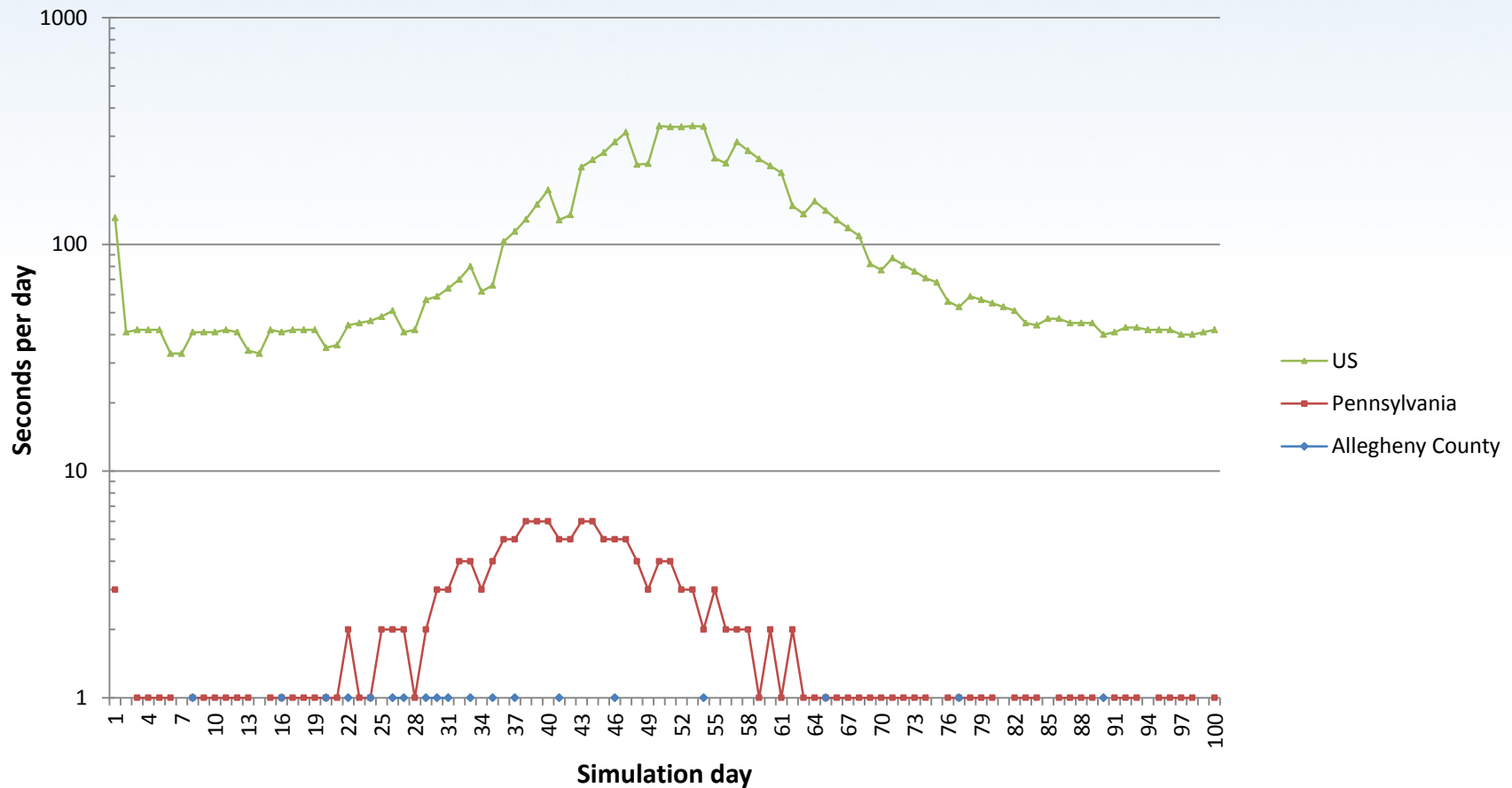
- US model input files
  - 8 GB total
- Memory usage
  - 210 GB (*TCMalloc*)
- *gperftools TCMalloc 2.0* (replaced Hoard)
- *dSFMT 2.2 update*, Snappy
- Run time
  - 4.3 hours at 4 threads

# FRED – Final

- US model input files
  - 8 GB total (Snappy, */local disk*)
- Memory usage
  - 210 GB
- dSFMT, Snappy, TCMalloc
- *Dynamic scheduling, chunk=10*
- Run time
  - 3 hours at 16 threads

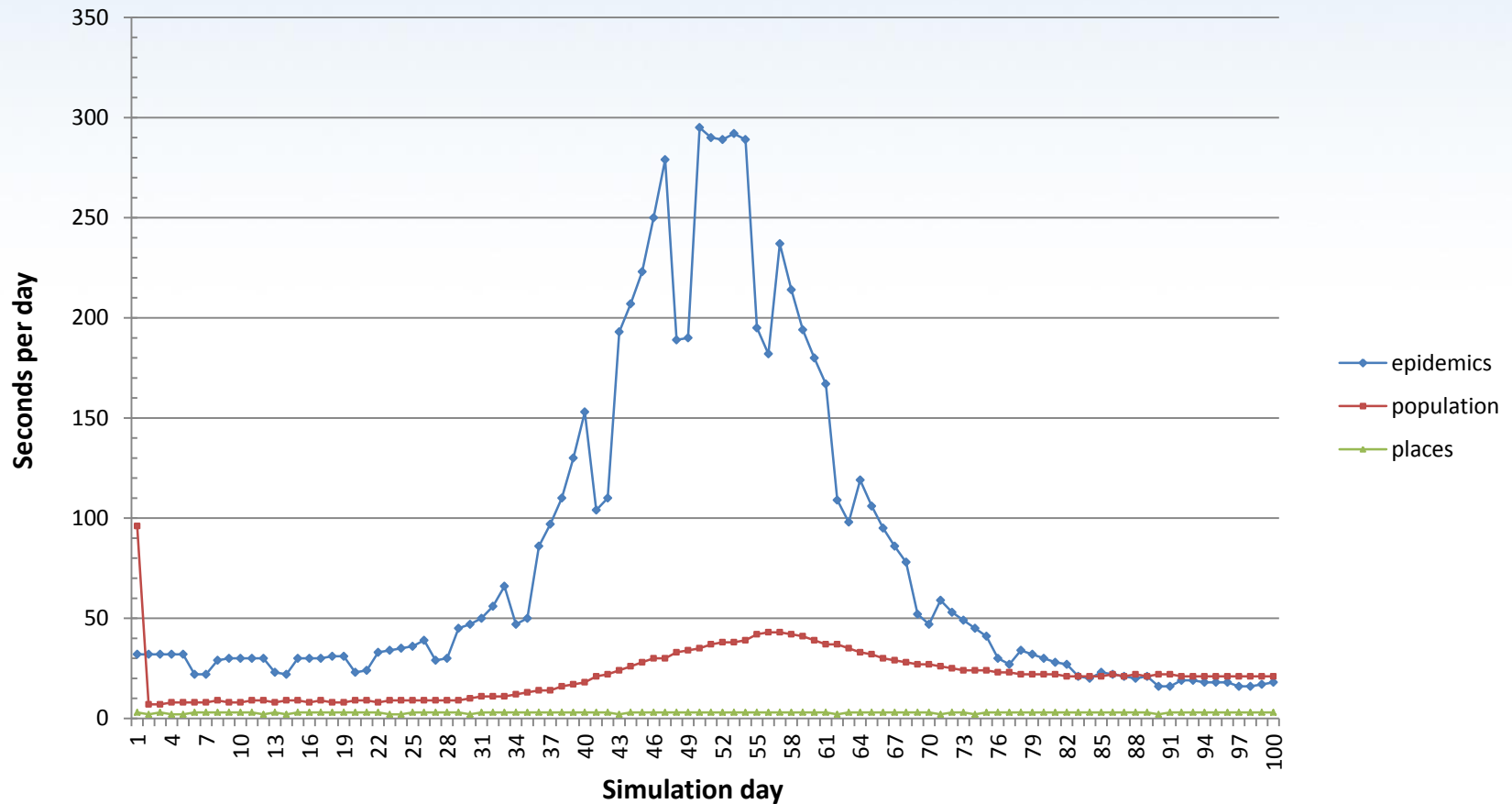
# FRED – Final Chart, Multiple Domains

FRED - 16 threads, all updates



# FRED – Final Chart, US Updates


FRED - 16 threads, US updates





# Summary – US Domain

- Input files
  - 23 GB to 8 GB (3x)
- Wall time
  - 96 hours to 3 hours (32x)
- Memory
  - 540 GB to 210 GB (2.5x)
- Job charges
  - 7680 SUs to 96 SUs (80x)



Our reach will forever  
exceed our grasp, but,  
in stretching our horizon,  
we forever improve our world.

**XSEDE**

Extreme Science and Engineering  
Discovery Environment