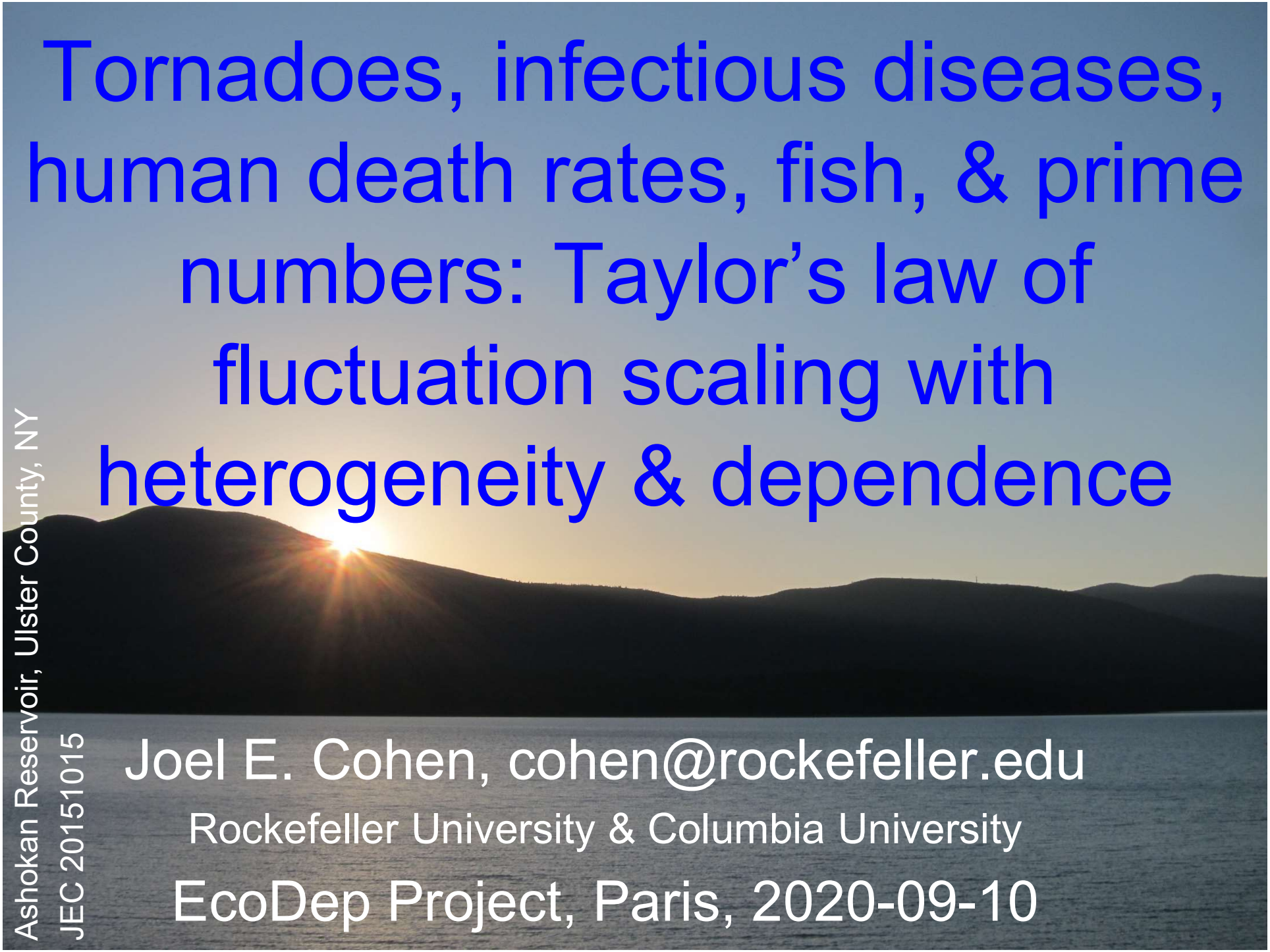
Tornadoes, infectious diseases, human death rates, fish, & prime numbers: Taylor's law of fluctuation scaling with heterogeneity & dependence



Ashokan Reservoir, Ulster County, NY
JEC 20151015

Joel E. Cohen, cohen@rockefeller.edu
Rockefeller University & Columbia University

EcoDep Project, Paris, 2020-09-10

Copyright page

Copyright © 2020 by Joel E. Cohen

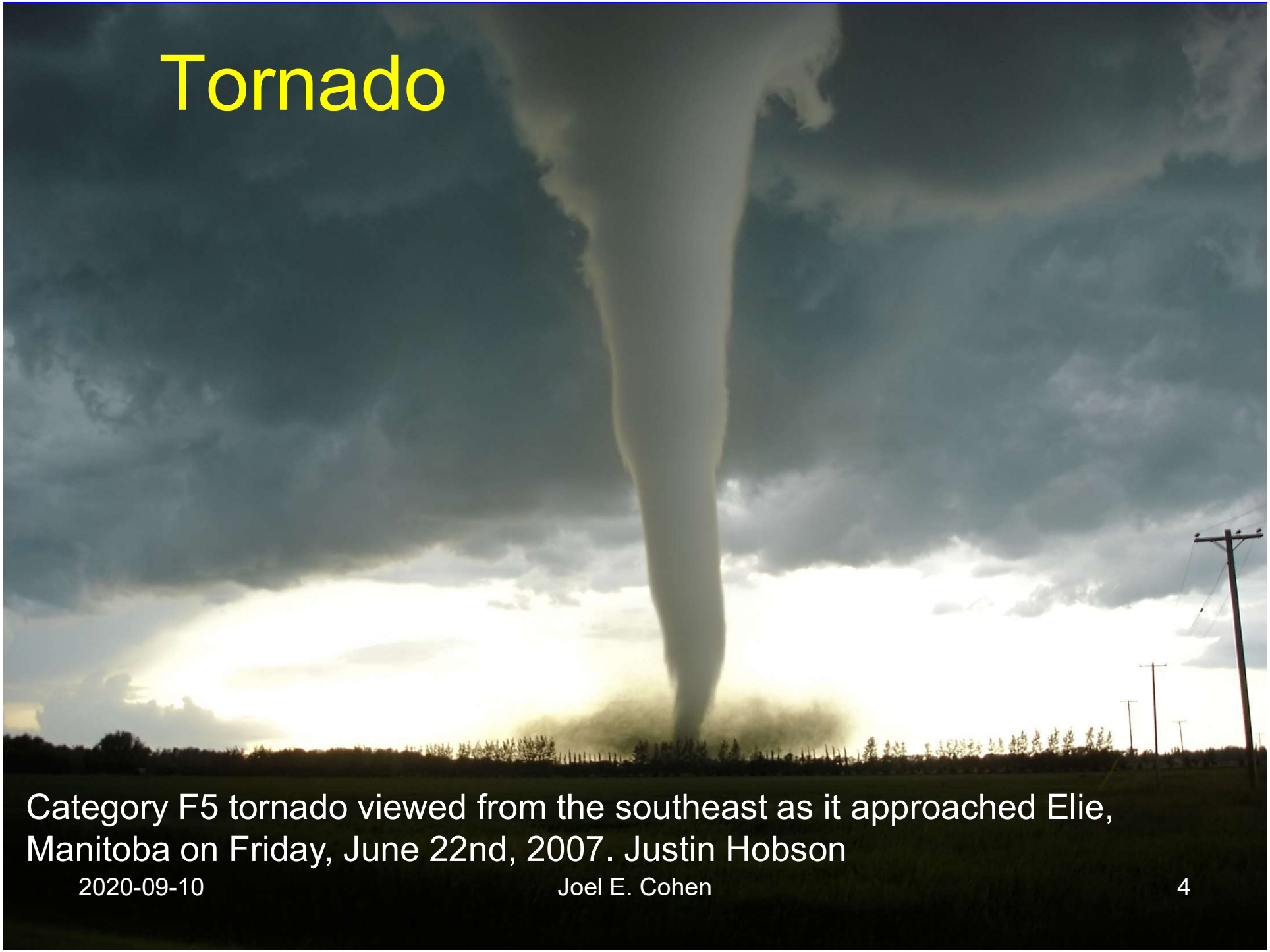
Certain copyrighted materials are included under "fair use" for educational, non-commercial purposes. The sources of these materials are credited in every case.

The photographs & text of these slides are not to be reprinted or further distributed in any medium, including via the Internet.

Outline

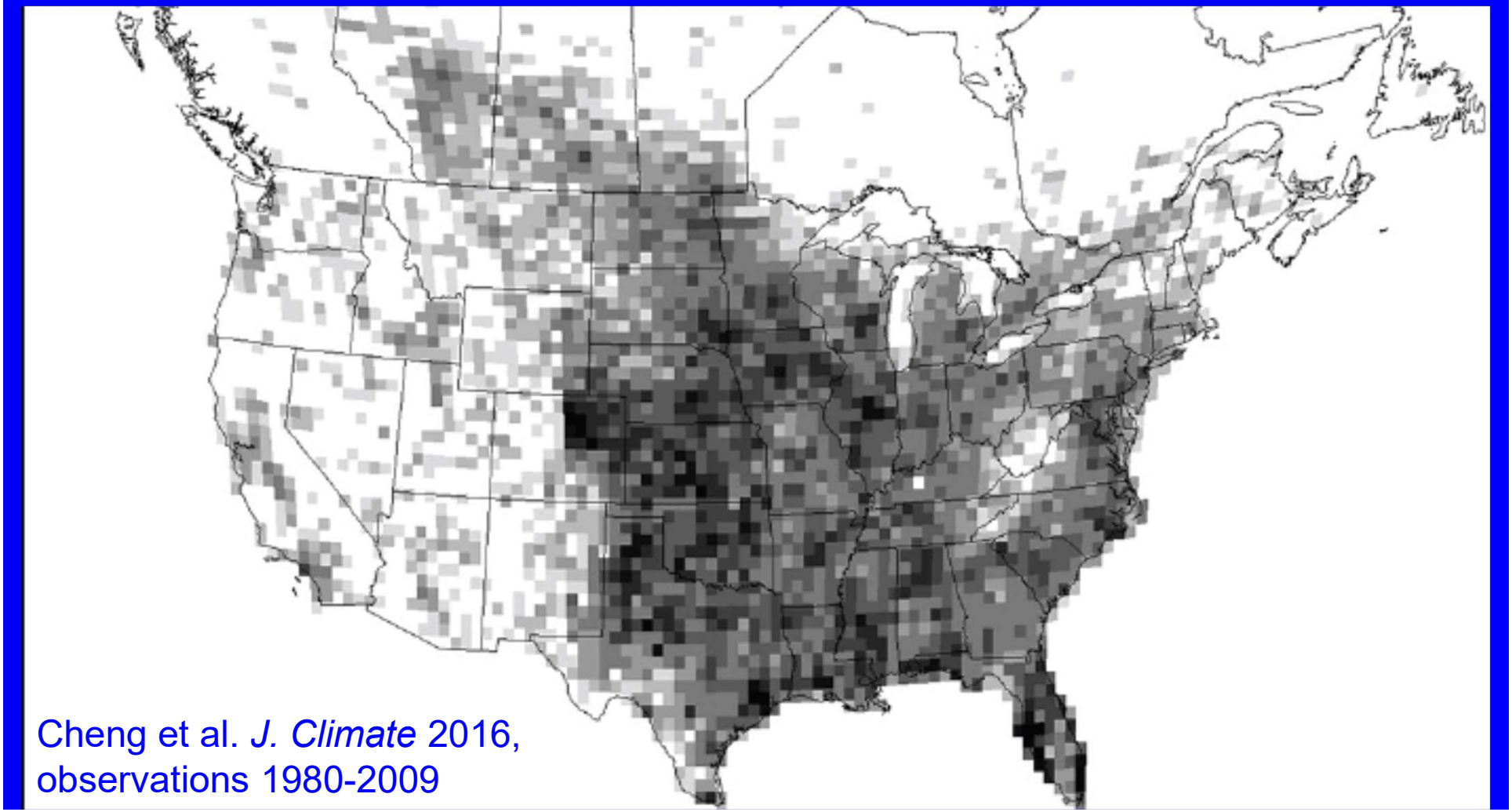
1. Examples
2. What is Taylor's law (TL)?
3. Recent mathematical results
4. Conclusions

Tornado



Category F5 tornado viewed from the southeast as it approached Elie, Manitoba on Friday, June 22nd, 2007. Justin Hobson

USA has more tornadoes than any other country. (Lloyd's)



Cheng et al. *J. Climate* 2016,
observations 1980-2009

Outbreaks (≥6 tornadoes) cause most damage.

Outbreak is defined as ≥6 tornadoes starting ≤6 hours apart.

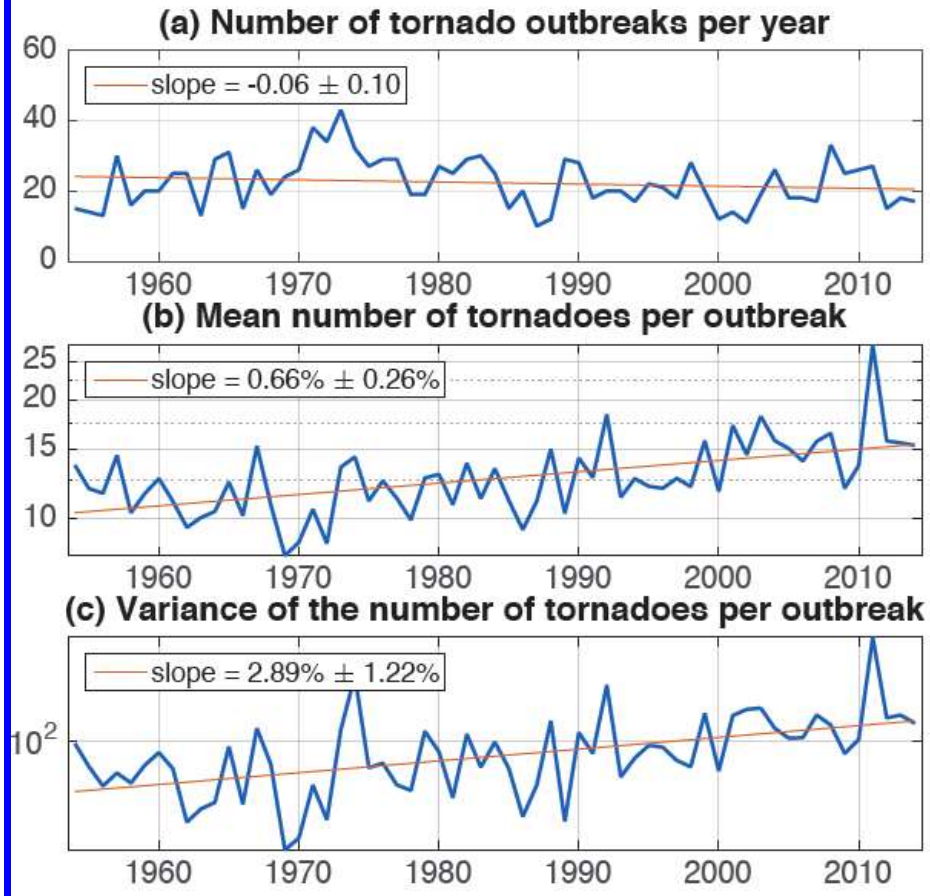
1972–2010: 79% of tornado fatalities & most economic losses occurred in outbreaks.

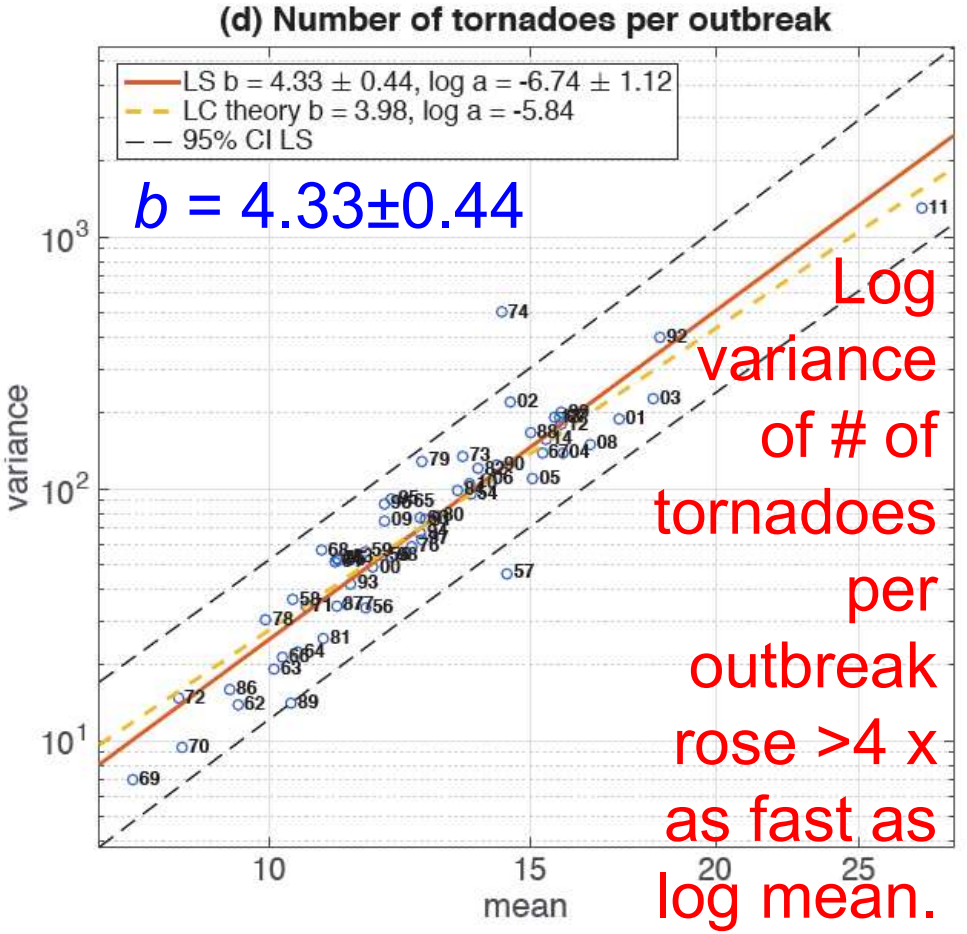
No trends in numbers of reliably reported tornadoes or outbreaks in last half century.

Mean & variance of number of tornadoes per outbreak, & insured losses, increased significantly in last half century.

F1+ tornadoes per outbreak in USA: variance~(mean)$^{4.3}$

Tippett & Cohen, *Nature Communications* 2016

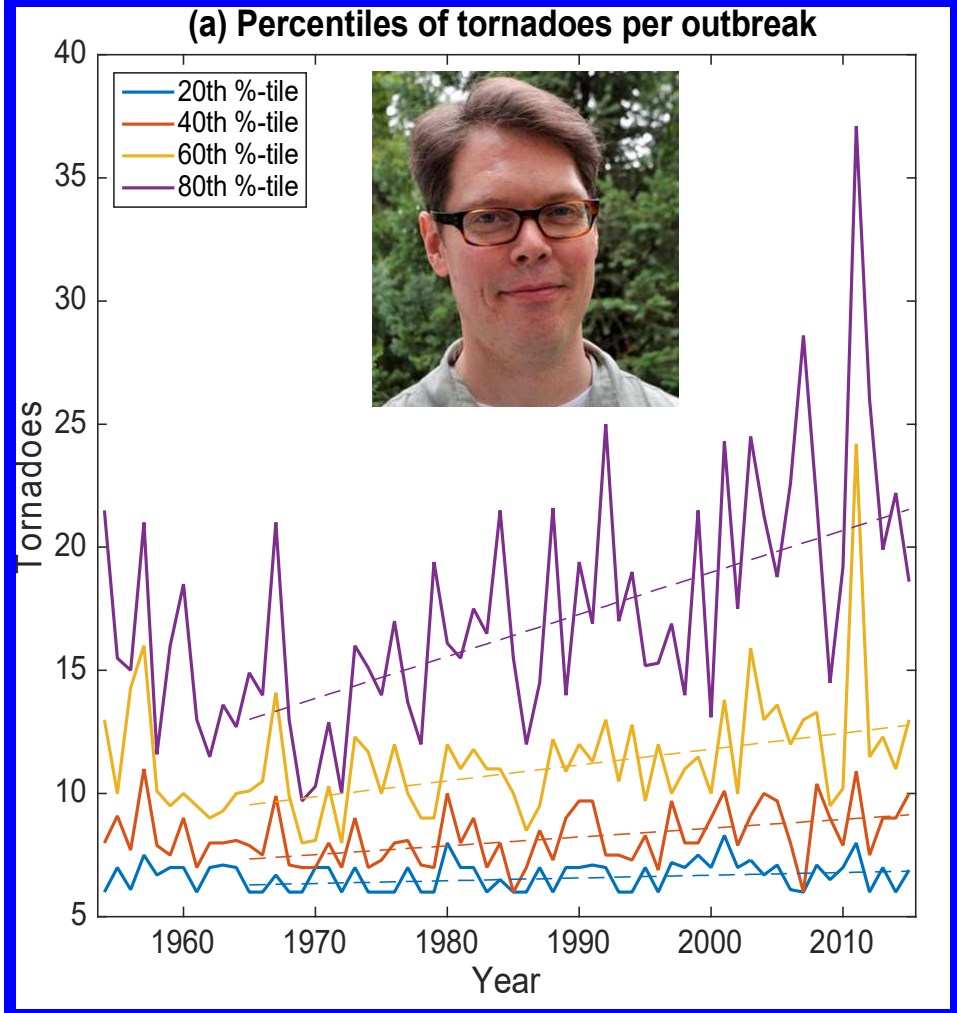


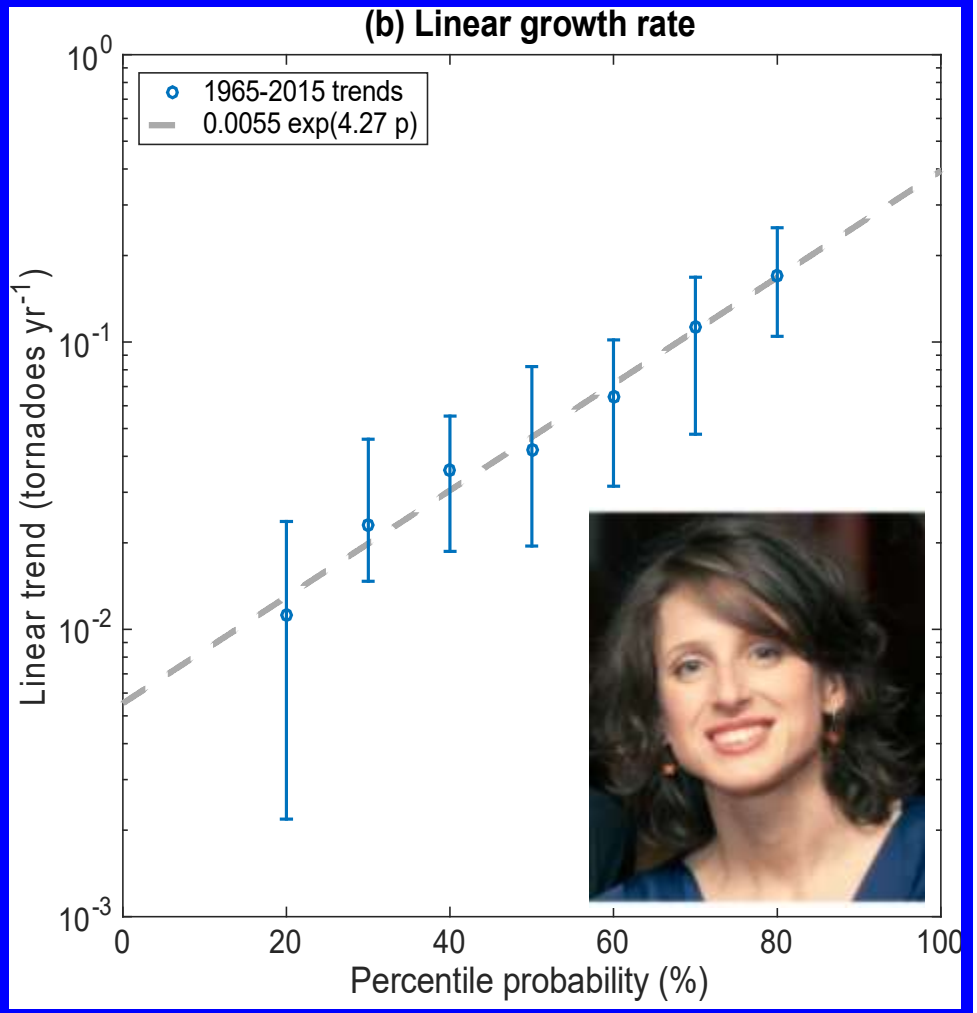


Log variance of # of tornadoes per outbreak rose >4 x as fast as log mean.

Higher percentiles increased faster.

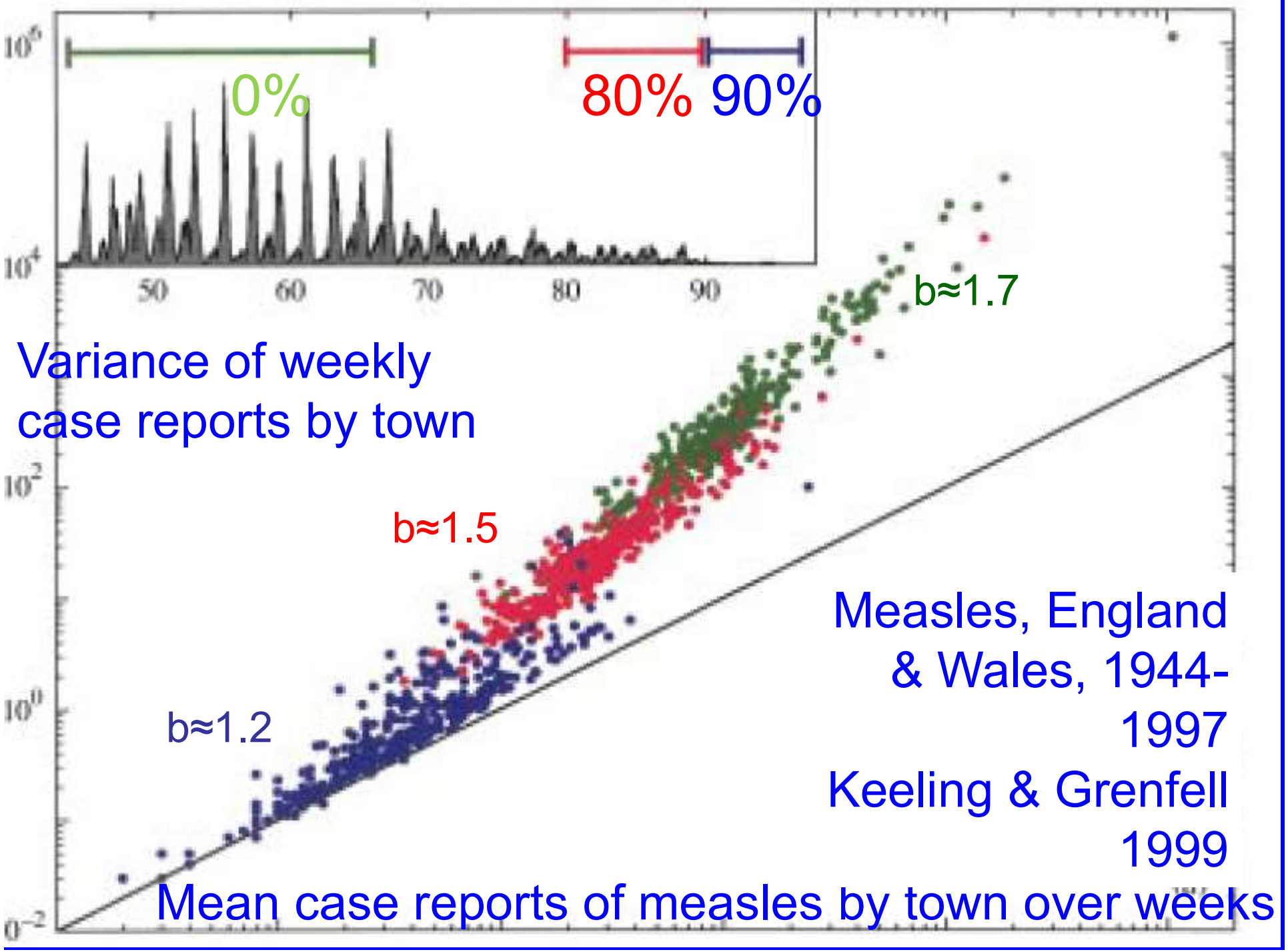
"quantile regression"





Tippett, Lepore, Cohen *Science* 2016

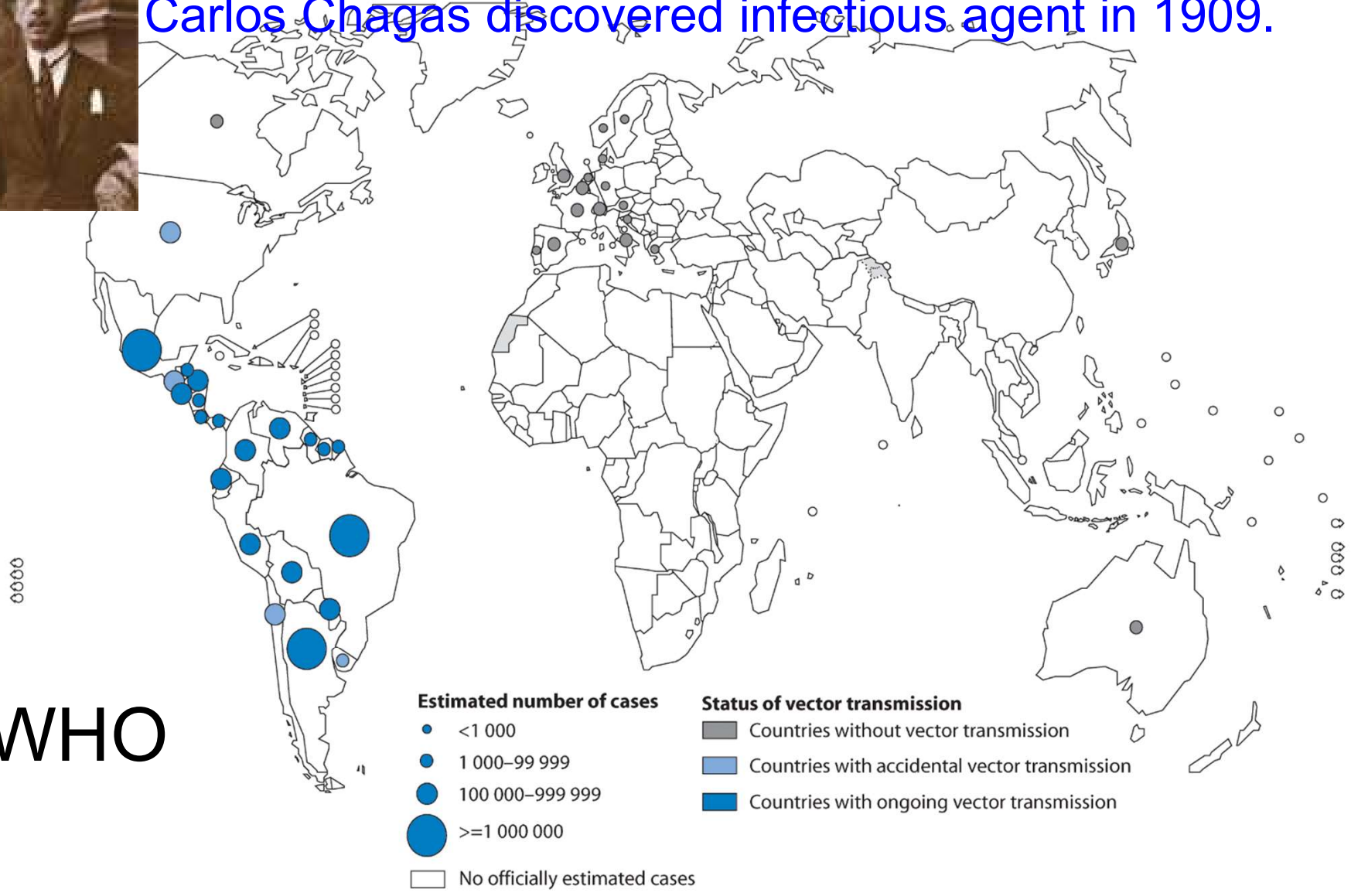
Infectious diseases


0%
80% 90%
50
60
70
80
90
Variance of weekly
case reports by town
b≈1.7
b≈1.5
b≈1.2
Measles, England
& Wales, 1944-
1997
Keeling & Grenfell
1999
Mean case reports of measles by town over weeks

6-7 million cases of Chagas disease 2019

Carlos Chagas discovered infectious agent in 1909.

Distribution of cases of *Trypanosoma cruzi* infection, based on official estimates and status of vector transmission, worldwide, 2006–2009

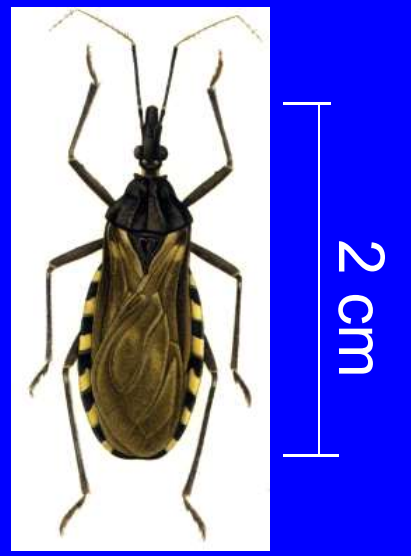


WHO

Parasite & vectors are widespread.

Parasite:
Trypanosoma cruzi

Vectors: multiple species, mainly *Triatoma infestans*



Source:
http://www.cpqrr.fiocruz.br/laboratorios/lab_triato/TriatoInfestans.html

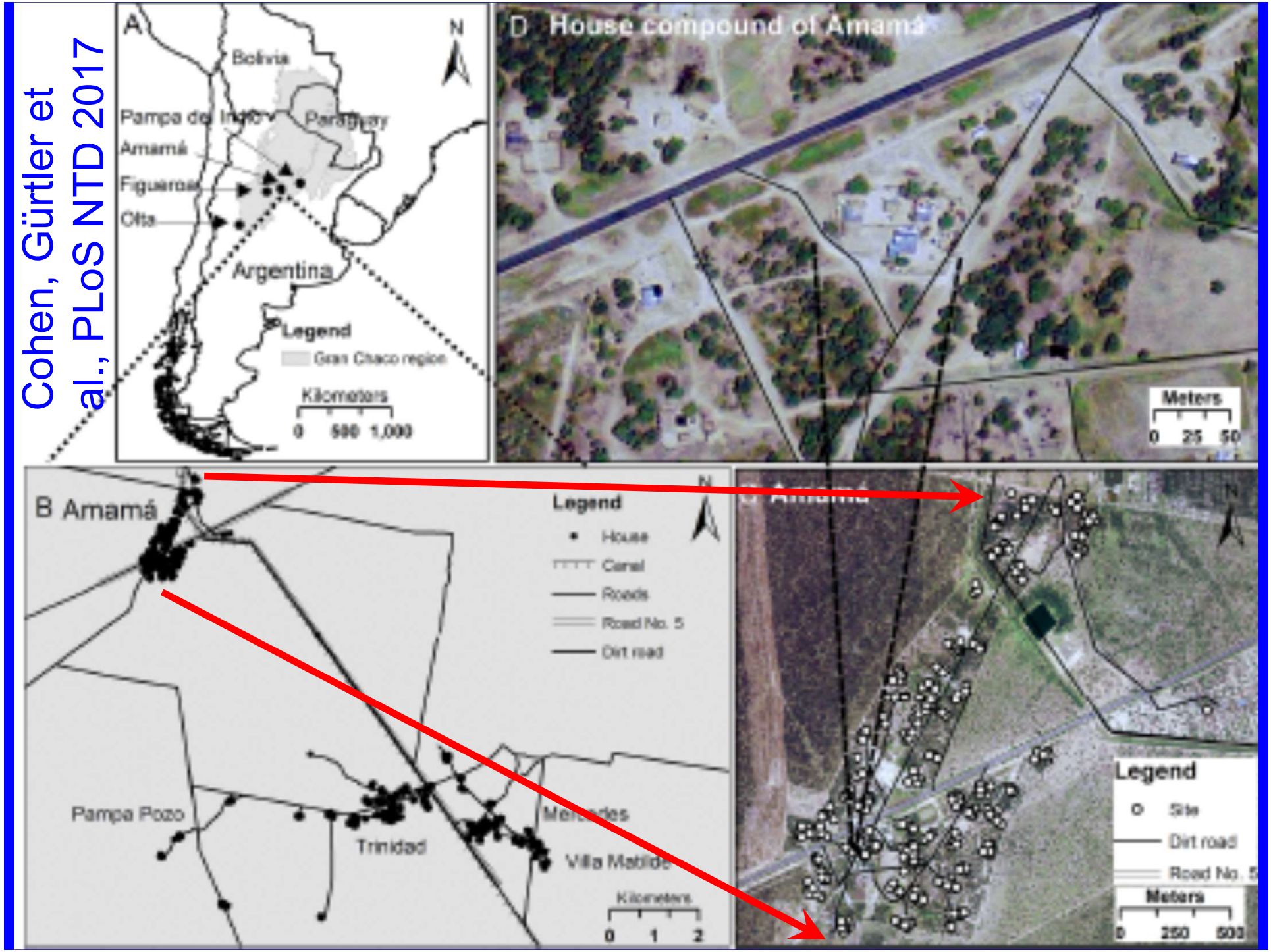
Clinical beginning & end

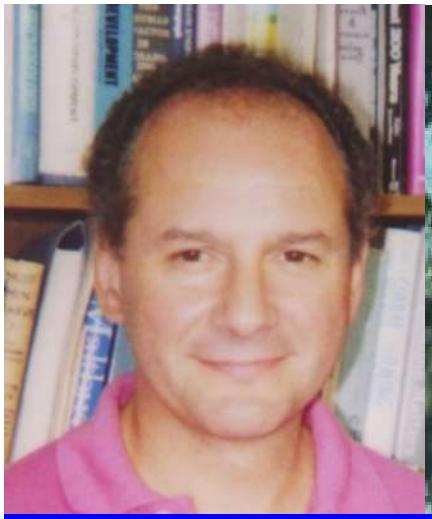
in 20-30%

Beginning:

Romaña's sign

End: megacolon, megaesophagus


Cohen, Gürtler et al., PLoS NTD 2017
A
Bolivia
Pampa del Indio
Paraguay
Amamá
Figueroa
Olta
Argentina
Legend
Gran Chaco region
Kilometers
0 500 1,000
N
D House compound of Amamá
Meters
0 25 50
B Amamá
Legend
House
Canal
Roads
Road No. 5
Dirt road
Pampa Pozo
Trinidad
Villa Matilde
Kilometers
0 1 2
Legend
Site
Dirt road
Meters
0 250 500

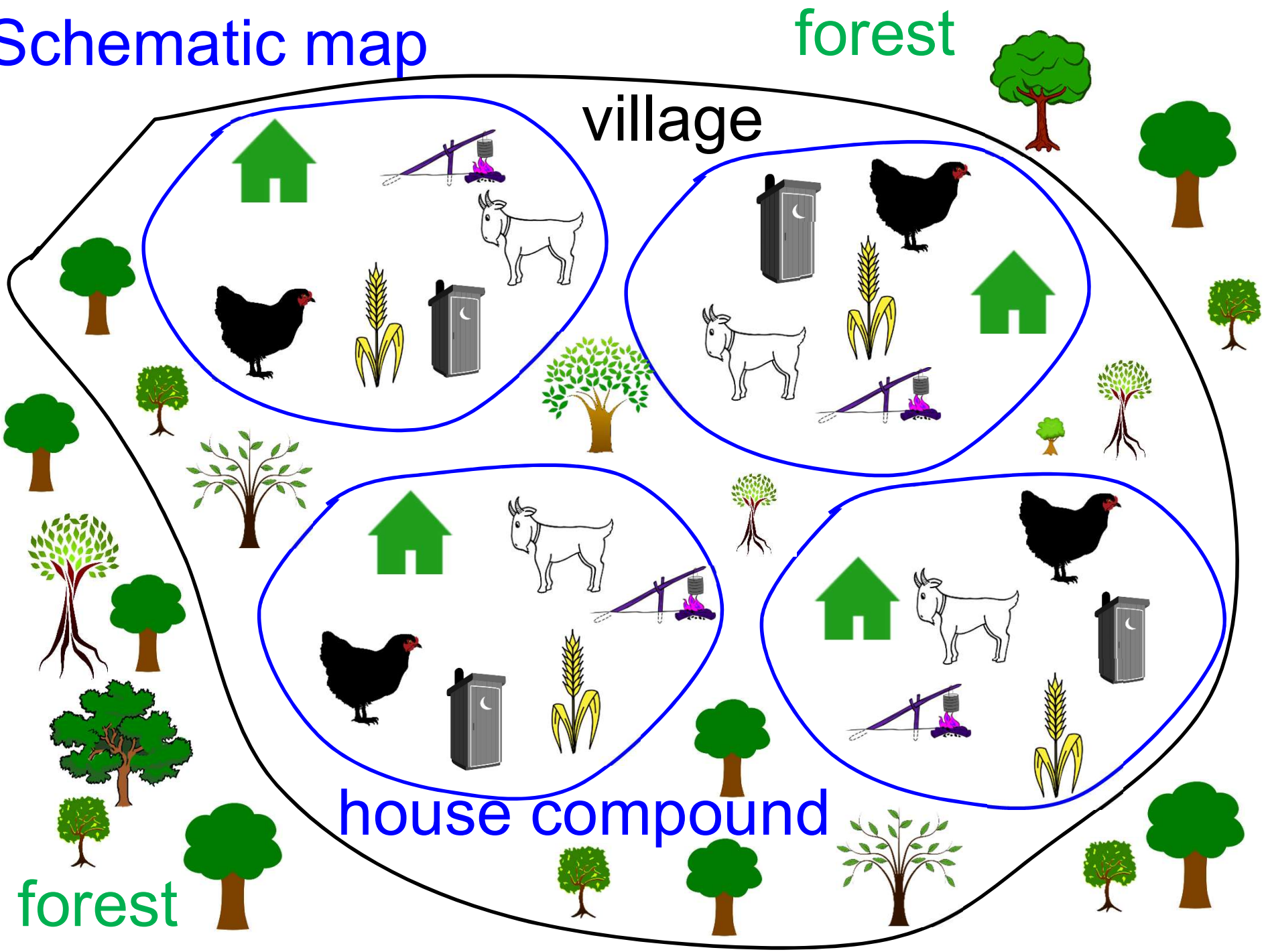


Ricardo E. Gürtler

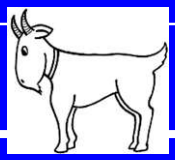
Homes & out-buildings are built of mud, sticks, & thatch, ideal for bugs.
Prov. Santiago del Estero, NW Argentina

Chagas' disease

Bug bites dog infected with *T. cruzi*, infected bug bites boy.

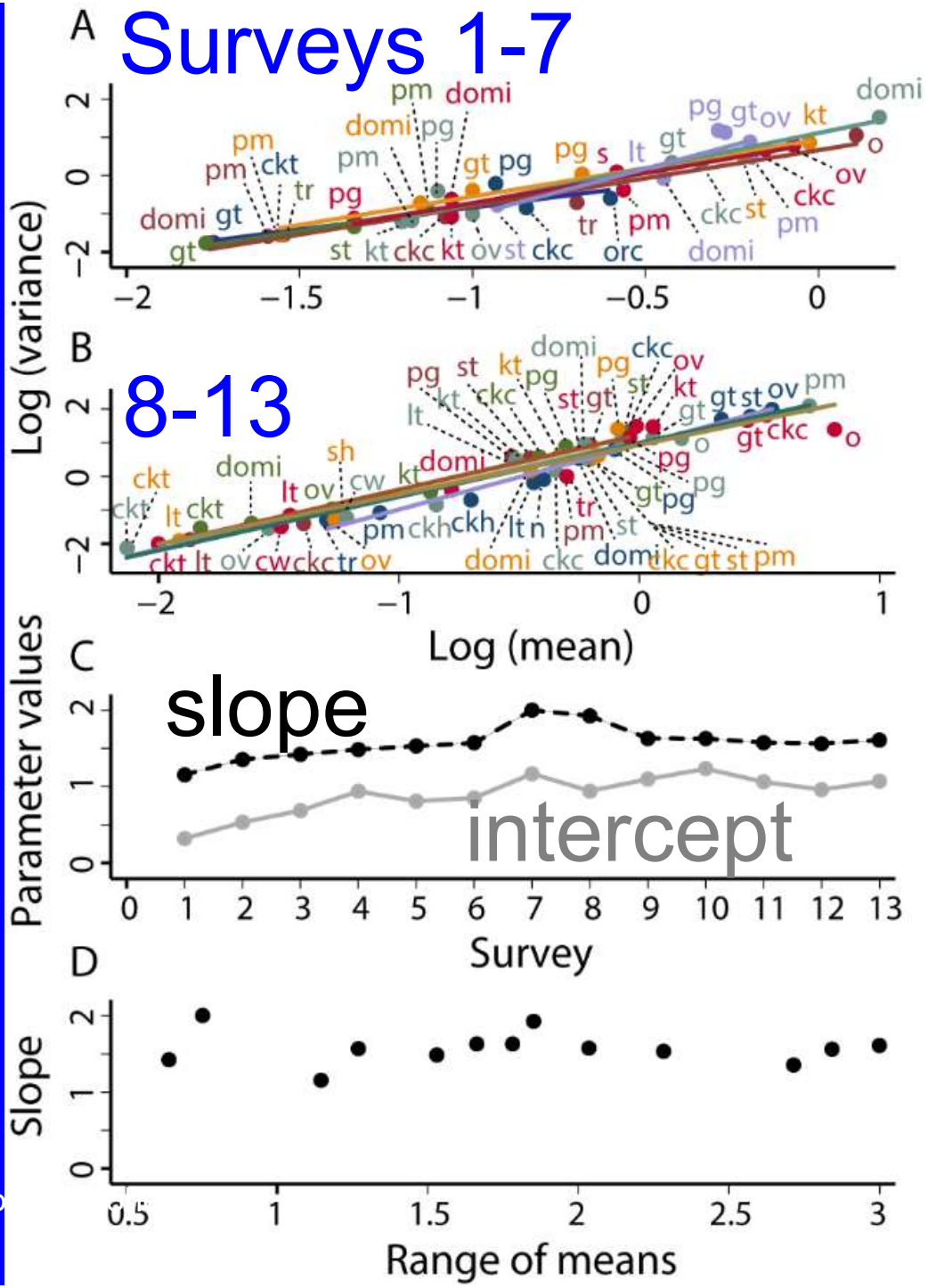

Schematic map
forest
village
house compound
forest

Each habitat (domicile, chicken coop, goat corral, granary) defines one sample.

Habitat→				
Bug population density (per hour of search) in exemplars of this habitat	x_{11}	x_{12}	x_{13}	$x_{...}$
	x_{21}	x_{22}	x_{23}	...
	x_{31}	x_{32}	x_{33}	...
		x_{42}	x_{43}	...
		x_{52}		...
Mean	m_1	m_2	m_3	m...
Variance	v_1	v_2	v_3	v...

Amamá core (sustained surveillance & control): *T. infestans* obeys spatial TL in 13 surveys 1993-2002.

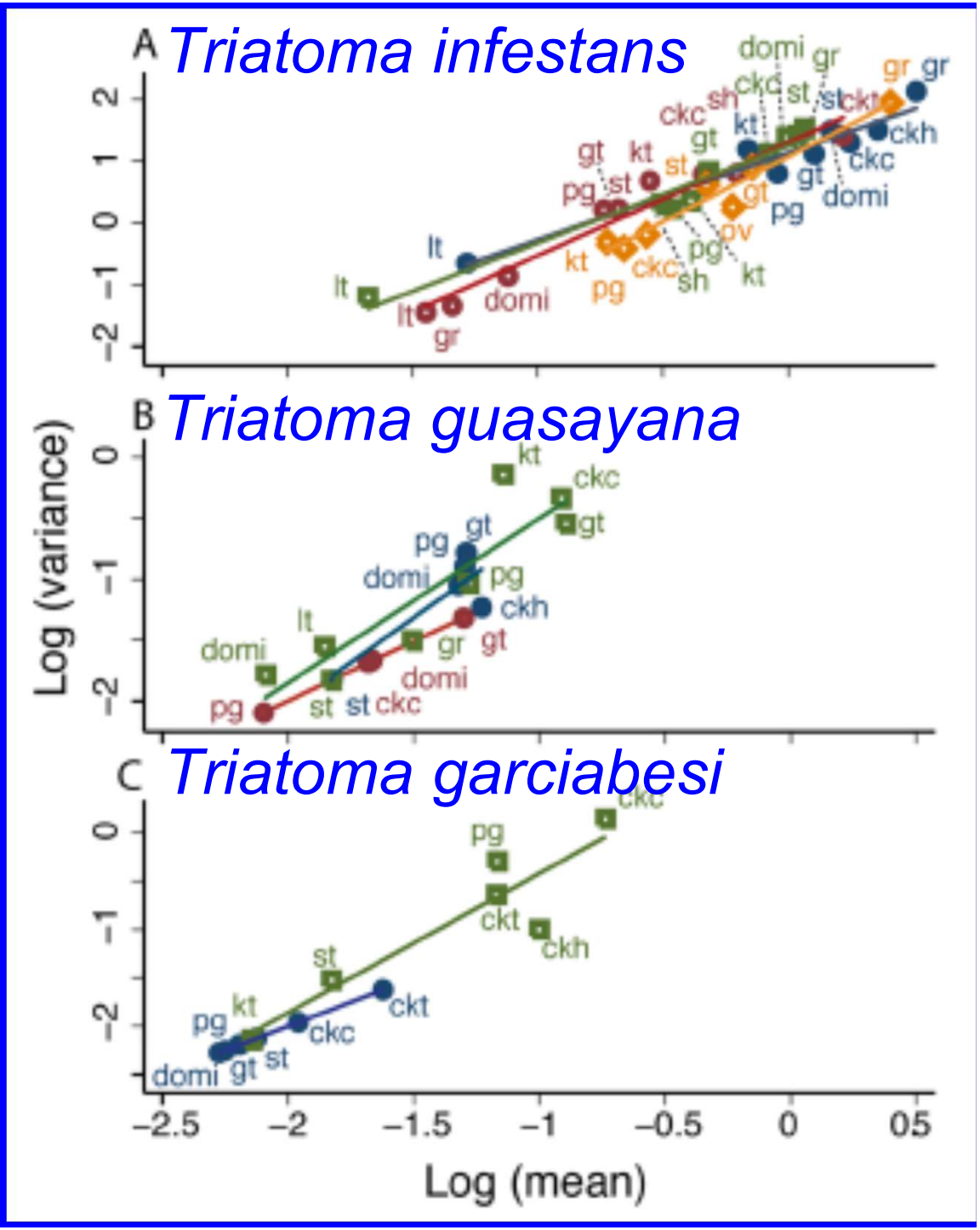
Cohen, Rodríguez-Planes, Gaspe, Cecere, Cardinal, Gürtler, *PLoS Neglected Tropical Diseases* 2017



Figueroa: Vectors of Chagas disease obey spatial TL.

Cohen, Rodríguez-Planes, Gaspe, Cecere, Cardinal, Gürtler, *PLoS Neglected Tropical Diseases* 2017

2020-09-10



Uses of TL in Chagas vector control

Flagging errors in data

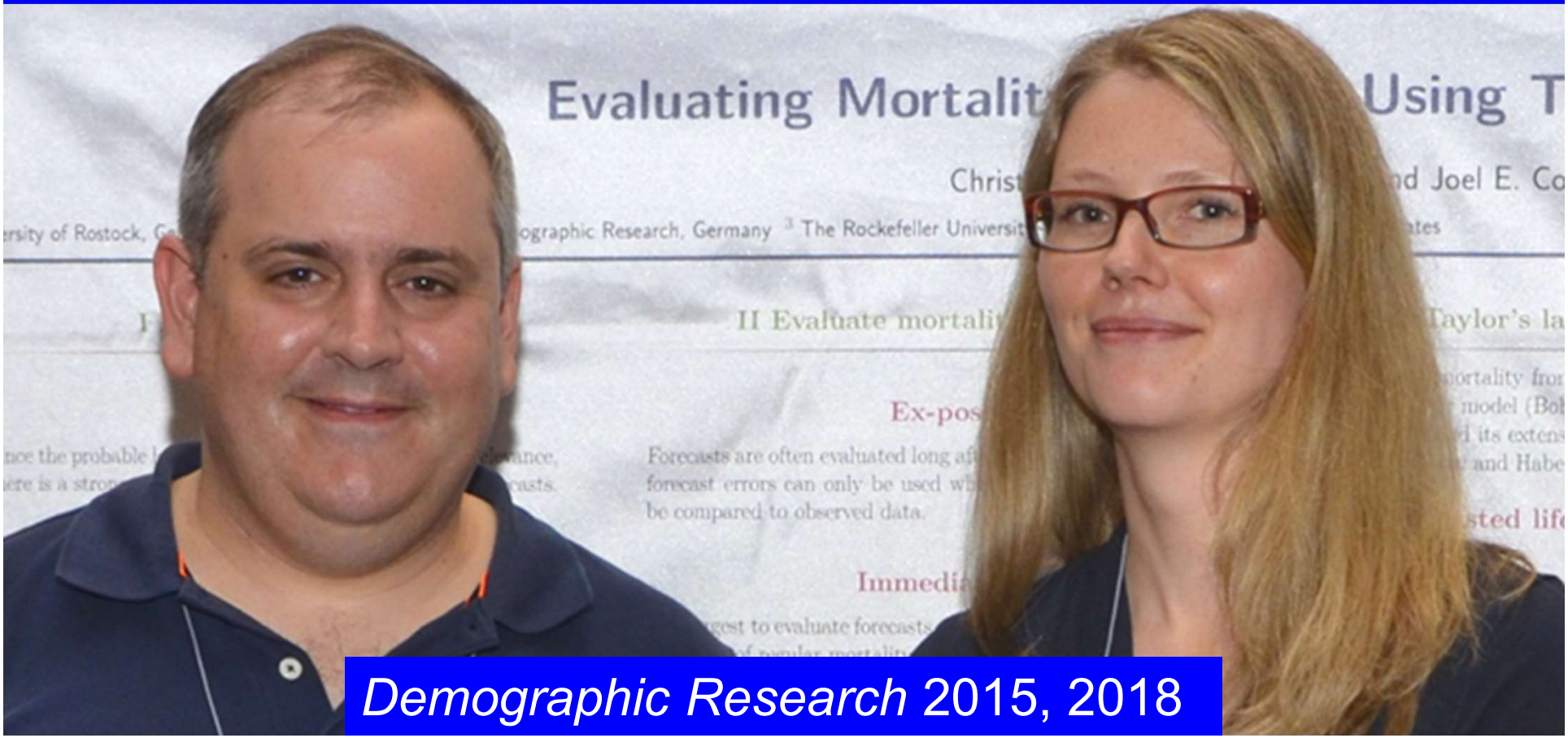
Identifying habitats of unusually high or unusually low variability

Designing sequential sampling to reach specified precision, using stopping line derived from TL

Community education workshop

How do human death rates vary over time, at each age, for each sex?

Roland Rau Christina Bohk-Ewald



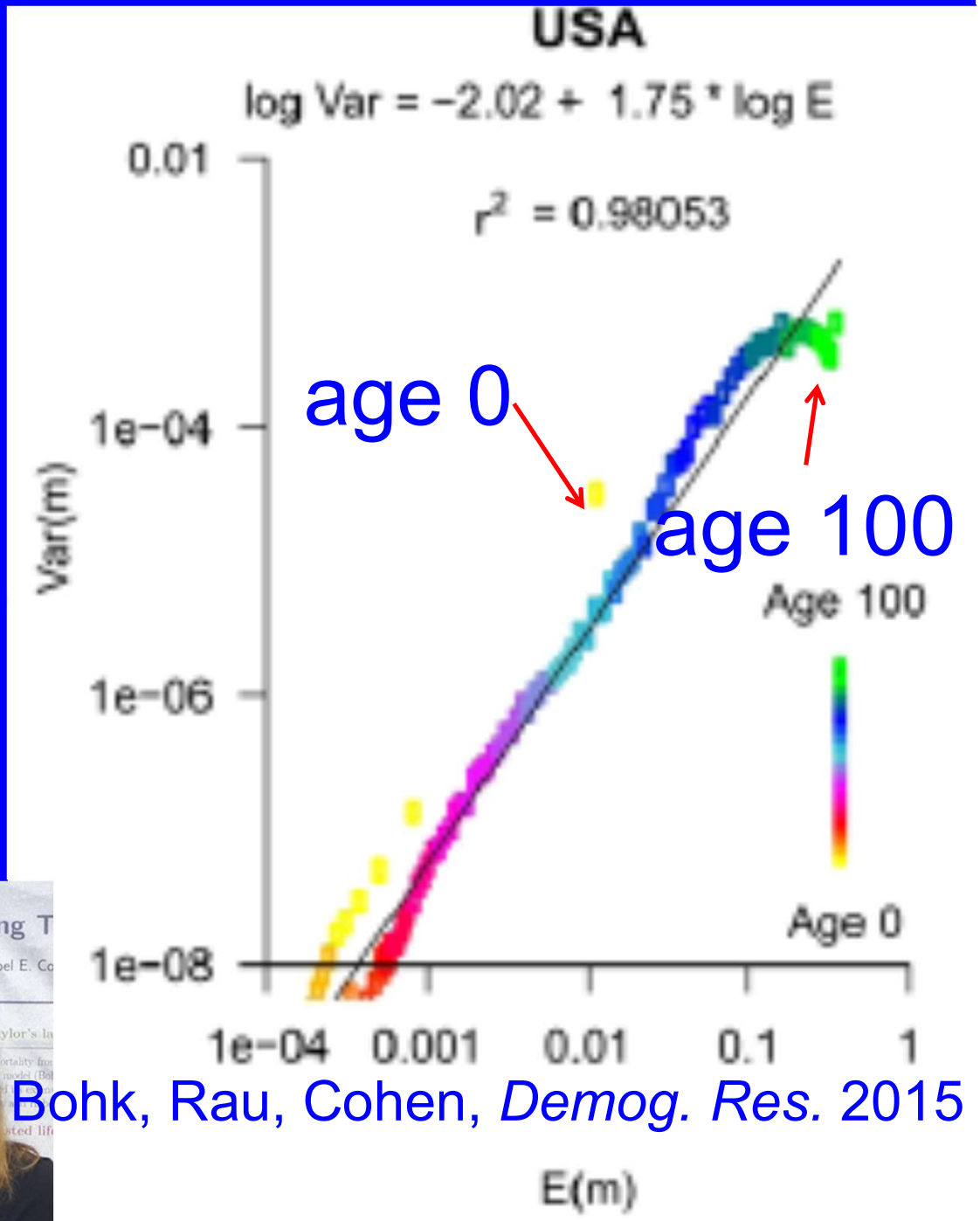
Demographic Research 2015, 2018

Age-specific death rates, USA

Human Mortality Database

Year $t\downarrow$ Age $x\rightarrow$	0	1	x	100
1960			…	
1961			…	
t			$m_{t,x} = \frac{d_{t,x}}{L_{t,x}}$	
2008			…	
2009			…	
Temporal mean			Mean(m_x)	
Temporal variance			Var(m_x)	

Female death rates for each year of age, mean & variance over 1960-2009, USA



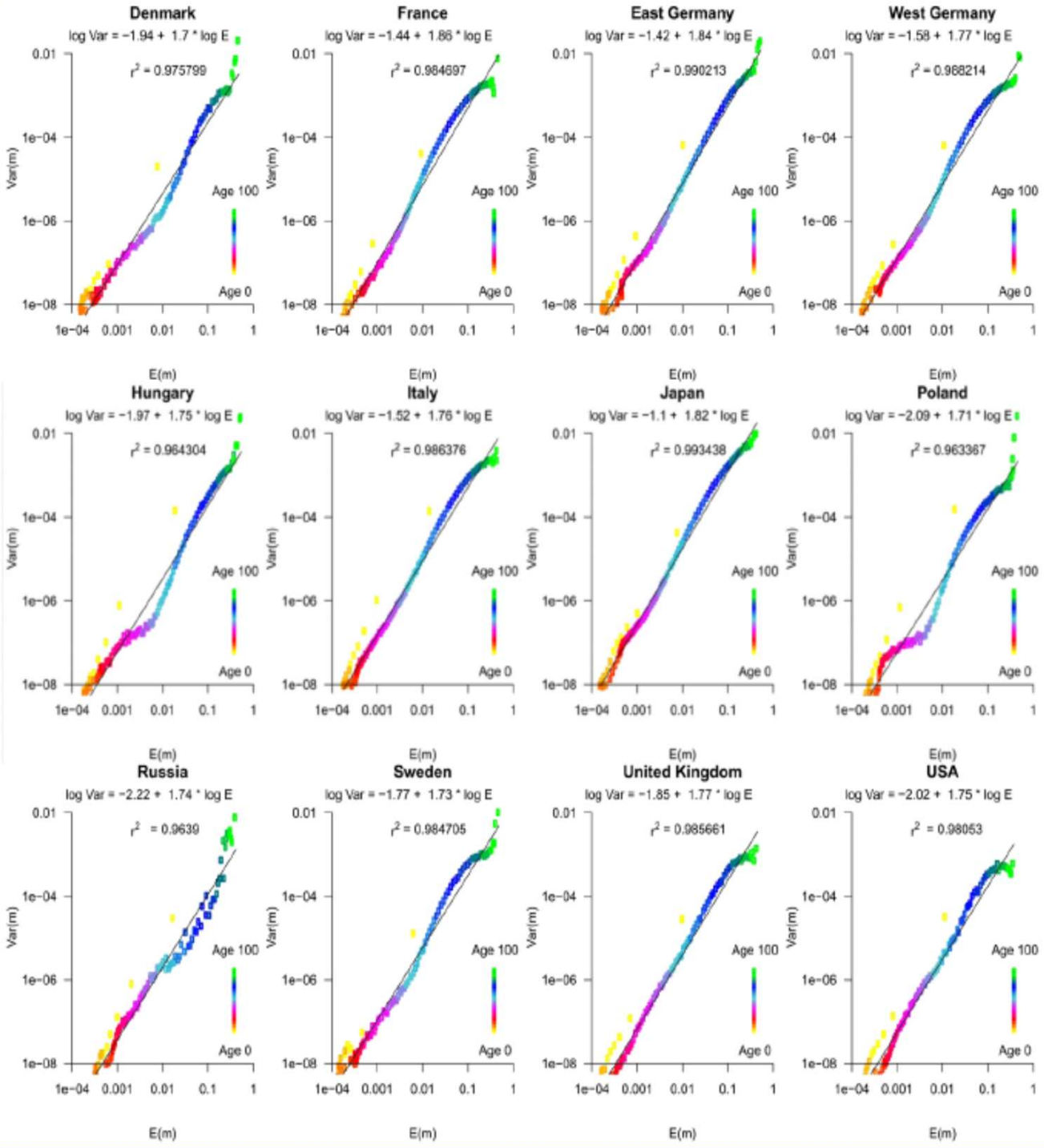
Bohk, Rau, Cohen, *Demog. Res.* 2015

Human ♀ mortality: temporal TL

At each age 0-100 years, mean, var of age-specific ♀ death rate over 1960-2009

Cohen, Bohk-Ewald, Rau, *Demog. Res.* 2018

2020-09-10

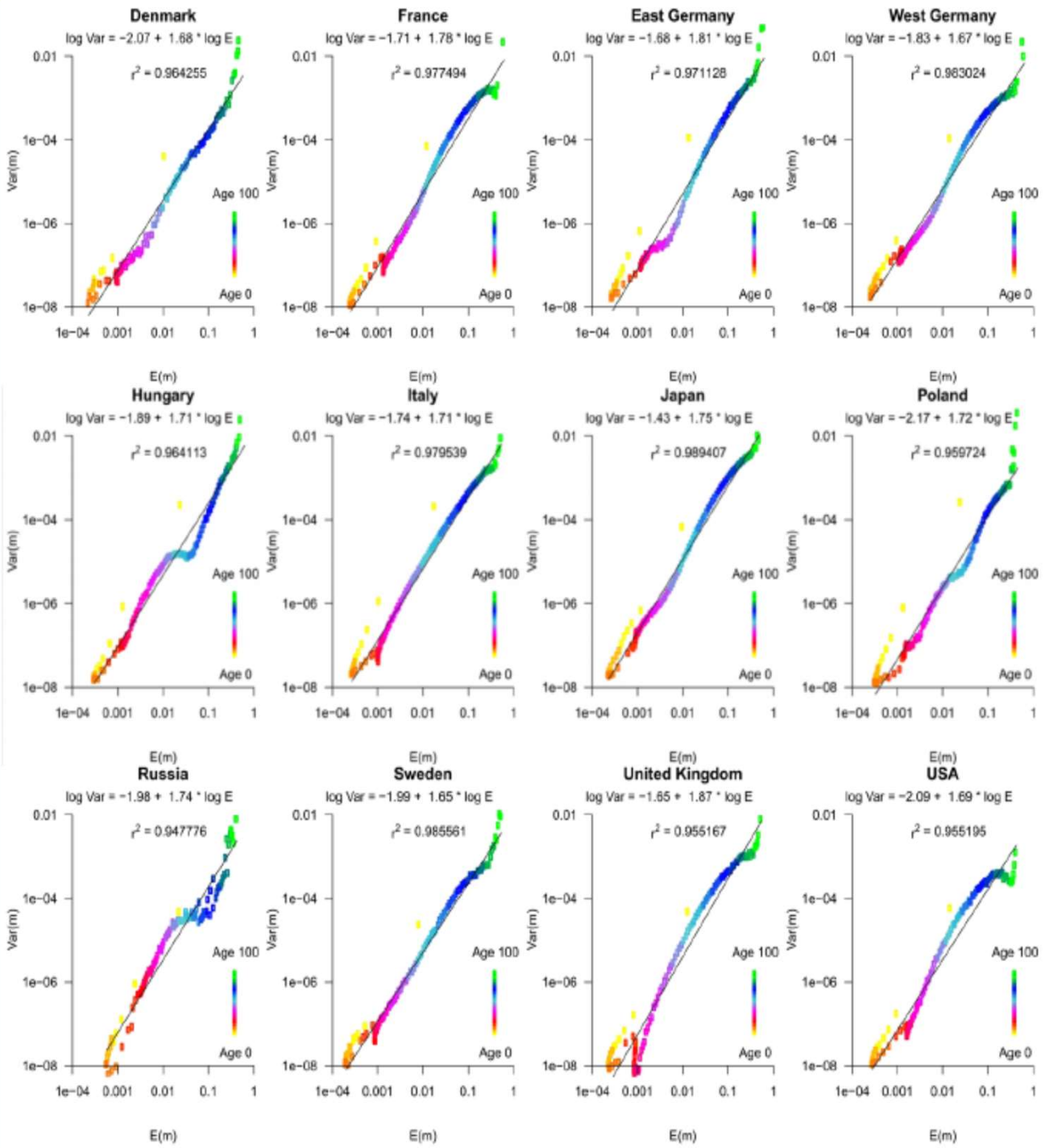


Human ♂ mortality: temporal TL

At each age 0-100 years, mean, var of age-specific ♂ death rate over 1960-2009

Cohen, Bohk-Ewald, Rau, *Demog. Res.* 2018

2020-09-10

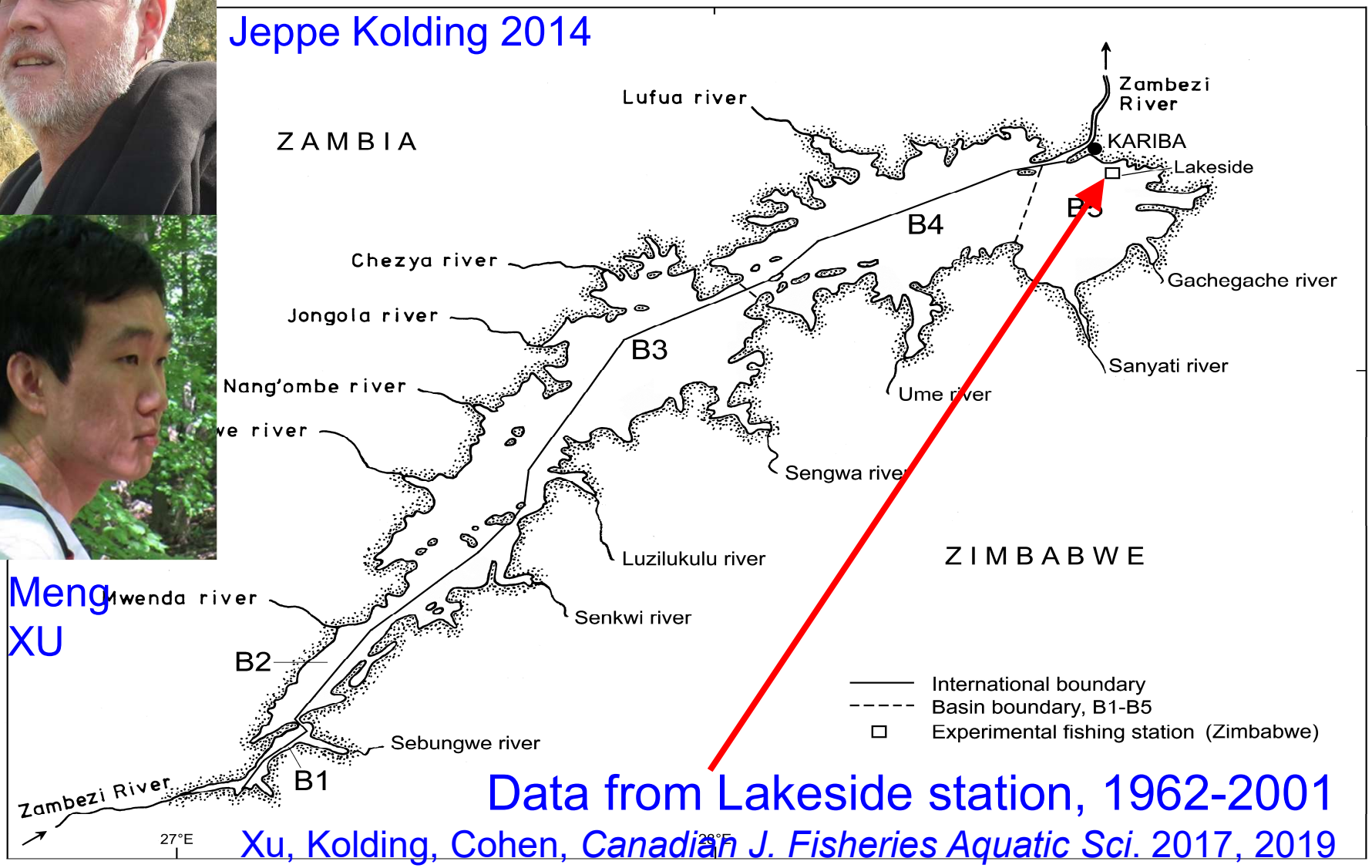


Fisheries

Lake Kariba, world's largest human-made reservoir

Meng XU

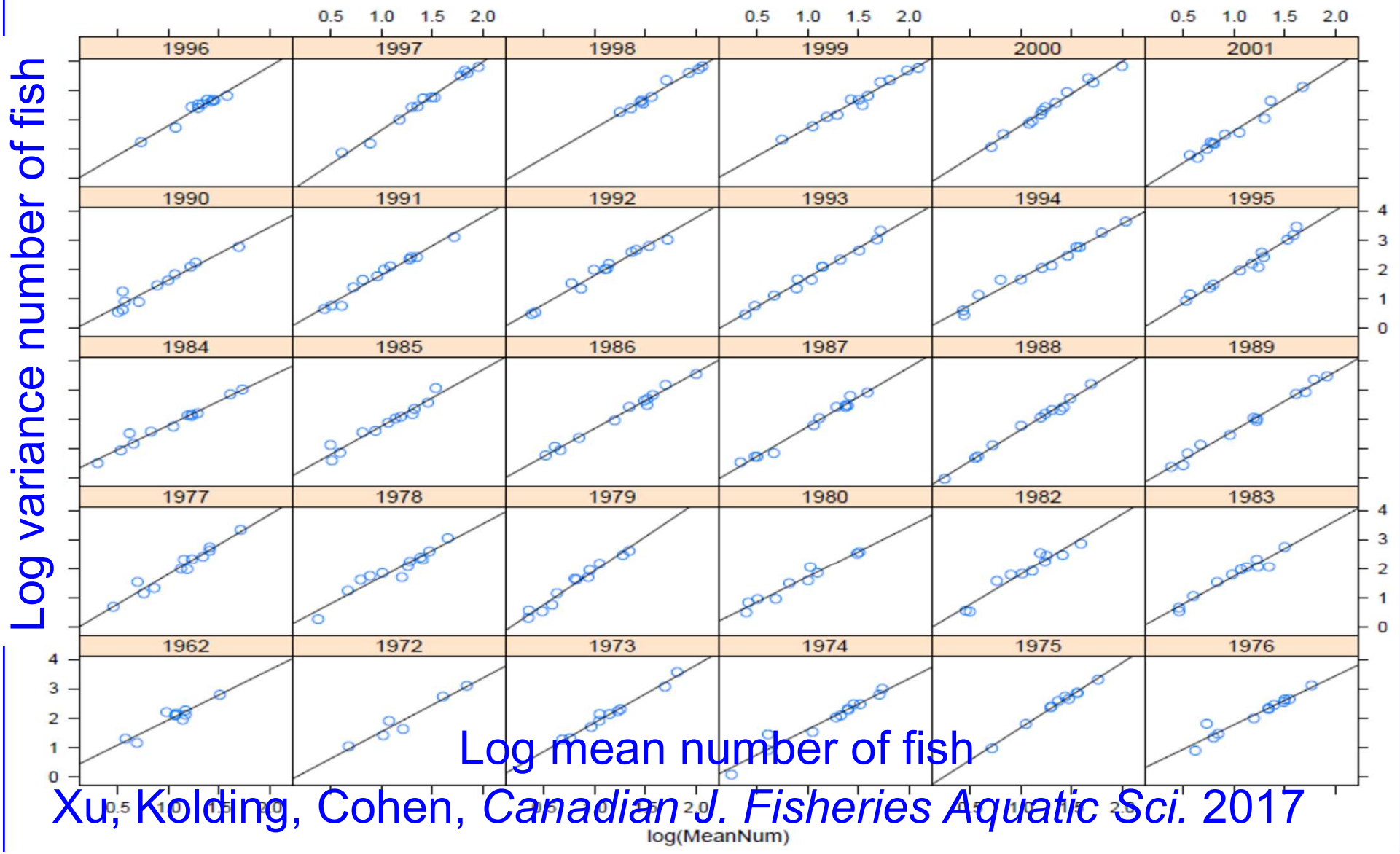
Jeppe Kolding 2014



Data from Lakeside station, 1962-2001

Xu, Kolding, Cohen, *Canadian J. Fisheries Aquatic Sci.* 2017, 2019

Multi-mesh fish sampling, Lake Kariba, obeys temporal Taylor's law. One dot per mesh size



Xu, Kolding, Cohen, *Canadian J. Fisheries Aquatic Sci.* 2017

Using TL to design sampling with fixed precision

Suppose we collect multiple samples daily on days $t = 1, 2, ...$

We want to stop sampling when the mean population density can be estimated with a fixed "precision" defined as

$$C = \frac{standard\ deviation\ of\ mean\ density}{mean\ density}.$$

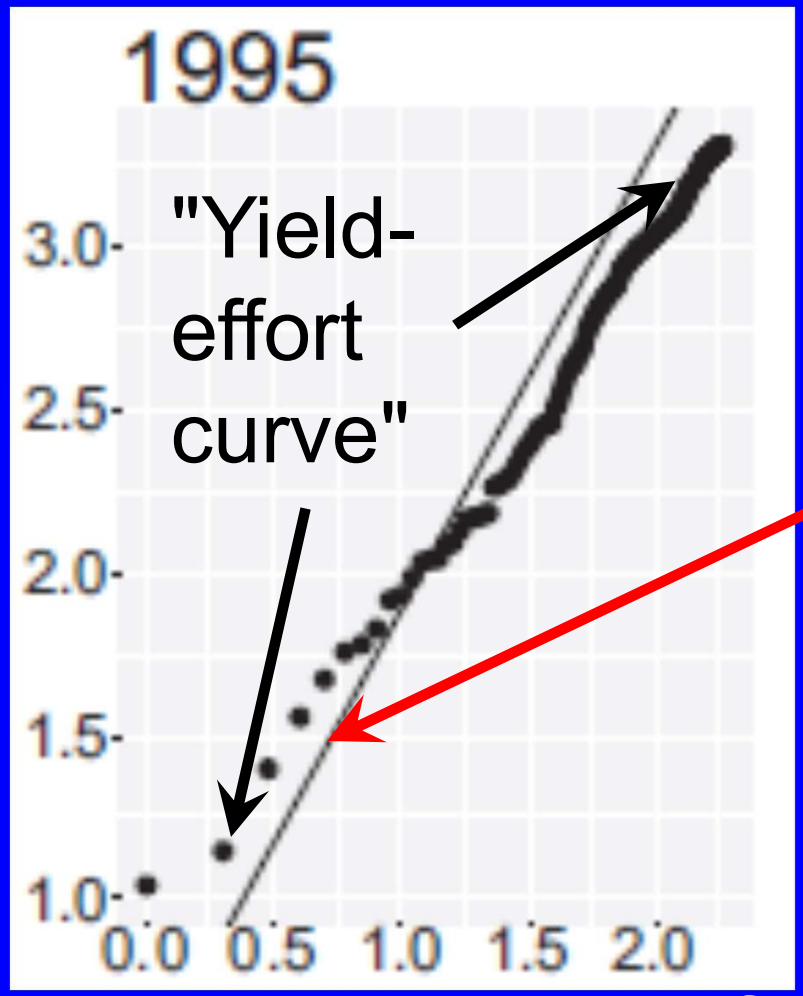
$C = 0.1$ when the standard deviation of mean density (SEM) is 10% of mean density.

TL $var = a \times mean^b$ specifies a stopping line.

Data structure: multiple samples, each with multiple observations

Sample day →	t=1	t=2	t=3
Number of individuals in samples on day t	x_{11}	x_{12}	x_{13}
	x_{21}	x_{22}	x_{23}
	x_{31}	x_{32}	x_{33}
Daily abundance	a_1	a_2	a_3
Cumulative abundance	$A_1 = a_1$	$A_2 = A_1 + a_2$	$A_3 = A_2 + a_3$
Daily no. samples	n_1	n_2	n_3

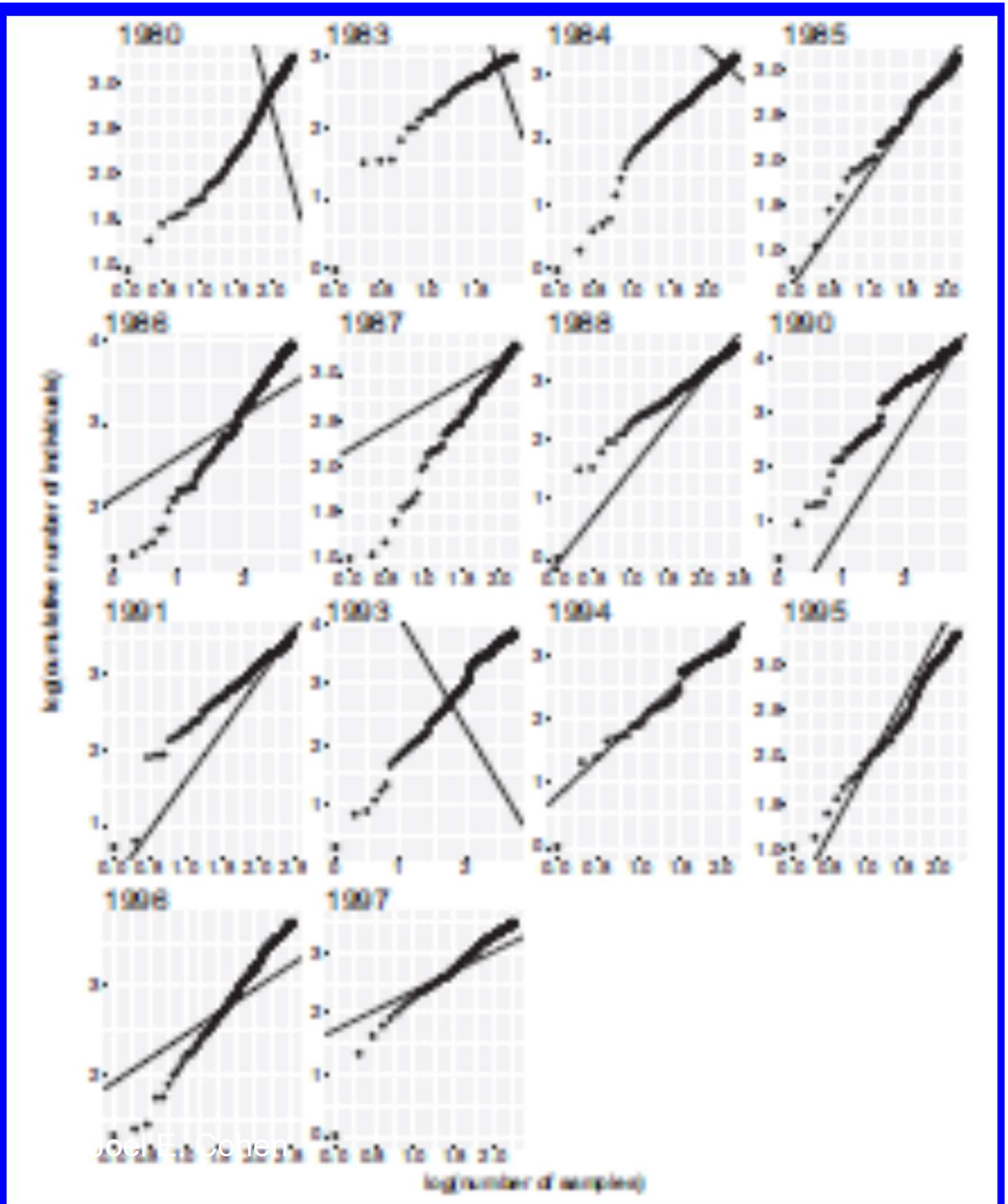
Example: Lake Kariba fisheries, fished area Xu, Kolding, Cohen 2019, CJFAS



Stopping lines from TL were updated each day. Shown is first stopping line with $C = 0.1$ that intersected yield-effort curve.

Example: Lake Kariba fisheries, fished area

Xu, Kolding, Cohen 2019, Canadian J. Fisheries & Aquatic Sciences



2020-09-10

This method could save work.

In Lake Kariba data, using updated stopping lines of fixed precision 0.1 after each new sample, the updated stopping-line method required 21% to 41% of the number of sampling days & 19% to 40% of number of samples that were planned under systematic sampling, depending on sampling area (fished vs unfished) & abundance measure (number vs weight).

Mean abundance estimates were similar to those from systematic sampling.

Primes

Primes not exceeding $x = 10$ are
$p_1 = 2, p_2 = 3, p_3 = 5, p_4 = 7$.

The number $\pi(x)$ of primes not exceeding x is $\pi(x) = 4$.

The mean of those 4 primes is
$m(x) = 4.25 = (2 + 3 + 5 + 7)/4$.

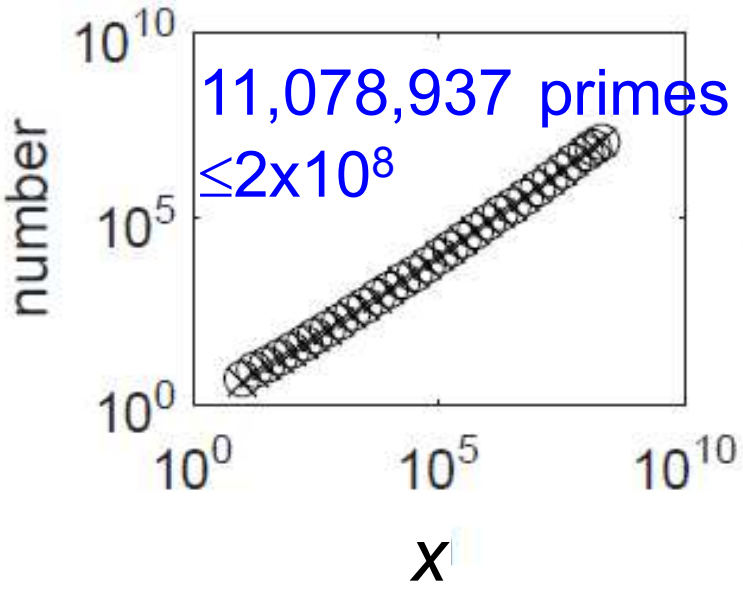
The variance of those 4 primes is $v(x) = 3.6875 = (2^2 + 3^2 + 5^2 + 7^2)/4 - 4.25^2$.

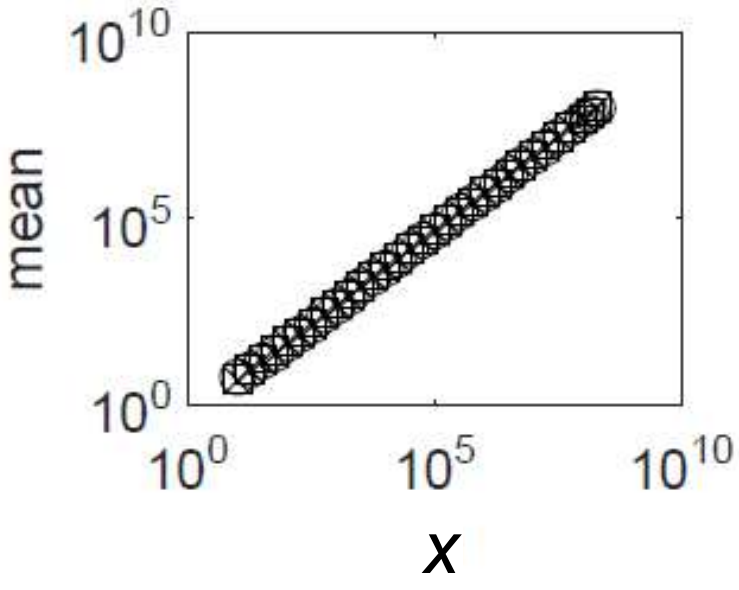
Primes obey $v(x) \sim \frac{1}{3} m(x)^2$ as $x \to \infty$.

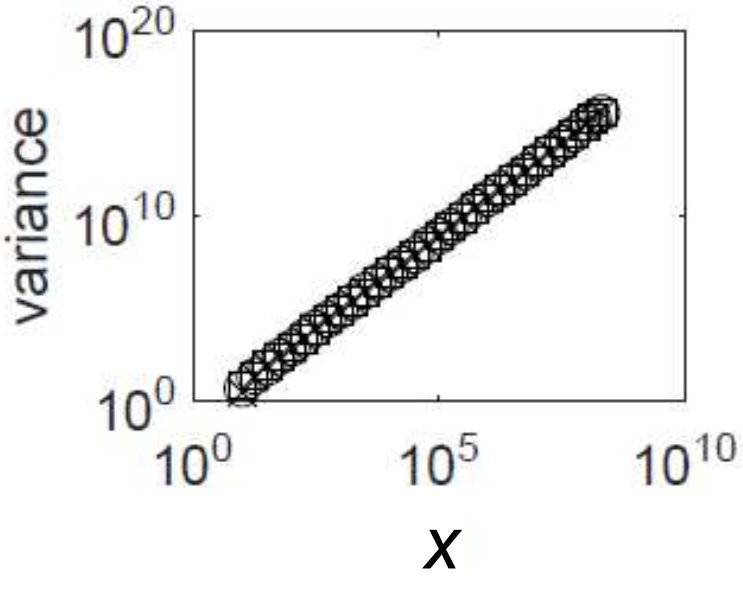
x, exact
enumeration
o, logarithmic
integral
□, asymptotic

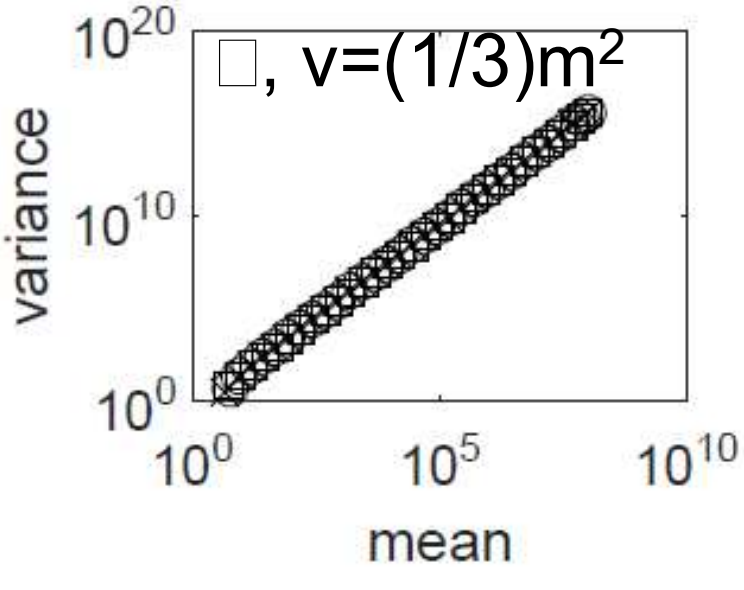
Cohen, *The American Statistician* 2016

2020-09-10









Twin primes

A natural number p is defined to be a twin prime iff p is a prime and $p+2$ is a prime.

3 and 5 are all the twin primes not exceeding 10. 7 is not a twin prime because 7 + 2 = 9 is not a prime.

No one knows whether the number of twin primes is finite or infinite.

Mean twin prime <10 is $m(10) = (3+5)/2 = 4$.

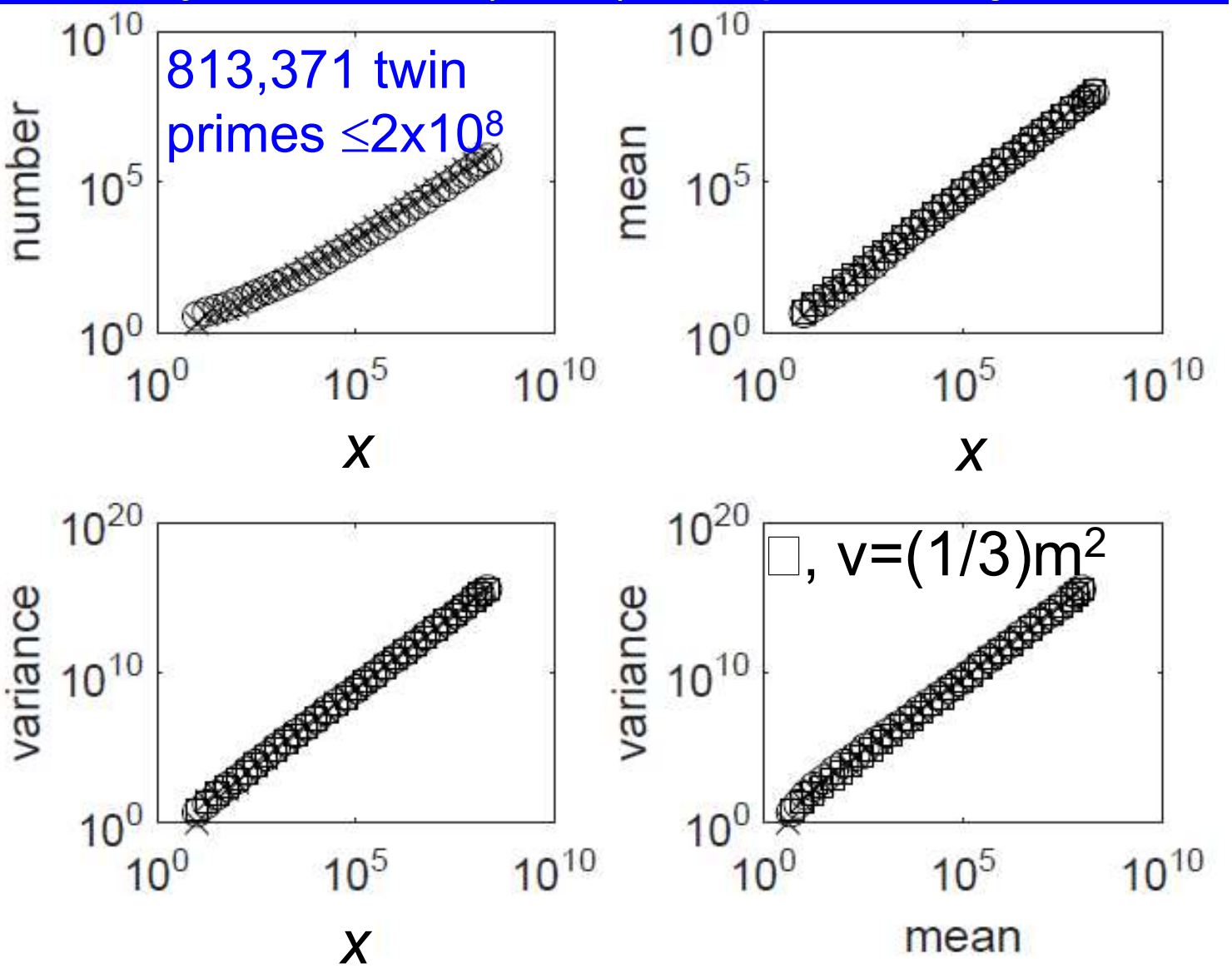
Variance is $v(10)=(1+1)/2=1$.

Twin primes obey $v(x) \sim \frac{1}{3} m(x)^2$ as $x \to \infty$, given Hardy-Littlewood (1923) twin-prime conjecture.

x, exact enumeration o, logarithmic integral □, asymptotic

Cohen, *The American Statistician* 2016

2020-09-10



Prime number theorem

If $f(x)$ and $g(x)$ are real-valued functions of a real x, then $f(x) \sim g(x)$ means that, as $x \to \infty$, $f(x) \to \infty$, $g(x) \to \infty$, $f(x)/g(x) \to 1$.

Hadamard & de la Vallée Poussin, 1896:

If $\pi(x)$ is the number of primes $\leq x$, then

$$\pi(x) \sim \mathrm{li}(x) \equiv \int_2^x \frac{dt}{\log t} \sim \frac{x}{\log x} \text{ as } x \to \infty.$$

Primes are asymptotically uniform on [2, x]. Cohen, *TAS* 2016

Let $X(x)$ be r.v., $X(x) = p \in \mathbb{P}(x)$={primes≤$x$} with probability $1/\pi(x)$, where $\pi(x)=\#\mathbb{P}(x)$.

For $r \in [0,1]$, the cdf of $X(x)$ is defined as $F_{X(x)}(rx) = \Pr\{X(x) \le rx\}$.

Theorem:

For any $r \in [0,1]$, as $x \to \infty$, $F_{X(x)}(rx) \sim r$.

Proof. $F_{X(x)}(rx) = \frac{\#\mathbb{P}(rx)}{\#\mathbb{P}(x)} = \frac{\pi(rx)}{\pi(x)} \sim \frac{(rx)/\log(rx)}{x/\log x}$

$= \frac{r \log x}{\log x + \log r} \to r$ as $x \to \infty$. QED

Hardy-Littlewood Twin Prime Conjecture (HLTPC)

HLTPC: for some constant $C_2 > 0$, for all $x \in [2, \infty)$, $\#\mathbb{P}_2(x) = \#\{3, 5, 11, 17, \ldots \leq x\} = \pi_2(x) \sim 2C_2 \mathrm{li}_2(x) \equiv 2C_2 \int_2^x \frac{dt}{(\log t)^2} \sim 2C_2 \frac{x}{(\log x)^2}$.

$C_2 \approx 0.6601618158\cdots$ Sebah & Gourdon 2002

Theorem Cohen, TAS 2016. If $\pi_2(x) \sim 2C_2 \frac{x}{(\log x)^2}$, then twin primes are asymptotically uniform on [2,x].

Primes satisfy asymptotic TL.

Corollary. Variance $v(x)$ & mean $m(x)$ of the primes $\leq x$ obey TL asymptotically as $x \to \infty$:

$$v(x) \sim (1/3)\left(m(x)\right)^2.$$

If HLTPC holds,

variance & mean of twin primes $\leq x$ obey same TL asymptotically as $x \to \infty$.

Cohen, *TAS*, 2016

Hardy-Littlewood, Bunyakovsky, Schinzel's H, Bateman-Horn (1962) conjectures on admissible prime configurations

(5,7,11), (11,13,17), …, (*p*, *p*+2, *p*+6): *n*=3

Among positive integers ≤*x*, the number of prime-generating integers for *n* polynomials under reasonable conditions is

$$P(x) \sim \frac{C}{D} \int_2^x \frac{dt}{(\log t)^n}.$$

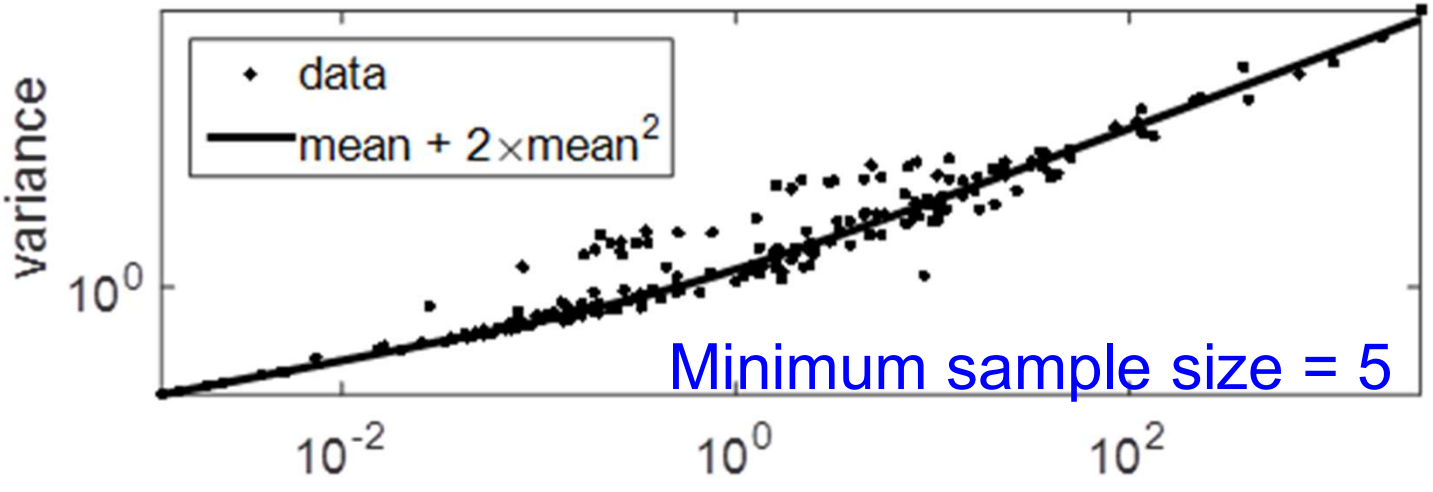
The mean *m*(*x*) & variance *v*(*x*) of these integers obey TL: $v(x) \sim \frac{1}{3} m(x)^2$ as $x \to \infty$.

Taylor's law is not magic.

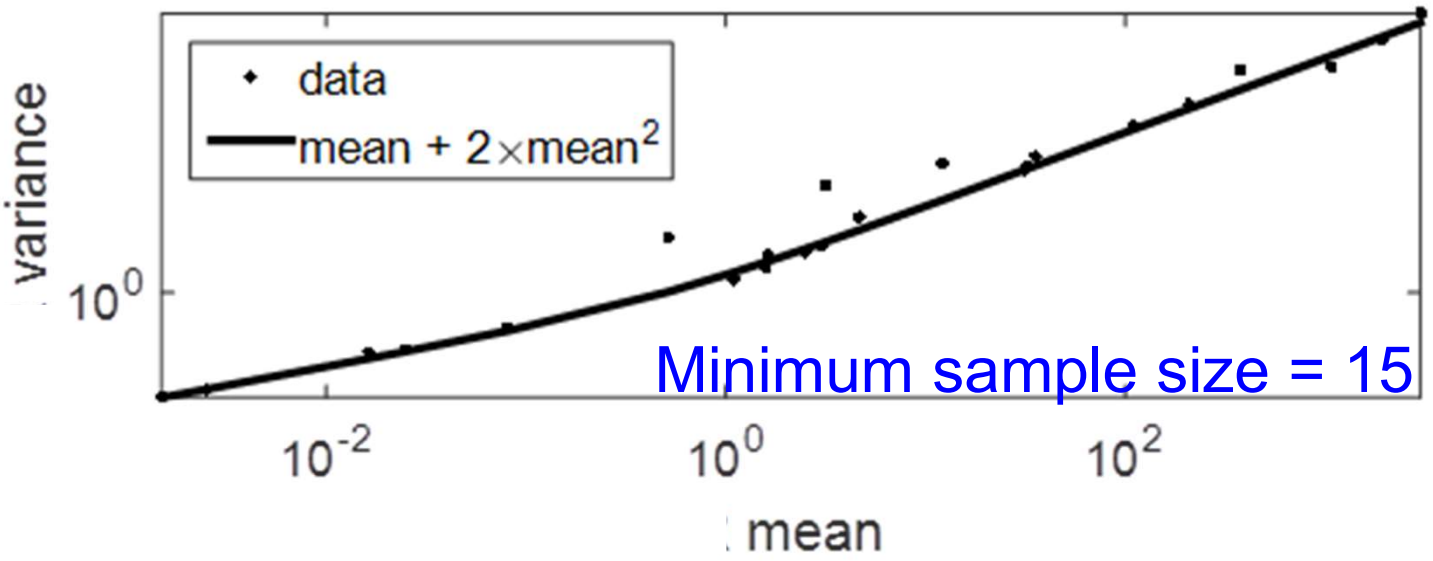
TL is very widely observed, but it is not always the best model.

Parasites/host follow negative binomial: variance=mean+2(mean^2).

Cohen, Lagrue, Poulin, *PNAS* 2016



Robert Poulin ↑



Clément Lagrue ↓

Outline

1. Examples
2. What is Taylor's law (TL)?
3. Recent mathematical results
4. Conclusions

TL data structure: multiple samples, each with multiple observations

Sample number →	*j*=1	*j*=2	*j*=3	*j*=…
Population size or density in units (quadrats, plots, transects, counties, states, years, days)	x_{11}	x_{12}	x_{13}	$x_{\ldots}$
	x_{21}	x_{22}	x_{23}	…
	x_{31}	x_{32}	x_{33}	…
		x_{42}	x_{43}	…
		x_{52}		…
Mean (weighted)	m_1	m_2	m_3	m…
Variance (weighted)	v_1	v_2	v_3	v…

Taylor's law *Nature* 1961

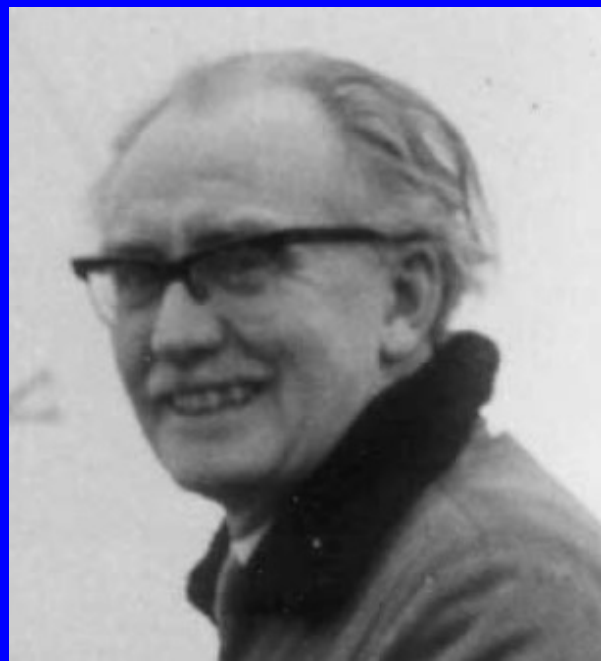
In multiple sets of samples, the variance of population density is proportional to a power of the mean population density.

$$\text{variance} = a(\text{mean})^b, a > 0.$$

$$\log(\text{variance}) = \log(a) + b \cdot \log(\text{mean}).$$

$$\text{variance}/(\text{mean})^b = a, \quad a > 0.$$

Taylor stated no model of error (deviations from exact equality).



Lionel Roy Taylor
(1924–2007)

2020-09-10

TL matters biologically because variability is fundamental.

Fluctuations of forests, fisheries, infectious diseases, disease vectors, tornados, rainfall

Conservation of endangered species

Sampling insect pests of cotton & soybeans, fishery stocks

Evaluation of human population projections

Variability is also fundamental in finance, insurance, computer science, earth science.

Taylor's law(s) mathematically

Given an index set $S \neq \emptyset$ with element(s) s, $\{X(s)\}_{s \in S}$ is a family of nonnegative random variables with finite, positive first 2 moments.

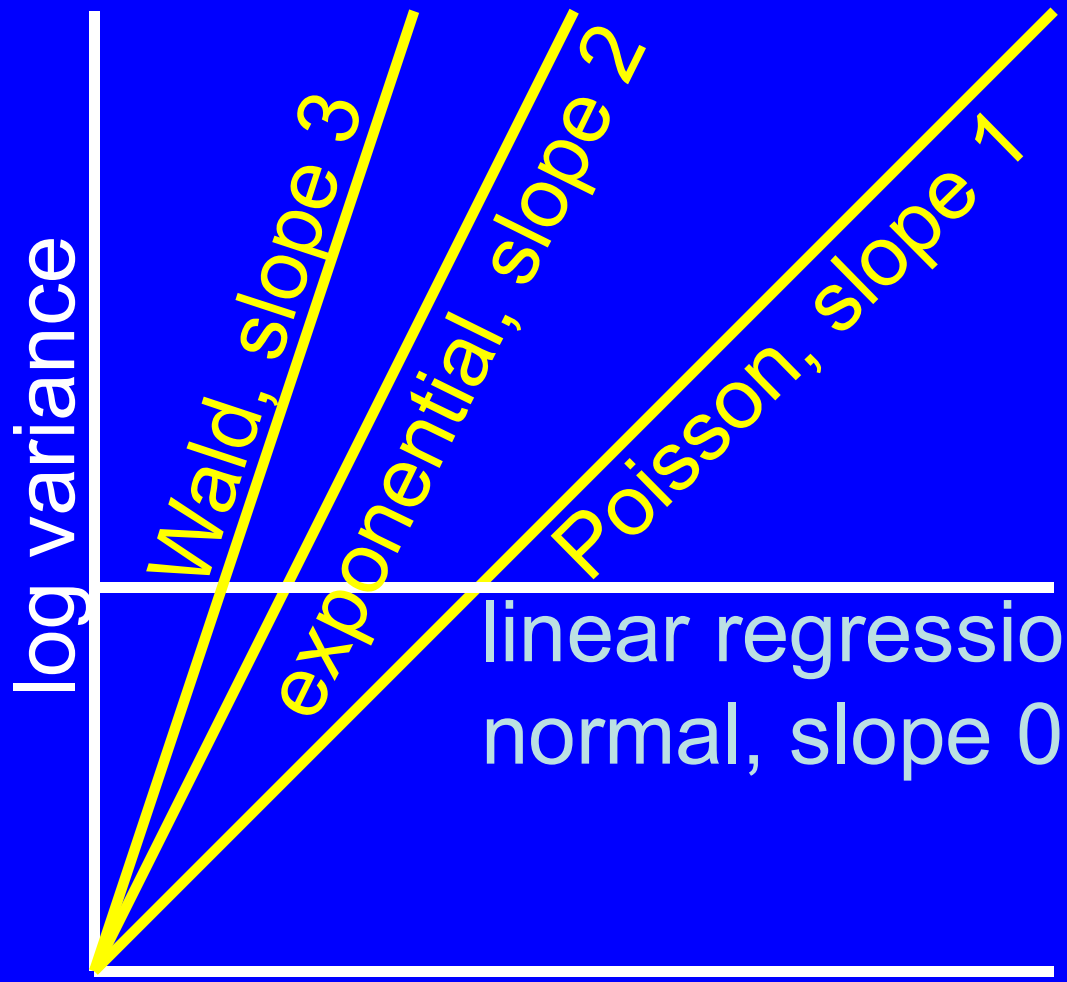
Population TL: ∃ real constants $a > 0, b$, such that $\forall s \in S, \quad Var(X(s)) = a(E(X(s)))^b$.

Sample of size $n > 1$ is a set $\{X_1(s), \ldots, X_n(s)\}$ of n iid copies of $X(s)$, with sample mean $m_n := (X_1 + \cdots + X_n)/n$, sample variance s_n^2.

Sample TL: $\log s_n^2 \approx \log a + b \log m_n$ or

$$\frac{s_n^2}{(m_n)^b} \text{ "is close to" } a > 0.$$

Taylor's law holds exactly for some classical probability distributions.



Other mathematical examples of TL

- Galton-Watson discrete generation branching process
- Continuous-time linear birth & death process
- Stochastic multiplicative process in independent or Markovian environments
- Random sampling from a skewed distribution with finite first 3 moments

And many more!

Outline

1. Examples
2. What is Taylor's law (TL)?
3. Recent mathematical results
4. Conclusions

Recent mathematical results

Weaker assumptions on data or distributions

- No finite moments
- Not identically distributed
- Not independent

Additional measures of variation

- Lower & upper semivariance
- Local lower & local upper semivariance
- Sortino, Sharpe, Farinelli-Tibiletti ratios

Joint work

Mark Brown (Columbia U.), Joel E. Cohen, Victor H. de la Peña (Columbia U.) (2017) Taylor's law, via ratios, for some distributions with infinite mean. *J. Applied Probability* **54**, 657–669.

Brown, Cohen, Chuan-Fa Tang (U. Texas Dallas), Sheung Chi Phillip Yam (Chinese U. Hong Kong), Taylor's law of fluctuation scaling for semivariances and higher moments of heavy-tailed data

Cohen, Richard A. Davis (Columbia U.), Gennady Samorodnitsky (Cornell U.), Heavy-tailed distributions, correlations, kurtosis, and Taylor's law of fluctuation scaling

TL can hold for sample moments even when mean is infinite.

Mark Brown

Victor de la Peña

Stable laws with support $[0,\infty)$

A nonnegative random variable (rv) X is (strictly) stable with tail index α iff $\forall n, X, X_1, \ldots, X_n$ iid, $X_1 + \cdots + X_n =_d n^{1/\alpha} X$.

If X is α-stable with $\alpha \in (0,1)$, then X has infinite mean & all higher moments infinite.

For $c > 0, 0 < \alpha < 1$, define $X =_d F(c, \alpha)$ iff $\Pr\{0 \le X < \infty\} = 1$ & for any $s \ge 0$,

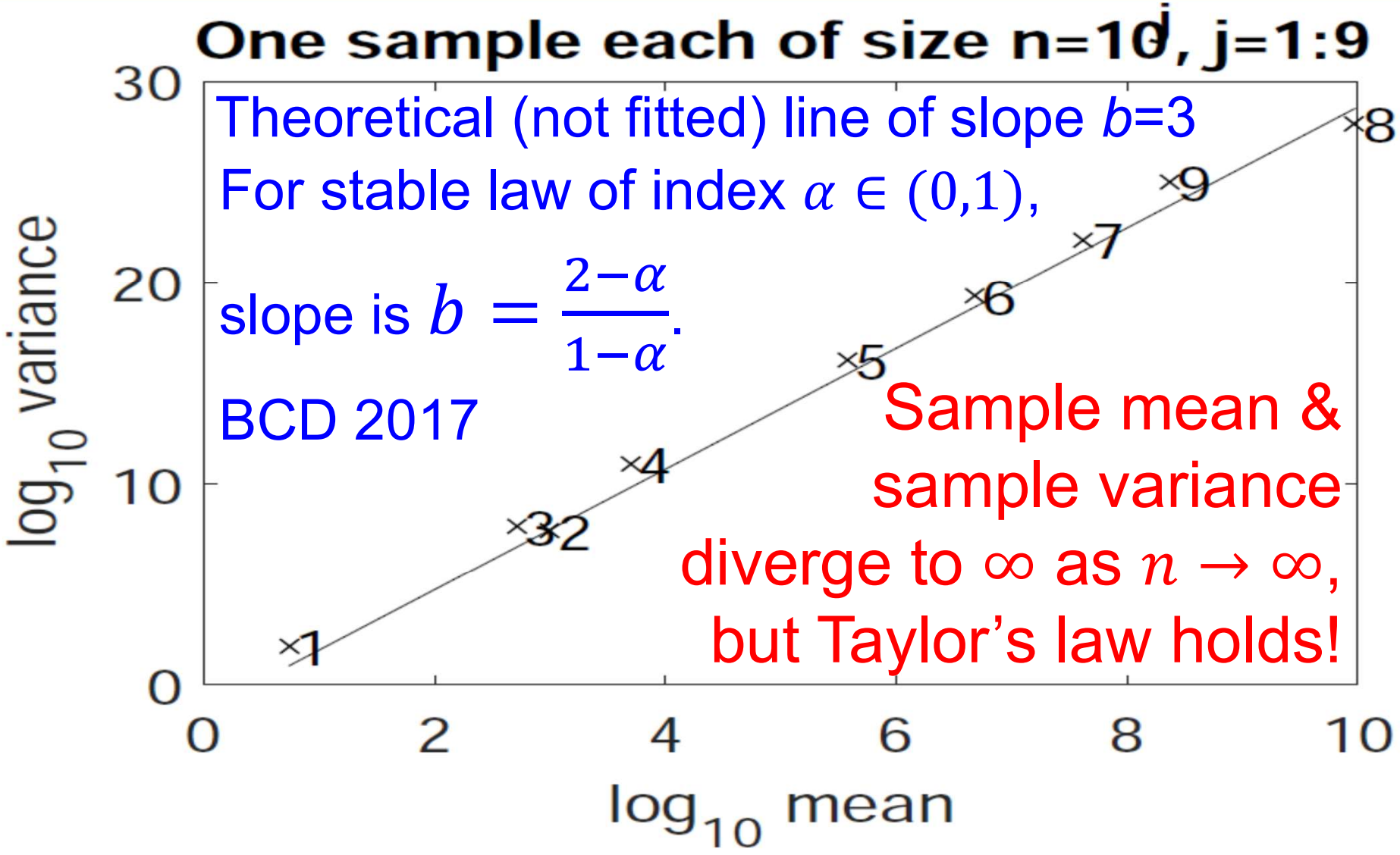
Laplace transform $Ee^{-sX} = e^{-(cs)^\alpha}$.

Every α-stable law X has this form.

Example: stable law with $\alpha = \frac{1}{2}$

If $Z =_d \mathcal{N}(0,1), c > 0$, then $\frac{c}{2Z^2} =_d F(c, \alpha = \frac{1}{2})$ is the distribution of the first passage time to $\sqrt{\frac{c}{2}}$ in standard Brownian motion.

Stable law with $\alpha = 1/2$ obeys TL with increasing sample sizes.



Regularly varying random variables

Y is a regularly varying (RV) rv with index α iff $\forall a > 0, \lim_{x\to\infty} F_Y(ax)/F_Y(x) = a^{-\alpha}$.

$L(x)$ is slowly varying at ∞ iff

$\forall a > 0, \lim_{x\to\infty} L(ax)/L(x) = 1.$

Every $Y \in RV(\alpha)$ obeys $\Pr\{Y > x\} = x^{-\alpha}L(x)$ for some $L(x)$ slowly varying at ∞.

The domain of attraction of α-stable X is exactly every $Y \in RV(\alpha)$.

We write: $Y \in RV(\alpha)$ or $Y \approx_d F(c, \alpha)$.

Example

If $Z =_d \mathcal{N}(0,1), \alpha \in (0,1)$, then

$$X := |Z|^{-1/\alpha} = \left(Z^2\right)^{-1/(2\alpha)} \in RV(\alpha).$$

Taylor's law for iid $RV(\alpha)$ rvs

Brown, Cohen, de la Peña, *J. Appl. Prob.* 2017

For $\alpha \in (0,1)$, let iid $X_1, \ldots, X_n \in RV(\alpha)$, $n > 1$. If

$$b := \frac{2-\alpha}{1-\alpha},$$

then as $n \to \infty$,

$W_n := \frac{s_n^2}{m_n^b} \to_d W(b)$, (non-degenerate limit distribution!)

$$\lim_{n\to\infty} EW_n = 1-\alpha;\ Var(W_n) \to (1-\alpha)^2;$$

$$\frac{\log s_n^2}{\log m_n} \to_p b.$$

We have any more results,
but I have no more time.

Conclusions: many roads lead to TL.

- Many models yield TL exactly or asymptotically.
- Power-law form & parameter values of TL do not determine underlying mechanisms.
- If the underlying mechanism is known, parameter values of TL can discriminate modes of operation of the mechanism.
- Interpreting the parameters of TL in terms of a specific mechanism requires testing the assumptions against detailed data.
- Many mathematical questions remain open.

Selected references, page 1

Cohen, J.E., Rodríguez-Planes, Lucía, Gaspe, María S., Cecere, María C., Cardinal, Marta V., Gürtler, Ricardo E. 2017 Chagas disease vector control and Taylor's law. PLoS Neglected Tropical Diseases 11(11):e0006092, November. https://doi.org/10.1371/journal.pntd.0006092

Cohen, J.E., Poulin, Robert, Lagrue, Clément 2016 Linking parasite populations in hosts to parasite populations in space through Taylor's law and the negative binomial distribution. Proceedings of the National Academy of Sciences, USA 114(1):E47-E56. doi/10.1073/PNAS.1618803114

Xu, Meng, Kolding, Jeppe & Cohen, J.E. 2019 Sequential analysis and design of fixed-precision sampling of Lake Kariba fishes using Taylor's power law. Canadian J. of Fisheries & Aquatic Sciences 76(6):904-917 dx.doi.org/10.1139/cjfas-2018-0091

Selected references, page 2

Cohen, Joel E., Lai, Jiangshan, Coomes, David A. & Allen, Robert B. 2016 Taylor's law and related allometric power laws in New Zealand mountain beech forests: the roles of space, time and environment. Oikos 125(9):1342-1357. doi: 10.1111/oik.02622

Cohen, Joel E., Xu, Meng & Brunborg, Helge 2013 Taylor's law applies to spatial variation in a human population. Genus 69(1):25-60. DOI 10.4402/genus-491.

Xu, Meng, Brunborg, Helge & Cohen, Joel E. 2017 Evaluating multi-regional population projections with Taylor's law of mean-variance scaling and its generalization. Journal of Population Research 34(1):79-99. DOI 10.1007/s12546-016-9181-0

Selected references, page 3

Bohk, Christina, Rau, Roland & Cohen, J. E. 2016 Taylor's power law in human mortality. Demographic Research 33(21):589–610, September 17. DOI: 10.4054/DemRes.2015.33.21.

Cohen, Joel E., Bohk-Ewald, Christina, Rau, Roland 2018 Gompertz, Makeham, and Siler models explain Taylor's law in human mortality data. Demographic Research 38(29):773-842. DOI: 10.4054/DemRes.2018.38.29

Cohen, Joel E., Bohk-Ewald, Christina, Rau, Roland 2018 Why does Taylor's law in human mortality data have slope less than 2, contrary to the Gompertz model? Demographic Research August. https://www.demographic-research.org/volumes/vol38/29/letter.htm, https://www.demographic-research.org/volumes/vol38/29/letter_67.pdf

Selected references, page 4

Tippett, Michael K. & Cohen, Joel E. 2016 Tornado outbreak variability follows Taylor's power law of fluctuation scaling and increases dramatically with severity. Nature Communications 7:10668. DOI: 10.1038/ncomms10668

Tippett, Michael K., Lepore, Chiara & Cohen, Joel E. 2016 More tornadoes in the most extreme U.S. tornado outbreaks. Science DOI: 10.1126/science.aah7393

Brown, Mark, Cohen, Joel E. & de la Peña, Victor 2017 Taylor's law, via ratios, for some distributions with infinite mean. J. of Applied Probability 54(3):1-13. doi:10.1017/jpr.2017.25

Reuman, Daniel C., Zhao, Lei, Sheppard, Lawrence W., Reid, Philip C. & Cohen, Joel E. 2017 Synchrony affects Taylor's law in theory and data. Proc. National Academy of Sciences, USA 114(26):6788–6793. www.PNAS.org/cgi/doi/10.1073/PNAS.1703593114

Selected references, page 5

Cohen, Joel E. 2013 Taylor's power law of fluctuation scaling and the growth-rate theorem. Theoretical Population Biology 88:94–100.

Cohen, Joel E. 2014 Taylor's law and abrupt biotic change in a smoothly changing environment. Theoretical Ecology 7(1):77-86. DOI 10.1007/s12080-013-0199-z.

Cohen, Joel E. 2014 Stochastic population dynamics in a Markovian environment implies Taylor's power law of fluctuation scaling. Theoretical Population Biology 93:30-37. DOI:10.1016/j.tpb.2014.01.001

Cohen, Joel E. 2016 Statistics of primes (and probably twin primes) satisfy Taylor's law from ecology. American Statistician 70(4):399-404. DOI: 10.1080/00031305.2016.1173591

Selected references, page 6

Cohen, Joel E. 2019 Every variance function, including Taylor's power law of fluctuation scaling, can be produced by any location-scale family of distributions with positive mean and variance. Theoretical Ecology doi: 10.1007/s12080-019-00445-7

Cohen, Joel E. 2020 Species-abundance distributions and Taylor's power lawof fluctuation scaling. Theoretical Ecology. https://doi.org/10.1007/s12080-020-00470-x

Cohen, Joel E. & Saitoh, Takashi 2016 Population dynamics, synchrony, and environmental quality of Hokkaido voles lead to temporal and spatial Taylor's laws. Ecology 97(12):3402-3413. DOI: 10.1002/ecy.1575

Selected references, page 7

Saitoh, Takashi & Cohen, Joel E. 2018 Environmental variability and density dependence in the temporal Taylor's law. Ecological Modeling 387:134-143. https://doi.org/10.1016/j.ecolmodel.2018.07.017

Xu, Meng, & Cohen, Joel E. 2019 Analyzing and interpreting spatial and temporal variability of US county population distributions using Taylor's law. PLoS ONE 14(12):e0226096. https://doi.org/10.1371/journal.pone.0226096.

Tippett, Michael K. & Cohen, Joel E. 2020 Seasonality of Taylor's law of fluctuation scaling in all-India daily rainfall. npj Climate and Atmospheric Science 3:3. https://doi.org/10.1038/s41612-019-0104-6

Thank you!
Questions?

cohen@rockefeller.edu

20190906 La Fage To Florac Cévennes "Cham des Bondons Chabusse"

2020-09-10