



Characterizing bioterrorist attacks from a short time series of diagnosed patient data: A Bayesian approach

SAND2006-7944C

J. Ray, Y. M. Marzouk, M. Kraus and P. Fast

jairay@ca.sandia.gov

Sandia National Laboratories, Livermore CA,
NORAD-NORTHCOM, Colorado Springs, CO
Lawrence Livermore National Laboratory, Livermore, CA



Problem and motivation

- **Consider a bioattack**
 - Atmospheric release of an aerosolized pathogen
 - Not caught on sensors
 - Not terribly big – $O(10^3)$ infected people
 - First intimation : successful diagnosis of an infected individual
- **The technical challenge**
 - *Infer* ($\tau, N, \langle D \rangle$)
 - *Inputs*: $\{t_i, n_i\}, i = 1 \dots M$, time series of new symptomatics every day / every 6 hrs.
- **Restrictions**
 - Can only use 3-4 days of data, past 1st diagnosis i.e. M is small
 - Quantify uncertainty due to incomplete observation / limited data
 - Noise – stochastic data
 - Expect model errors – i.e. model (used for inference) is approximate



Methodology

- **Research Challenge**

- Little prior work – 2 published papers on the general topic
- No contagious diseases, simplified models for non-contagious ones
- All recent publications (oldest is 2004)

- **Bayesian Inference**

- Likelihood Λ of observing a $\{t_i, n_i\}$, sequence given a $(\tau, N, \langle D \rangle)$ attack can be analytically derived [1]
- Exploits the dose-dependent incubation period distribution of a disease

$$P(N, \tau, \langle D \rangle | \{t_i, n_i\}) \propto \Lambda(\{t_i, n_i\} | N, \tau, \langle D \rangle) \pi_N(N) \pi_\tau(\tau) \pi_D(\langle D \rangle)$$

- **Simulated aerosol attacks to generate data**

- Assume a city with a generic population distribution
- Lay down a plume, infect people with different dosages
- Dose dependent anthrax incubation period models [2; stochastic !]
- Sources of errors – noise, model errors, incomplete observation

- **Also invert the Sverdlovsk anthrax incident of 1979**

1. Ray et al, Sandia Technical. Report., SAND2006-1492
2. Wilkening, PNAS, 103(20):7589-7594, May 2006.

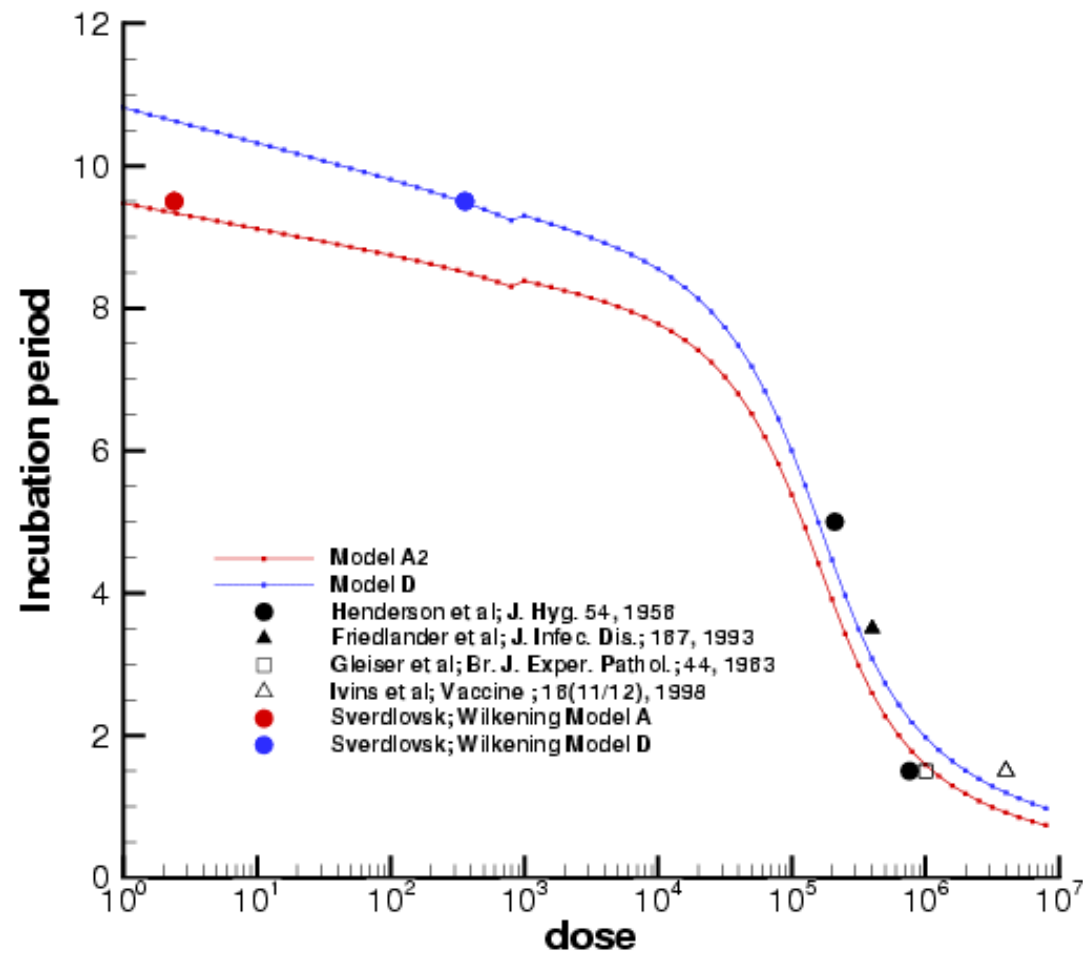



Anthrax incubation period models

- Spores are subjected to competing processes
 - Clearance by immune system and germination into vegetative cells (rates obtained from non-human primate expts.)
 - PDF for time to germination (PDF #1)
- Vegetative cells reproduce at various rates (random variable)
- A threshold number of vegetative cells triggers symptoms
- Time from germination to symptoms, s , has a log-normal distribution (PDF #2)
- Convolution of PDF # 1 and PDF #2 gives incubation period distribution
- Parameters calculated from non-human primate experiments and Sverdlovsk, 1979.



Attack and inference models



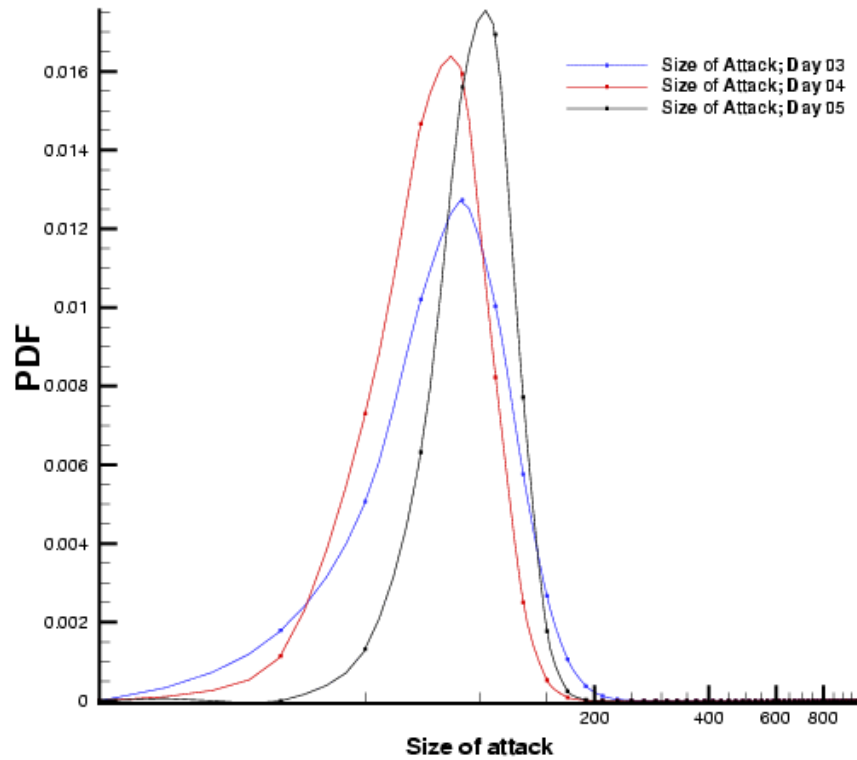


Check No. 1 – Ideal case

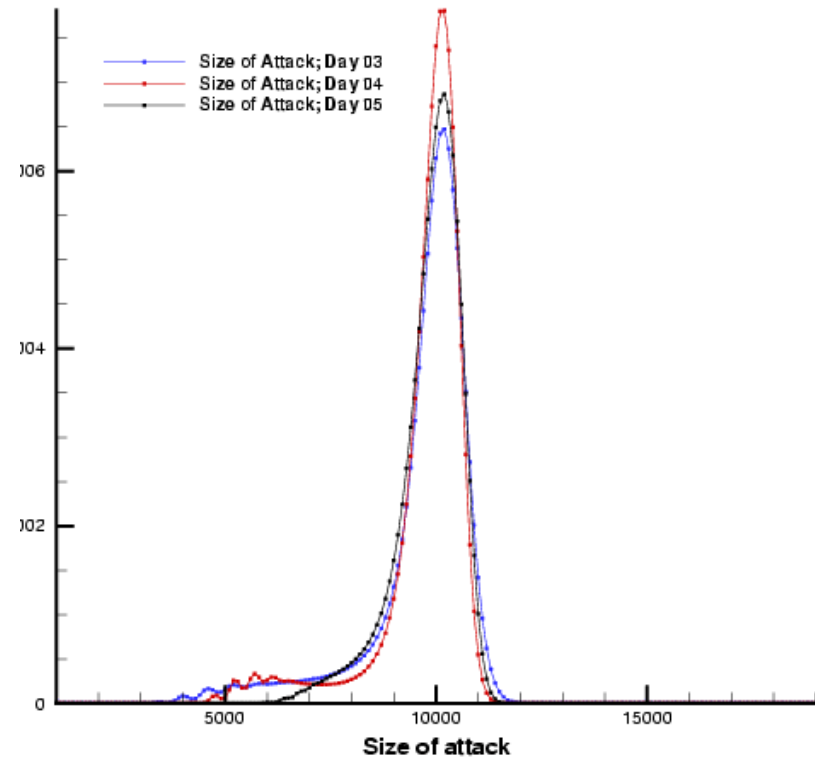
- Does the method work in the ideal case?
- **Approach :**
 - Simulate 2 “ideal” attacks
 - Case B : 100 infected people
 - Case E : 10,000 infected people
 - Every infected person receives a dose of 100 spores
 - The disease progresses as per the blue model
 - Collect observations (# of symptomatic people) over 6-hr intervals
 - Inference as per blue mode too
 - No model errors !
 - Infer characteristics of attack based on 3-5 days of data
- **Discrepancy between characterization and simulation due to:**
 - Noise in the observations
 - Incomplete observation



Inference of size of attack



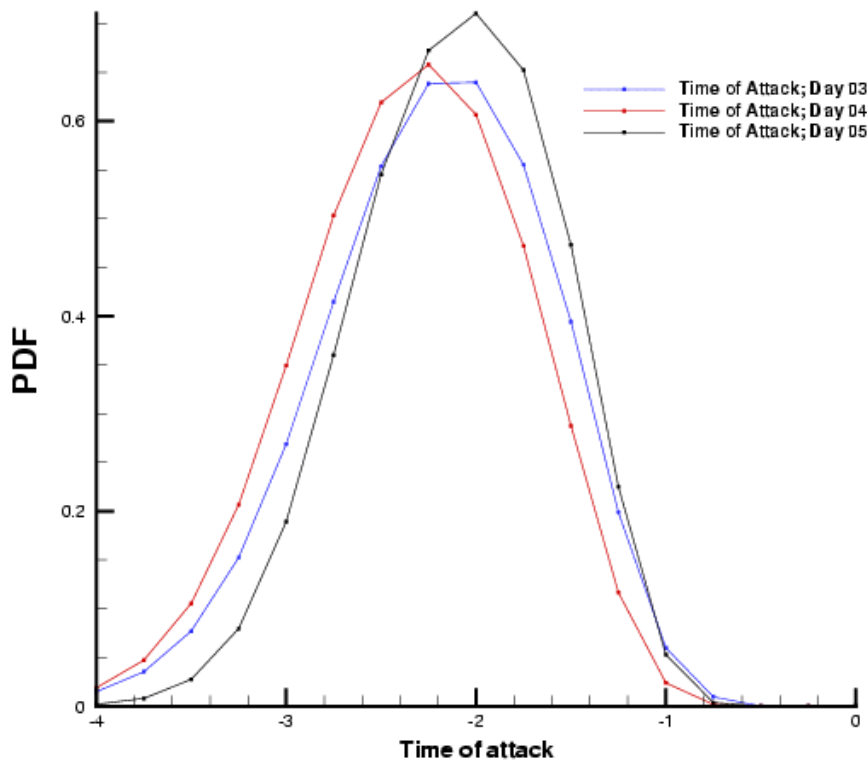
Case B : $N = 100$, $\tau = -2.25$,
 $\log_{10}(D) = 2$



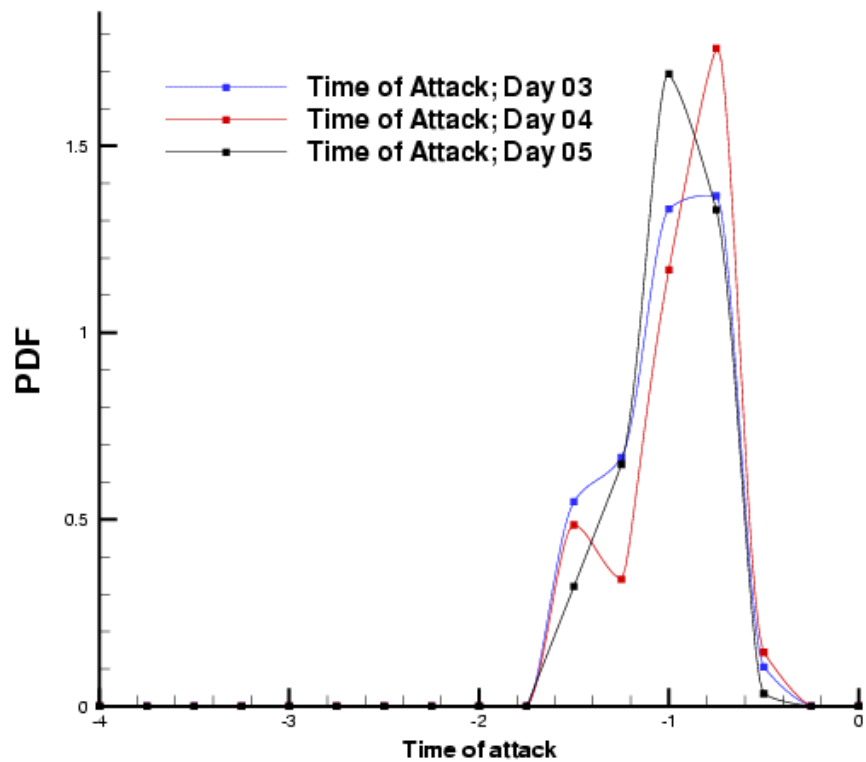
Case E : $N = 10,000$, $\tau = -1.0$,
 $\log_{10}(D) = 2$



Inference of time of attack



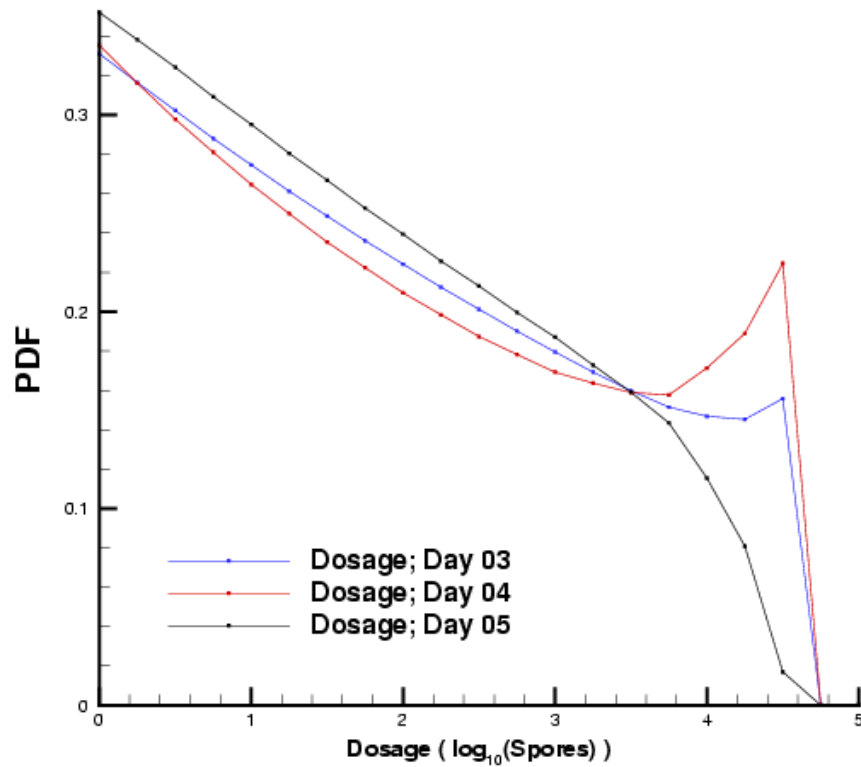
Case B : $N = 100$, $\tau = -2.25$,
 $\log_{10}(D) = 2$



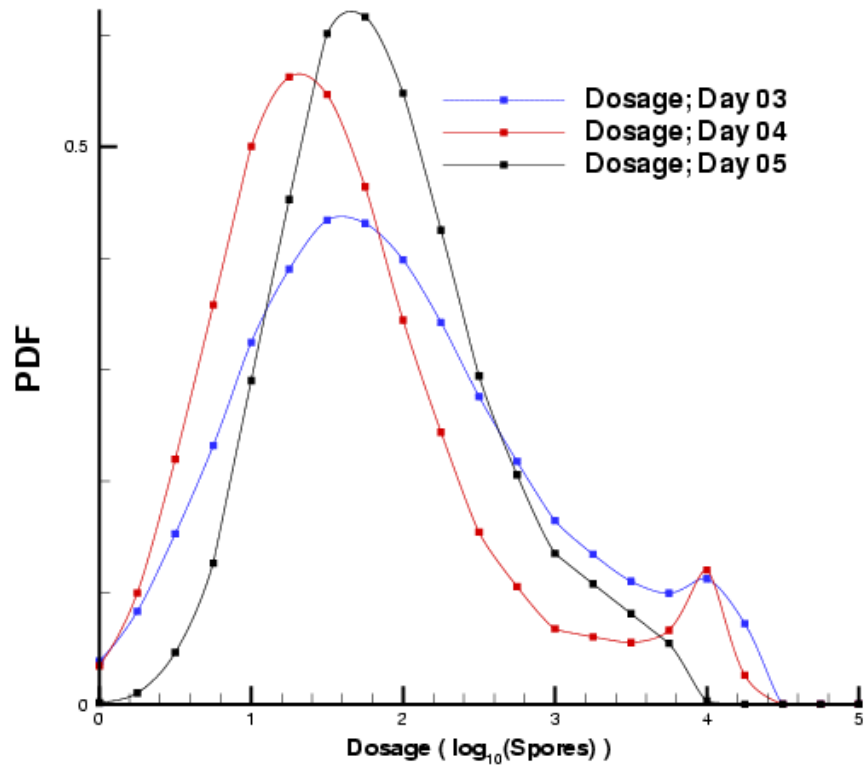
Case E : $N = 10,000$, $\tau = -1.0$,
 $\log_{10}(D) = 2$



Inference of dosage received during attack



Case B : $N = 100$, $\tau = -2.25$,
 $\log_{10}(D) = 2$

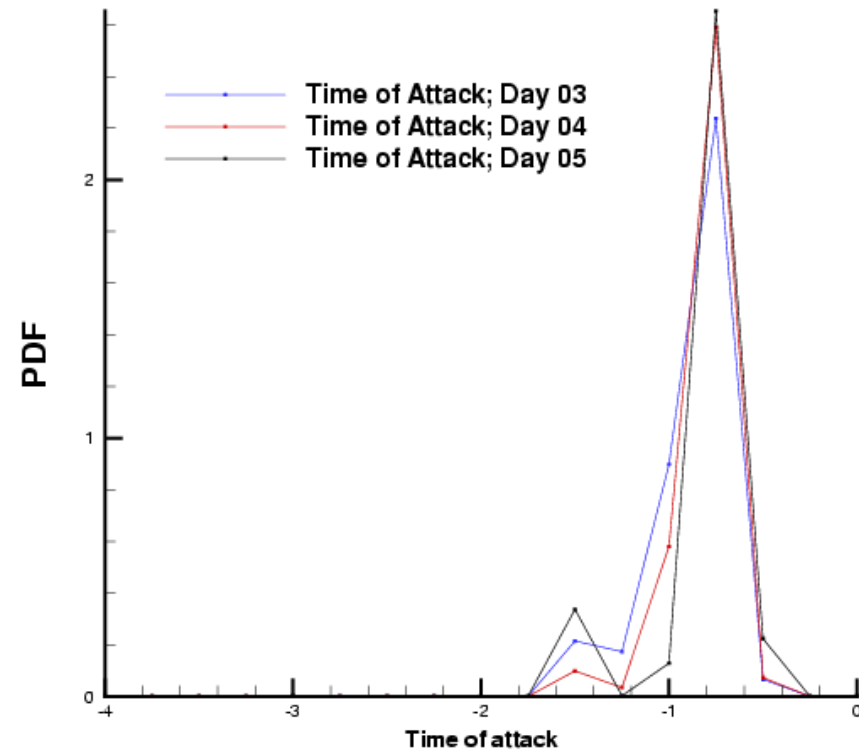
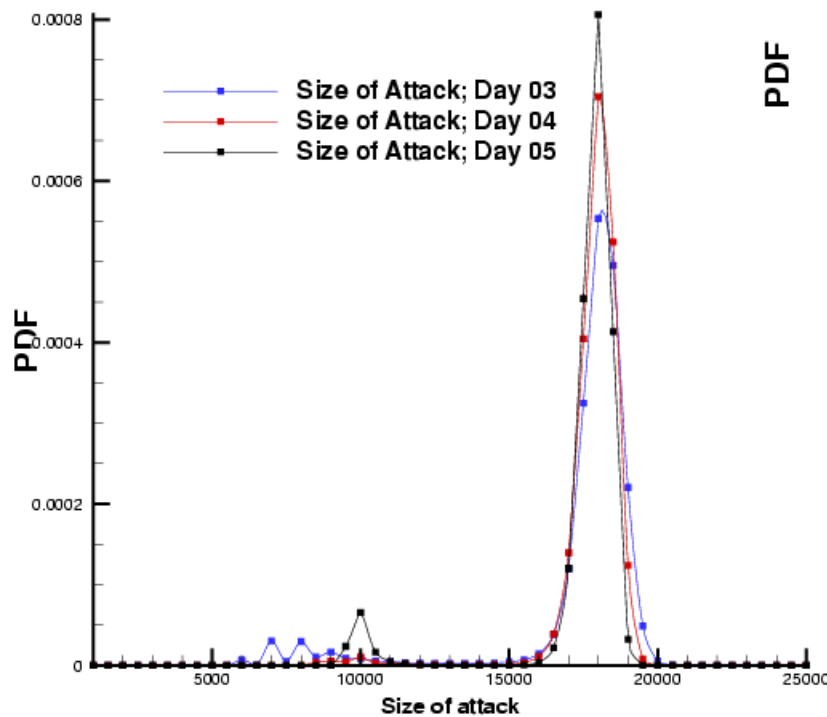


Case E : $N = 10,000$, $\tau = -1.0$,
 $\log_{10}(D) = 2$



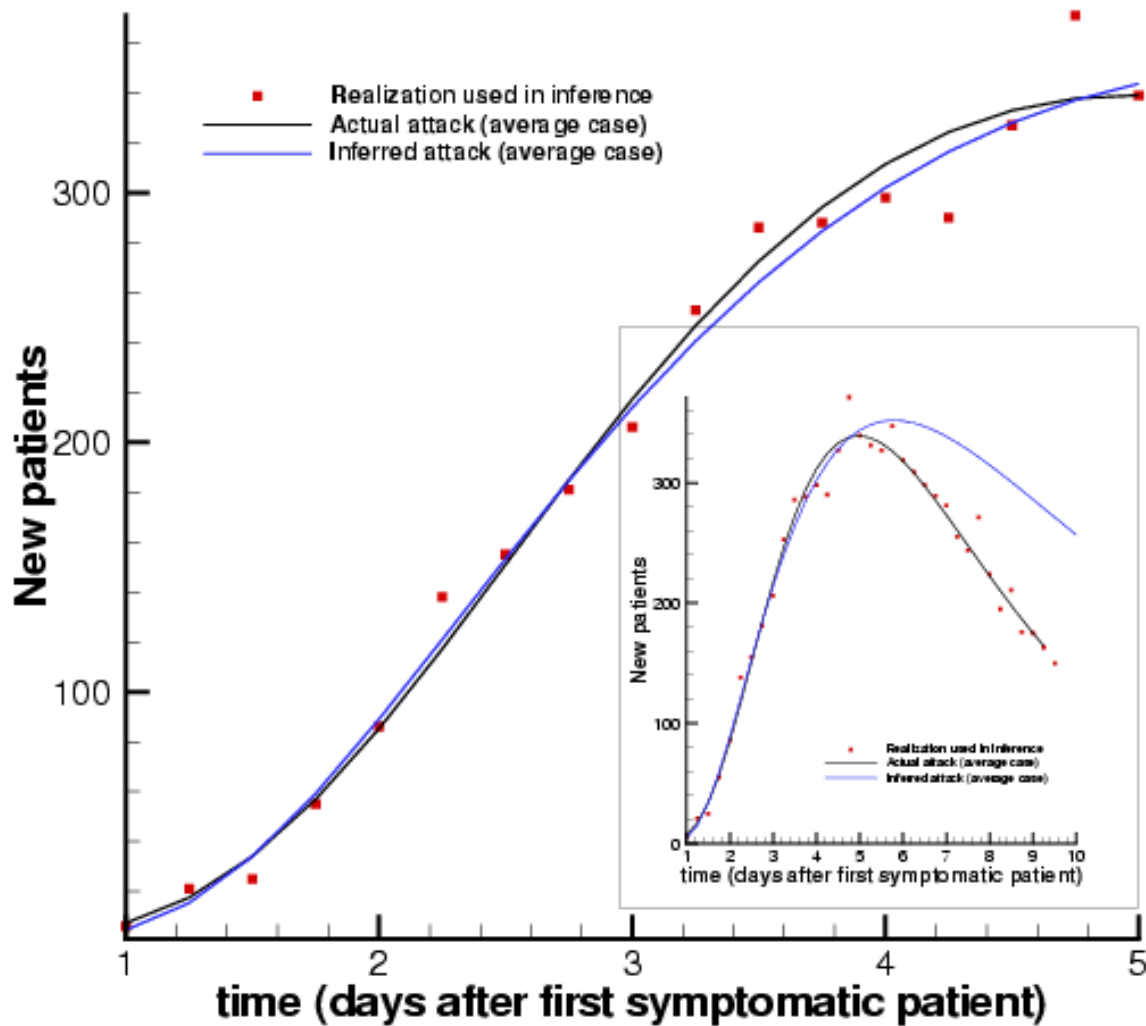
A spectacular failure

- Inferring with partial observations can lead to spectacular failures
- Time series : {2, 369, 938, 1102, 958}
- Attack : $N = 10^4$, $\tau = -1.5$, $D = 10^4$





Why?





Synopsis of the first check

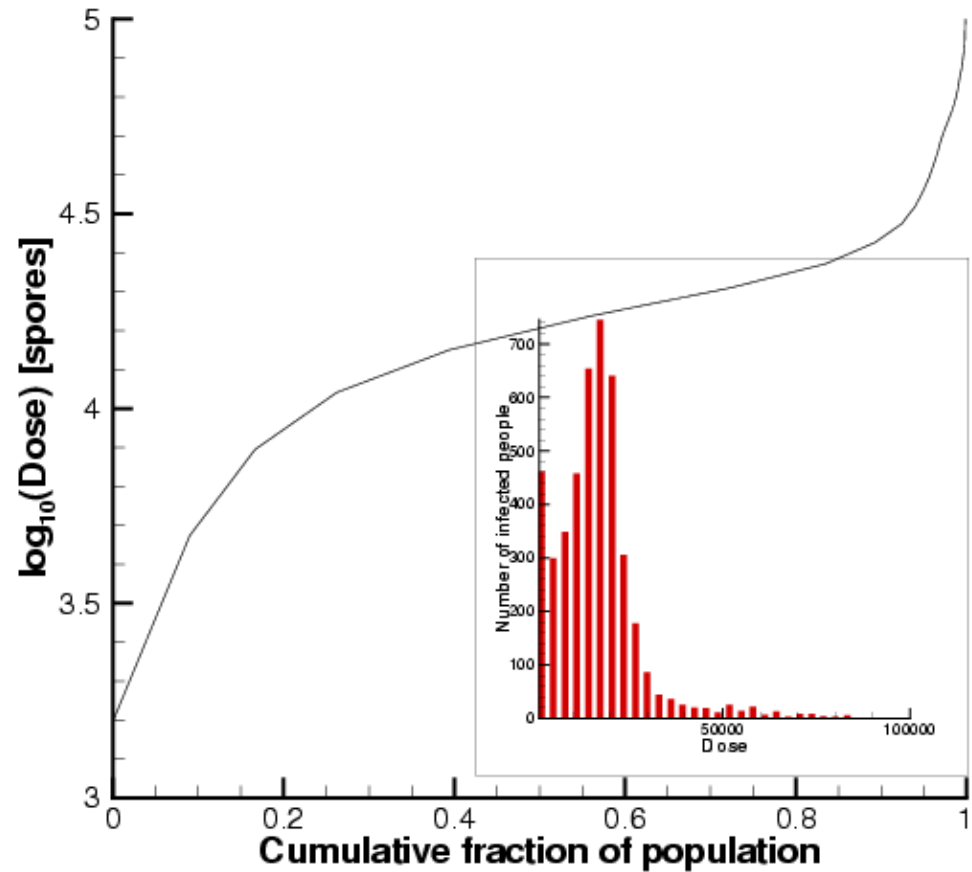
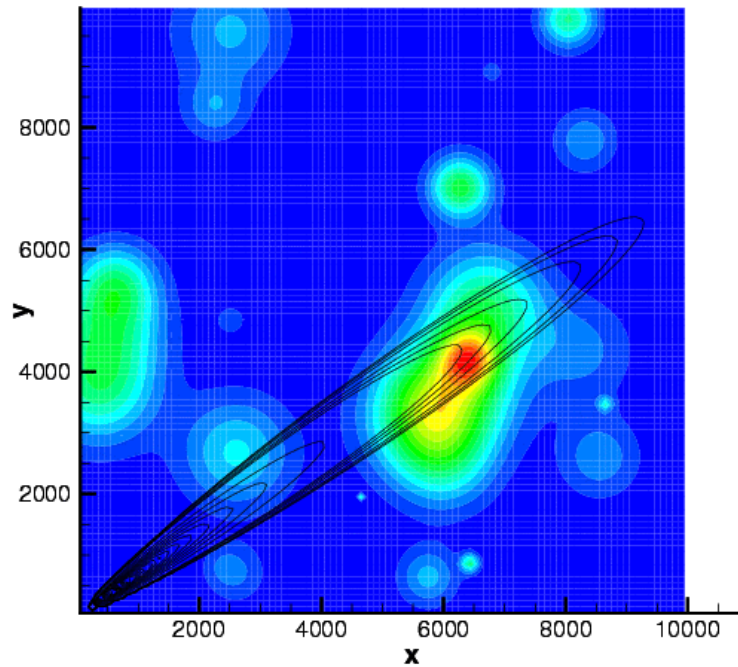
- **Given ideal case (accurate model and uniform dose), the inverse problem**
 - Reliably infers size and time
 - Dosage is hard for small attacks
 - Large attacks are easier to infer
 - Characterizations can go wrong when based on incomplete observations, *but....*
 - Always recovers to correct one when more data becomes available.
- **The method is mathematically consistent, *but....***
- ***Is it useful / applicable in non-ideal cases?***



Simulated attack example

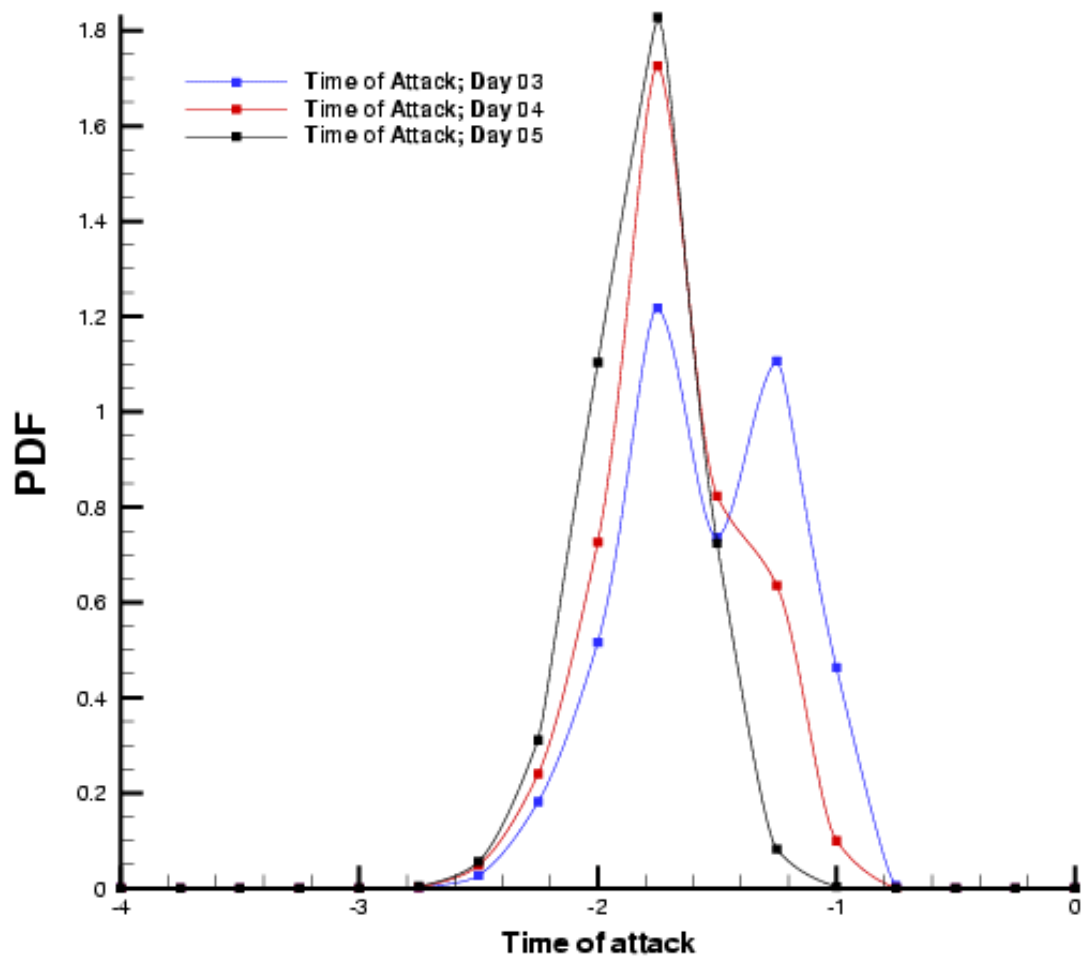
- **Simulated attack**

- Case: $N = 453$, $t = -0.75$, $\log_{10}(\langle D \rangle) = 4.23$
- Time series: $\{1, 36, 57, 55, 56\}$



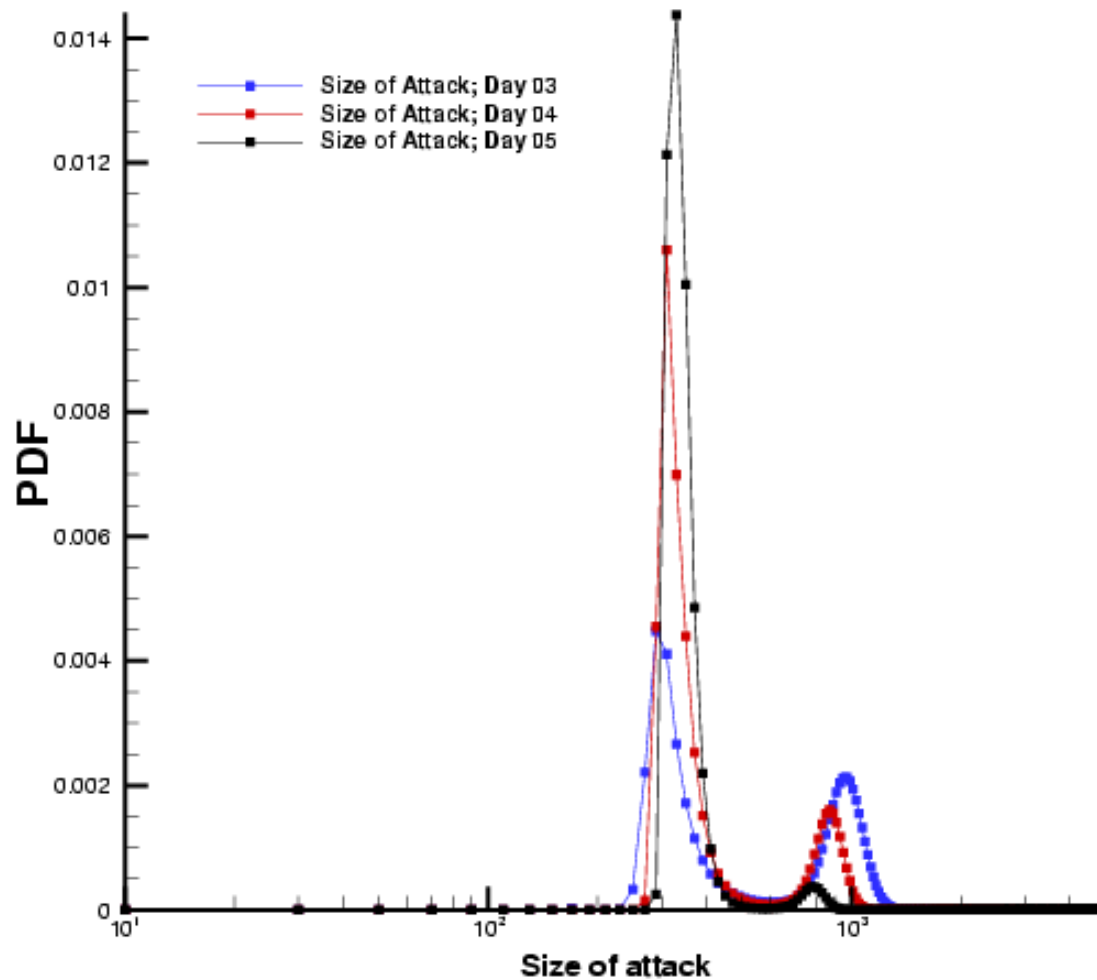


Comparison of inferred time



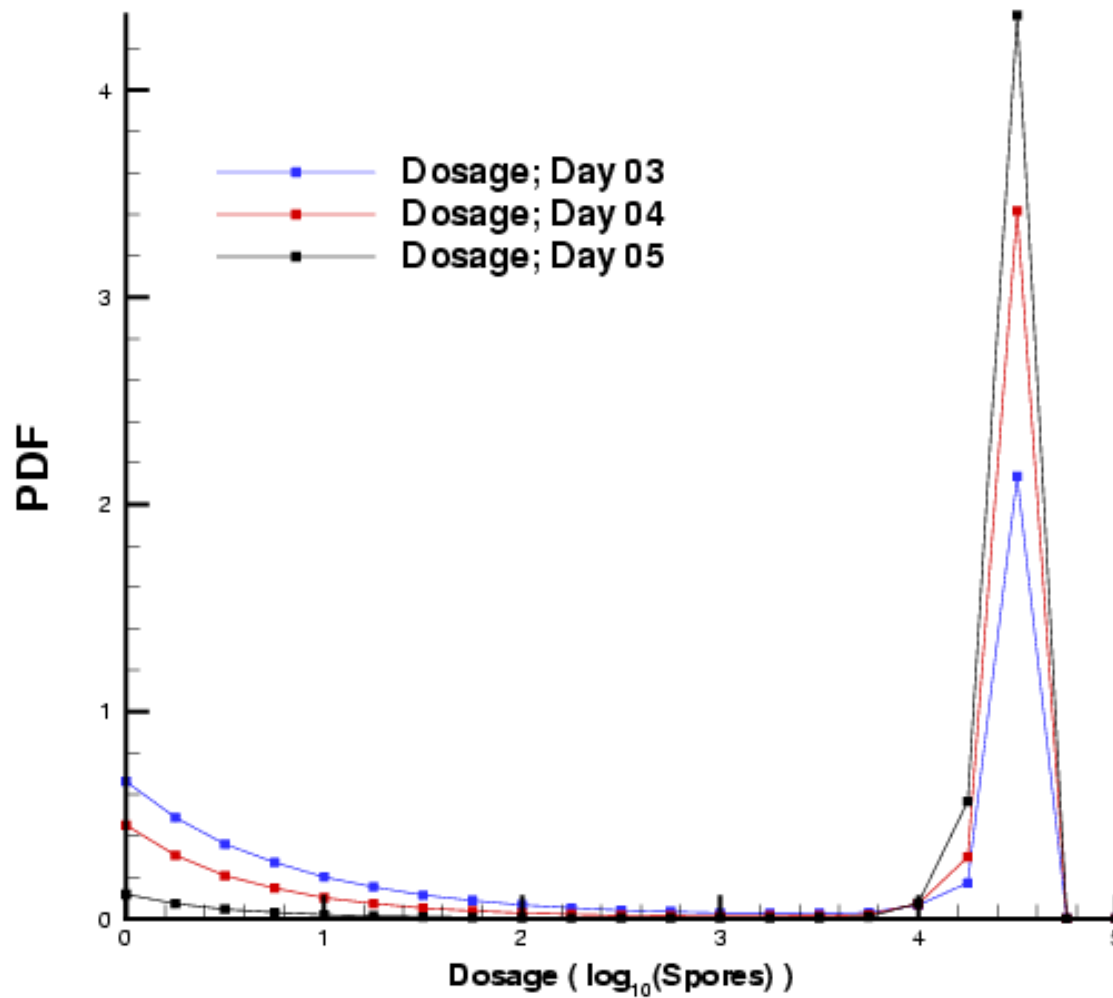


Comparison of inferred size





Comparison of inferred dosage



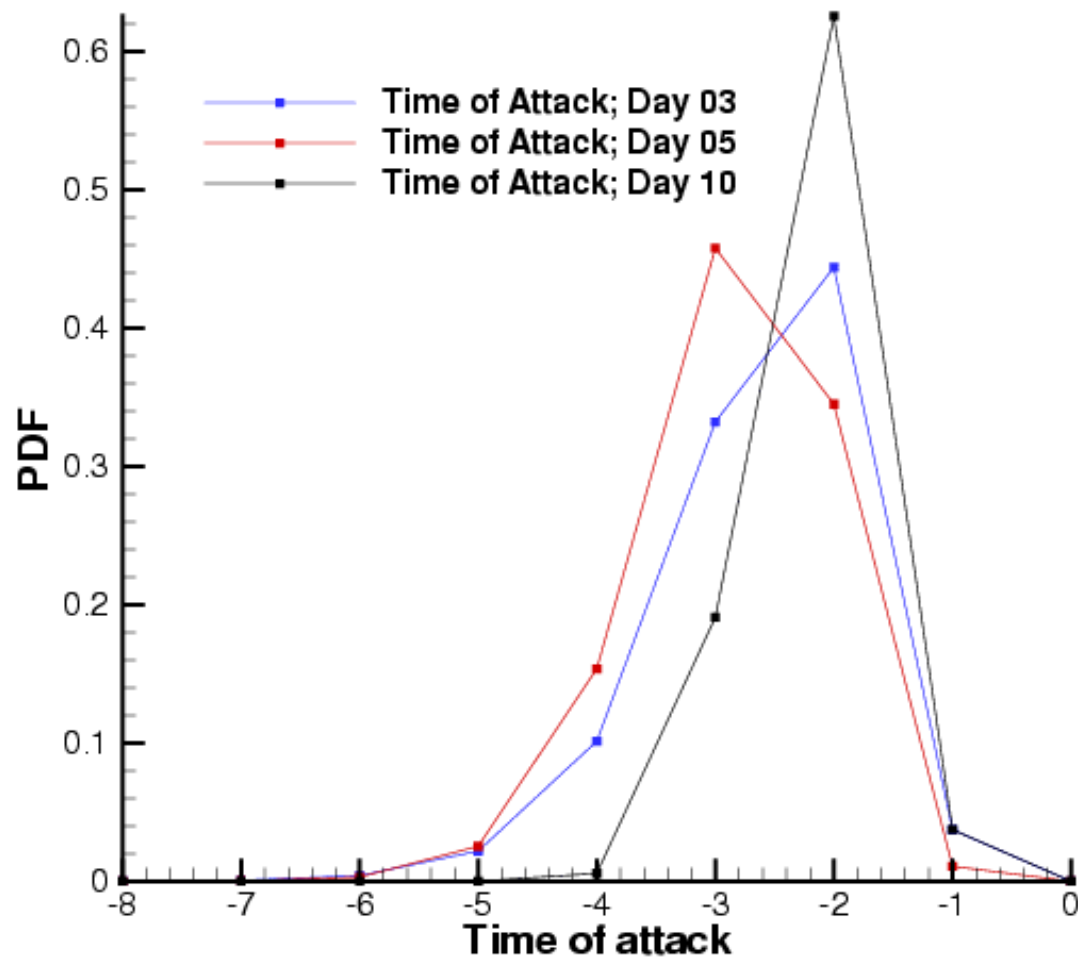


Sverdlovsk, 1979

- **Suspected atmospheric release of weapon-grade anthrax formulation from a military compound**
 - Estimated date : April 2nd, 1979.
 - First symptomatic: April 4th, 1979
 - Estimated number of infected people: 75 ; 70 died
- **Challenges**
 - Small size
 - Reconstructed data
 - Low dose; estimated dose per person:
 - 9 spores (Meselson, *Science*, 1994, using Glassman's numbers)
 - 1-10 spores (Wilkening, PNAS, 103(20), 2006)
 - Effect of prophylaxis (initiated April 12th, 1979)
 - Vaccination (started : April 15th, 1979 (approx))

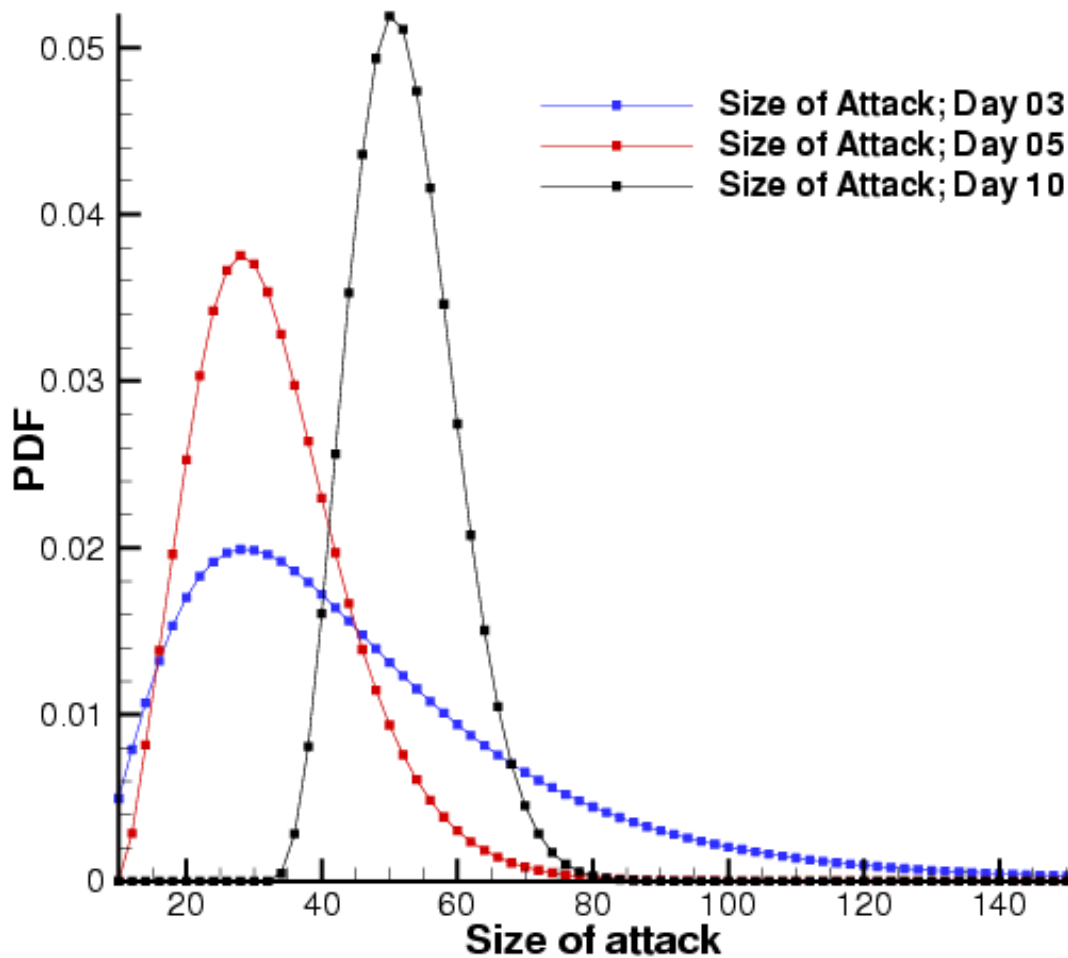


Sverdlovsk, 1979 - Time of infection



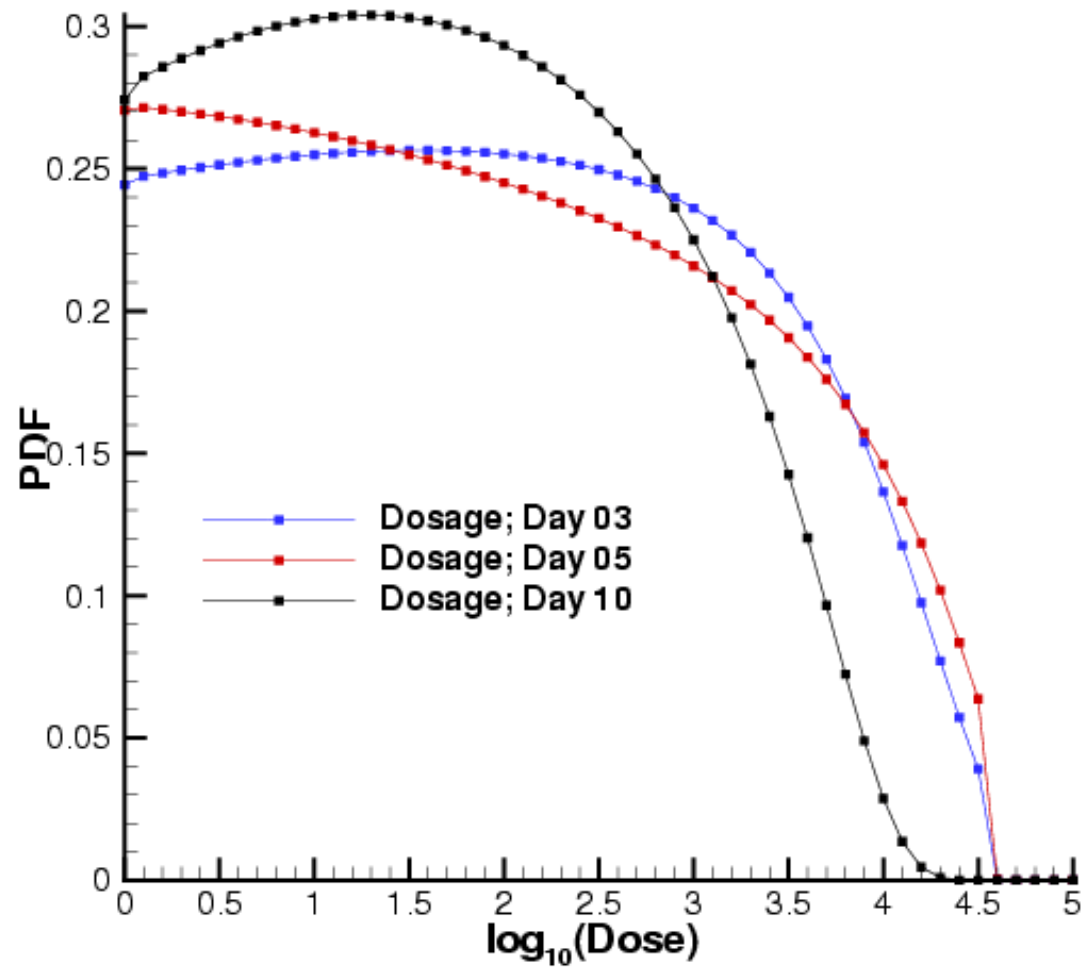


Sverdlovsk, 1979 – Size of infected population





Sverdlovsk, 1979 – Dosage





Conclusions

- **We have**
 - A rigorous Bayesian formulation to characterize bioterrorist attacks (anthrax)
 - Can be extended to smallpox, plague and other disease with a *symptomatic* contagious period.
- **We need, in short order,**
 - To bring in a spatial component into the inverse problem,
 - Ditto, contagious diseases
- **Ultimately, need to design a risk-based response plan**
 - Characterization not very useful if the cavalry rides in every time someone sneezes.
- **More Information :**
 - Ray et al, “A Bayesian method for characterizing distributed micro-releases”, Sandia Technical Report, SAND2006-7568, Printed January 2007. Unclassified, unlimited release.