Modelling Gas Escapes in Milano

(based on 2 B.Sc. dissertations at Politecnico di Milano)

Fabrizio Ruggeri

Istituto di Matematica Applicata e Tecnologie Informatiche
Consiglio Nazionale delle Ricerche
Via Bassini 15, I-20133, Milano, Italy, European Union
fabrizio@mi.imati.cnr.it
www.mi.imati.cnr.it/fabrizio/

WHAT IS RELIABILITY?

- Probability of system operating correctly, under specified conditions, for a given time
- Quality over time
- $P(T \geq t)$ (reliability function), with T failure time, nonnegative r.v.

RELIABILITY IN OUR LIFE

- Alarm clock
- Coffee machine
- Car/bus/train
- PC/machinery/phone
- Car/bus/train
- Oven
- Shower
- TV set
- Bed

RELIABILITY IN OUR SOCIETY

- Transportation (cars, airplanes, ships)
- Bridges and roads
- Buildings
- Dams
- Health devices (cardiac valves)
- Nuclear plants
- Chemical plants
- Missiles
- Appliances

SOME ISSUES IN RELIABILITY

- *System* performance
- Monetary costs
- Social costs
- Warranty (length, cost, forecast)
- Inventory of spare parts
- Maintenance and replacement policy
- Product testing
- Degradation up to failure
- Safety and security

POISSON PROCESS

- One of the simplest and most applied stochastic processes

- Used to model occurrences (and counts) of rare events in time and/or space, when they are not affected by past history

- Applied to describe and forecast incoming telephone calls at a switchboard, arrival of customers for service at a counter, occurrence of accidents at a given place, visits to a website, earthquake occurrences and machine failures, to name but a few applications

- Simple mathematical formulation and relatively straightforward statistical analysis $\Rightarrow$ very practical model for describing and forecasting many random events, although sometimes approximate

POISSON PROCESS

- $N(t)$, $t \geq 0$: stochastic process counting number of events occurred up to time t
- $N(s,t]$, $s < t$: number of events occurred in time interval $(s,t]$
- Poisson process with intensity function $\lambda(t)$: counting process $N(t), t \geq 0$, s.t.
  1. $N(0) = 0$
  2. Independent number of events in non-overlapping intervals
  3. $P\left(N(t, t+\Delta t] = 1\right) = \lambda(t)\Delta t + o(\Delta t)$, as $\Delta t \to 0$
  4. $P\left(N(t, t+\Delta t] \geq 2\right) = o(\Delta t)$, as $\Delta t \to 0$
- Definition $\Rightarrow P(N(s,t] = n) = \dfrac{(\int_s^t \lambda(x)dx)^n}{n!} e^{-\int_s^t \lambda(x)dx}$, for $n \in \mathbb{Z}^+$

  $\Rightarrow N(s,t] \sim \mathcal{P}(\int_s^t \lambda(x)dx)$

  $\Rightarrow N(t) \sim \mathcal{P}(\int_0^t \lambda(x)dx)$

POISSON PROCESS

- Intensity function: $\lambda(t) = \lim_{\Delta t \to 0} \dfrac{P(N(t, t+\Delta t] \geq 1)}{\Delta t}$
  - HPP (homogeneous Poisson process): constant $\lambda(t) = \lambda$, $\forall t$
  - NHPP (nonhomogeneous Poisson process): o.w.
- Mean value function
  - $m(t) = E[N(t)] = \int_0^t \lambda(x)dx, t \geq 0$
  - $m(s,t] = E[N(s,t] = \int_s^t \lambda(x)dx$
- HPP with rate λ
  - $N(s,t] \sim \mathcal{P}(\lambda(t-s))$ and $N(t) \sim \mathcal{P}(\lambda t)$
  - $m(t) = \lambda t$ and $m(s,t] = \lambda(t-s)$
  - Stationary increments (distribution dependent only on interval length)
  - Interfailure times $X_i \sim \mathcal{E}(\lambda)$: $P(X_1 > t) = P(N(t) = 0) = e^{-\lambda t}$
    (same for $X_i, i > 1$)

CASE STUDY: GAS ESCAPES

- Company responsible for a large metropolitan gas distribution network developed in the last century

- Distribution network characterized by very non-homogeneous technical and environmental features (material, diameter of pipes, laying location, etc.)

- Distribution network consists of several thousand kilometers of pipelines providing gas at low, medium and high pressure

- Most of the network is at low-pressure (20 mbar over atmospheric pressure) and attention will be concentrated on it

GAS ESCAPES: RISKS AND COSTS

- Possible explosions: casualties and destructions

- Cost of installation of pipelines

- Cost of maintenance and replacement of riskier pipelines

- Labor cost of emergency and inspection squads

GAS ESCAPES: ATTENUATION OF RISKS

- Gas smells by law (introduction of smelling chemicals) to favor gas escape detections
- Pipelines installation and pipes in house according to the law
- Material chosen according to characteristics of installation area (traffic, ground moisture, residential area or not, etc.)
- Efficient calling center to report possible escapes
- Sufficient size and training of emergency and inspection squads
- Identification of risk factors (kind of junction, material, laying conditions, etc.)
  $\Rightarrow$ *Statistics and Decision Analysis*

GAS ESCAPES: REPLACEMENT POLICY

- Setting up an efficient replacement policy in an urban gas distribution network
- i.e. change of a certain kind of pipelines with a safer one
- assessment of the propensity of failure of the different kinds of pipelines
- assessment of the **probability** of failure of the different kinds of pipelines
- change of pipelines with highest propensity/probability of failure

GAS ESCAPES: DEMERIT-POINT-CARDS

- Propensity to failure determined by demerit-point-cards in many companies (see, e.g., technical report by British Gas Corporation)
- Influence of various quantitative and qualitative factors (diameter, laying depth, etc.) is quantified by assigning a score to each of them (e.g. if laying depth is between 0.9 and 1.5 m, the score is 20)
- Positive aspects
  - easy to specify
  - highlighting the critical factors strongly correlated with pipeline failures

GAS ESCAPES: DEMERIT-POINT-CARDS

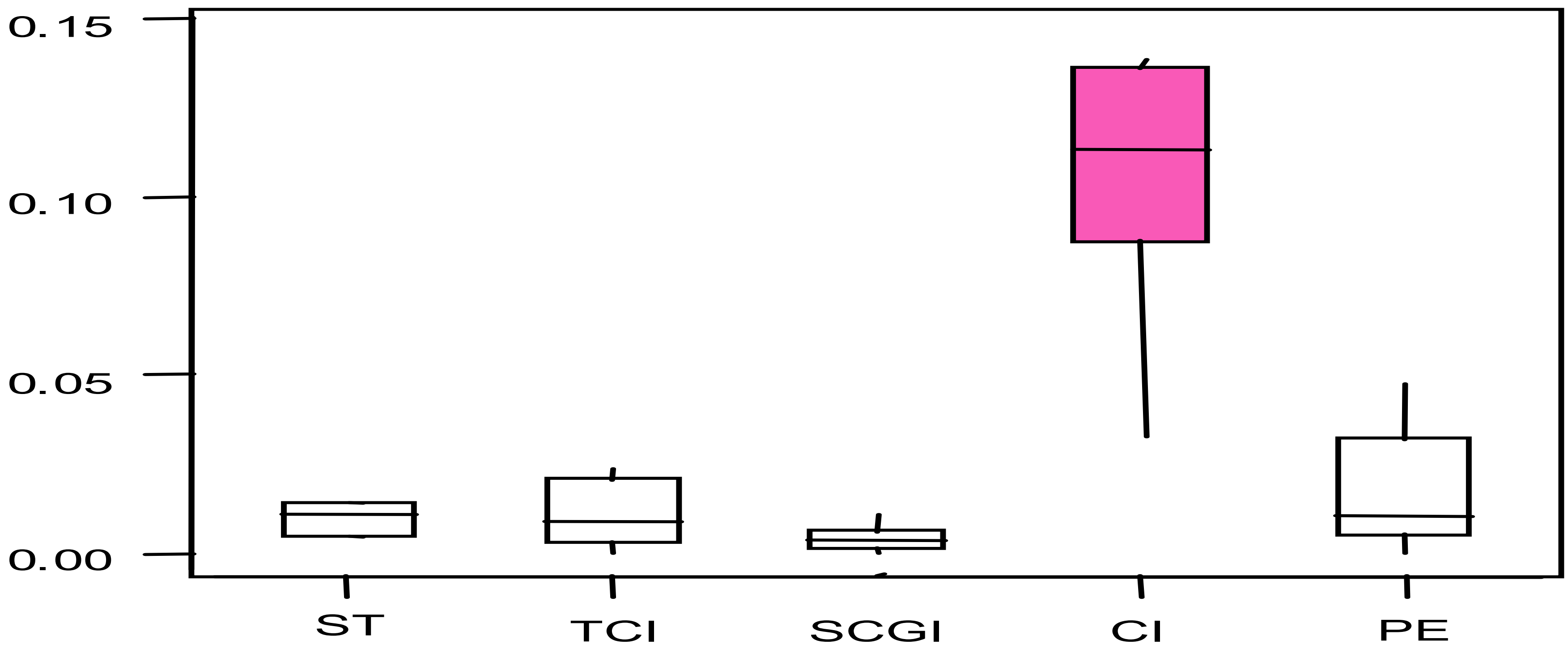
- Negative aspects
  - Setting of the scores, in our context, for the considered factors (and the choice of the factors) would be based only on previous empirical experience in other cities without any adjustment for the current context
  - An aggregate demerit score, obtained by adding the individual scores of the different factors, often hides possible interactions between the considered factors (e.g. diameter and laying depth), so losing other important information and worsening an already critical situation
  - Propensity-to-failure score of a given section of gas pipeline is considered independently of the length of the section and the planning period
  - *Stochastic nature of the phenomenon is ignored*

GAS ESCAPES: MATERIALS

Several materials for low-pressure pipelines

- traditional cast iron (CI)
- treated cast iron (TCI)
- spheroidal graphite cast iron (SGCI)
- steel (ST)
- polyethylene (PE)

GAS ESCAPES: YEARLY FAILURE RATES



GAS ESCAPES: INTEREST ON CI

- higher failure rate than other materials (even by one order of magnitude)

- covers more than a quarter of the whole network

- about 6000 different pipe sections with homogeneous characteristics, ranging in length from 3 to 250 meters for a total of 312 kilometers

GAS ESCAPES: FACTORS

Three groups of factors identified based upon

- studies in other companies
- reports in literature
- discussions with company's experts

GAS ESCAPES: FACTORS

- Intrinsic features of the pipeline-section
  - thickness
  - diameter
  - age
- factors concerning the laying of the pipeline-section
  - depth
  - location
  - ground characteristics
  - type and state of the pavement
  - laying techniques
- environmental parameters of the pipeline-section
  - traffic characteristics
  - intensity of underground services
  - external temperature and moisture

GAS ESCAPES: FACTORS

Reports on those factors rarely available and useful

- companies hardly disclose data on failures/escapes
- companies are in general responsible for a single city network, making very difficult any comparison between different situations and data re-utilisation/sharing
- data registration methods change over the long period of operation of the pipelines
- data are poorly registered or recorded for other purposes which are not sufficiently coherent with the requirements of a correct safety and reliability analysis

GAS ESCAPES: MANOVA ON FACTORS

- Multivariate analysis of variance (MANOVA) to consider mean differences on two or more dependent variables (i.e. failure factors) simultaneously

- CI distribution network (and the corresponding number of failures) over a significantly long period (10 years) was divided into different classes based on the previous factors

- For each factor, two levels were identified, where the notations ”high” and ”low” were qualitative rather than quantitative

GAS ESCAPES: MANOVA ON FACTORS

- The study identified diameter, laying depth and location as the most significant factors

- The other factors given above were not particularly important, as they turned out
  - either to be homogeneous for the analyzed distribution network (e.g. the soil used during installation works always had the same chemical and mechanical characteristics, while external temperature and moisture actually do not have a different effect on different pipeline sections, as the ground has a strong insulating capacity, even with shallow laying)
  - or to be strongly correlated to the above factors (e.g. the thickness of a pipe is fixed for a given diameter).

- Those considerations were fully shared and validated by company experts

GAS ESCAPES: FIRST FINDINGS

Following this preliminary data analysis, the most important conclusions were

- This industrial sector is characterized by a remarkable *shortage of data* because pipeline failures are rare and available information is often inadequate. This scarcity of data, together with some underlying ”noise” stemming from imprecise recording of data, suggested that information obtained from the company archives should be improved with expert judgements using a *Bayesian approach*

GAS ESCAPES: FIRST FINDINGS

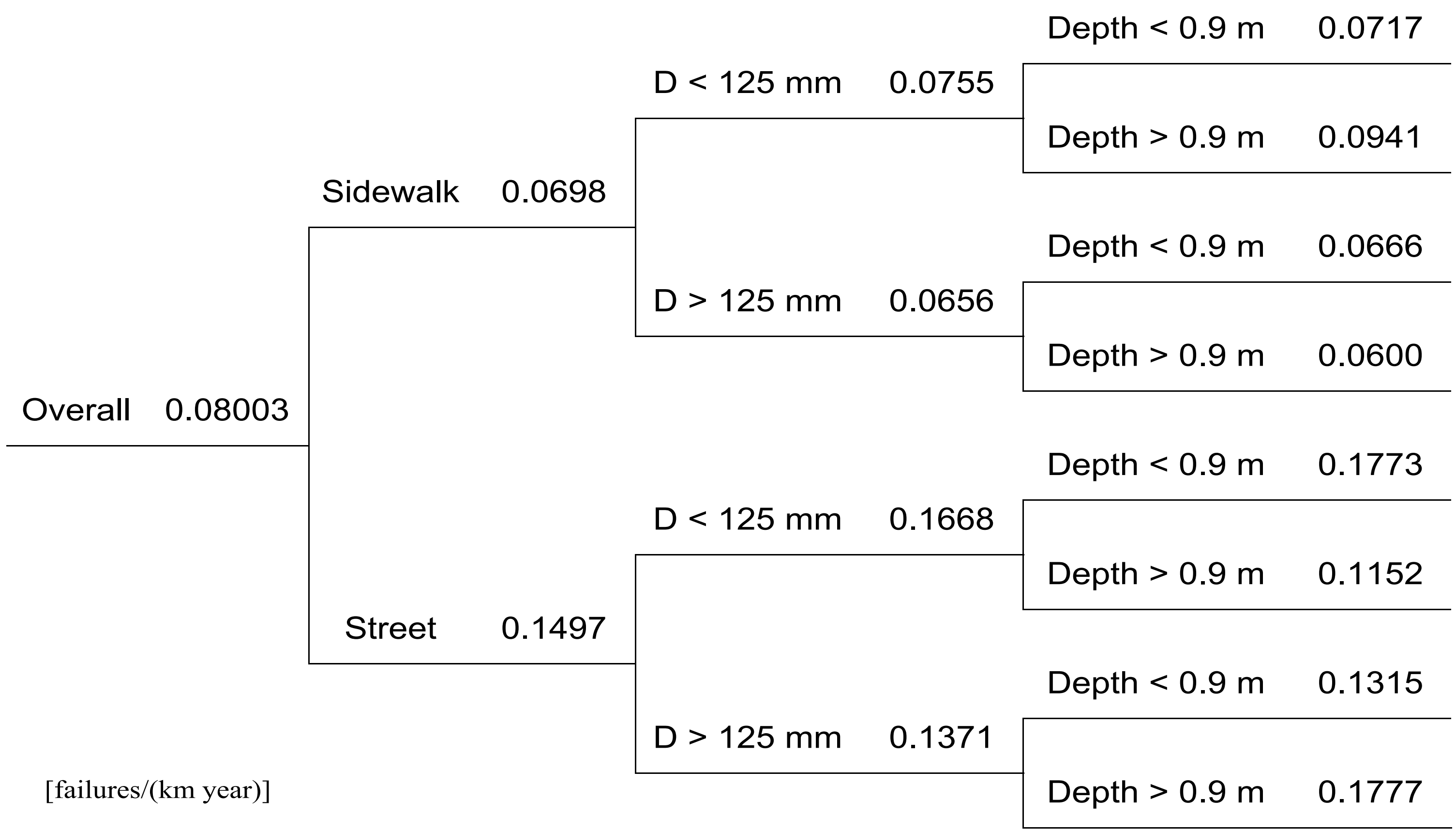
- CI pipeline failure rate seems to be *scarcely sensitive to wear* or proximity to a previous failure (or leak) in the same pipeline section, but it is mostly influenced by accidental stress, even if the useful life phase may be considered longer than 50 years. So propensity-to-failure in a unit time period or unit length does not vary significantly with time and space. Since failures are rare events (an assumption confirmed by available data), it was felt appropriate to model them with an *homogeneous (in time and space) Poisson process*

- The evident importance of interactions between factors (e.g. diameter is significant with shallow but not with deep laying) led to the abandonment of the additive approach, typical of the demerit-point-cards, which sums the effects of the factors, in favor of the determination of pipeline classes derived from the combination of the levels of the most significant factors. This proposed data organization also facilitates the *expression of experts judgements*

FAILURES IN CAST-IRON PIPES

- Some materials (e.g. steel) subject to aging $\Rightarrow$ NHPP
- Cast-iron is not aging $\Rightarrow$ HPP in space and time
- HPP with parameter λ (unit failure rate in time and space)
- n failures in $[0, T] \times \mathcal{S}, \Rightarrow L(\lambda|n, T, \mathcal{S}) = (\lambda s T)^n e^{-\lambda s T}/n!$, with $s = meas(\mathcal{S})$
- Data: $n = 150$ failures in $T = 6$ years on a net $\approx s = 312$ Km long
  $\Rightarrow L(\lambda|n, T, \mathcal{S}) = (1872\lambda)^{150} e^{-1872\lambda}/150!$
- MLE $\widehat{\lambda} = n/(sT) = 150/1872 = 0.080$
- 8 classes determined by two levels of relevant covariates (diameter, location and depth) $\Rightarrow$ 8 HPP's because of
- *Coloring Theorem*: Data from HPP(λ) allocated (independently from the HPP) with probability p_i, $i = 1, m$, to m classes $\Rightarrow$ m independent HPP(λ_i), $\lambda_i = \lambda p_i$

FAILURES IN CAST-IRON PIPE



BAYESIAN APPROACH

- Bayes Theorem: X r.v. with density $f(x|\lambda)$ and prior $\pi(\lambda)$
  - sample $\mathbf{X} = (X_1, \ldots, X_n) \Rightarrow f(\mathbf{X}|\lambda) = L(\lambda|\mathbf{X})$ likelihood
  - posterior $\pi(\lambda|\mathbf{X}) = \dfrac{L(\lambda|\mathbf{X})\pi(\lambda)}{\int L(\omega|\mathbf{X})\pi(\omega)d\omega}$

- HPP with parameter λ and n failures in $[0, T] \times \mathcal{S}$
  - likelihood $L(\lambda|n, T, \mathcal{S}) = (\lambda s T)^n e^{-\lambda s T}/n!$
  - $\lambda \sim \mathcal{G}(\alpha, \beta) \Rightarrow \lambda|n, T, \mathcal{S} \sim \mathcal{G}(\alpha + n, \beta + sT)$ (Conjugate prior)
  - posterior

$$\begin{aligned} \pi(\lambda|n, T, \mathcal{S}) &\propto L(\lambda|n, T, \mathcal{S})\pi(\lambda) \\ &\propto \lambda^n e^{-\lambda s T} \cdot \lambda^{\alpha-1} e^{-\beta\lambda} \end{aligned}$$

  - α and β chosen to match mean α/β and variance α/β^2

ELICITATION OF EXPERTS' OPINIONS

- The expert judgements were collected by an ad hoc questionnaire and integrated with historical data by means of Bayesian inference

- Experts from three areas within the company were selected to be interviewed:
  - pipeline design: responsible for designing the network structure (4 experts)
  - emergency squad: responsible for locating network failures (8 experts)
  - operations: responsible for the repair of broken pipelines (14 experts)

ELICITATION OF EXPERTS' OPINIONS

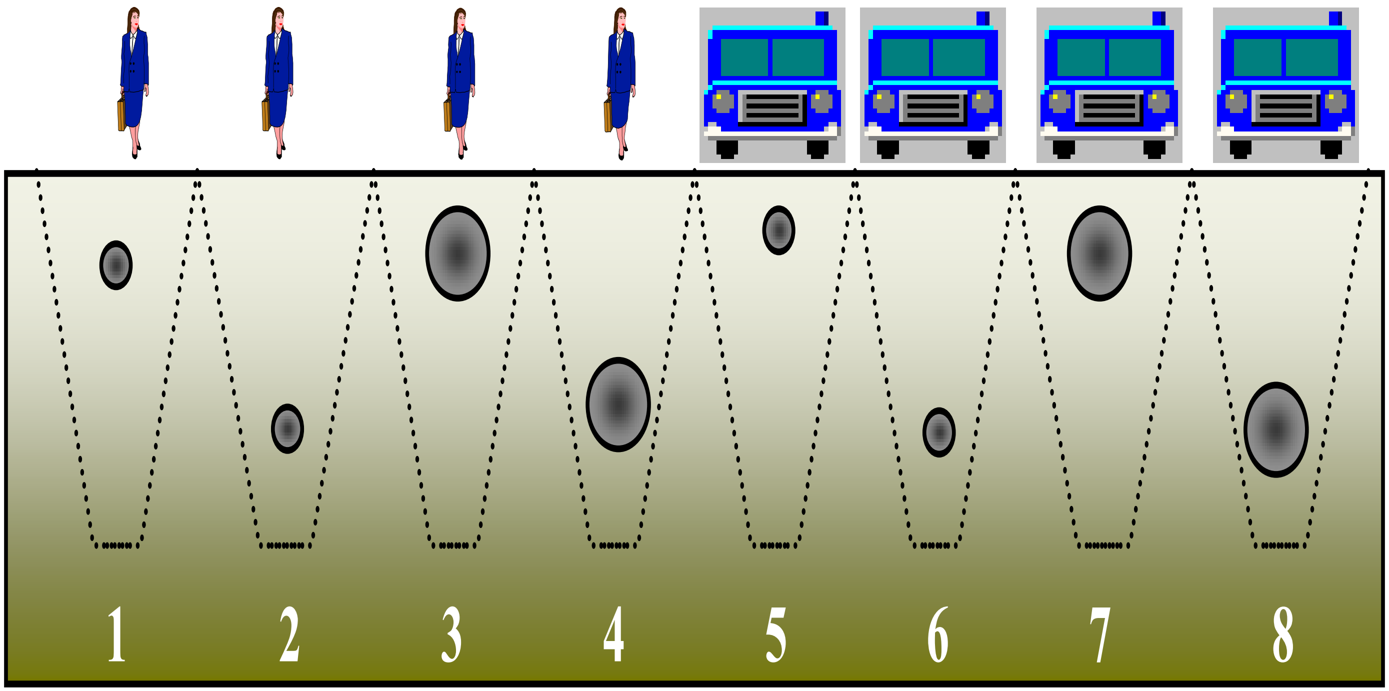
- Interviewees were actually not able to say how many failures they expected to see on a kilometer of a given kind of pipe in a year (the situation became even more untenable when they were asked to express the corresponding standard deviation or upper and lower bounds)

- The experts had great difficulty in saying how and how much a factor influenced the failure and expressing opinions directly on the model parameters while they were able to compare the performance against failure of different pipeline classes

- To obtain such a propensity-to-failure index, each expert was asked to compare the pipeline classes pairwise. In a pairwise comparison the judgement is the expression of the relation between two elements that is given, for greater simplicity, in a linguistic shape

- The linguistic judgement scale is referred to a numerical scale (Saaty's proposal: Analytic Hierarchy Process) and the numerical judgements can be reported in a single matrix of pairwise comparisons

ELICITATION OF EXPERTS' OPINIONS



ANALYTIC HIERARCHY PROCESS

Two alternatives A and B

B	"equally likely as"	A $\rightarrow$ 1
B	"a little more likely than"	A $\rightarrow$ 3
B	"much more likely than"	A $\rightarrow$ 5
B	"clearly more likely than"	A $\rightarrow$ 7
B	"definitely more likely than"	A $\rightarrow$ 9

Pairwise comparison for alternatives $A_1, \ldots, A_n$

$\Rightarrow$ square matrix of size n

$\Rightarrow$ eigenvector associated with the largest eigenvalue

$\Rightarrow (P(A_1), \ldots, P(A_n))$

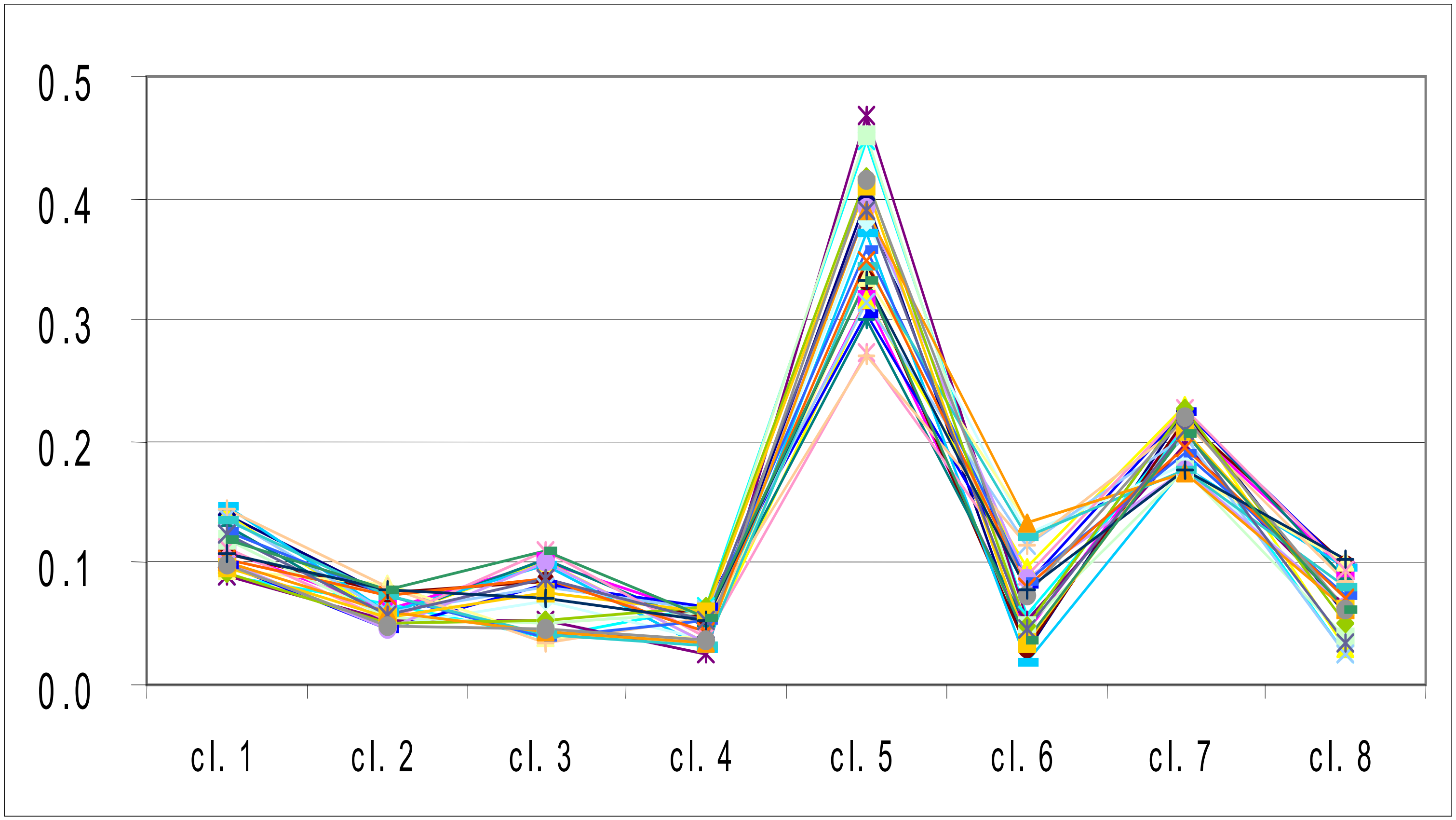
ANALYTIC HIERARCHY PROCESS

An expert's opinion on propensity to failure of cast-iron pipes

Class	1	2	3	4	5	6	7	8
1	1	3	3	3	1/6	1	1/6	3
2	1/3	1	1/4	2	1/6	1/2	1/5	1
3	1/3	4	1	1	1/4	1	1/6	2
4	1/3	1/2	1	1	1/5	1	1/5	1
5	6	6	4	5	1	4	4	5
6	1	2	1	1	1/4	1	1/6	1
7	6	5	6	5	1/4	6	1	4
8	1/3	1	1/2	1	1/5	1	1/4	1

ELICITATION OF EXPERTS' OPINIONS

Probabilities of one gas escape occurring in the eight classes



ESTIMATES' COMPARISON

- Location: **W** (under walkway) or **T** (under traffic)
- Diameter: **S** (small, $<$ 125 mm) or **L** (large, $\geq$ 125 mm)
- Depth: **N** (not deep, $<$ 0.9 m) or **D** (deep , $\geq$ 0.9 m)

Class	MLE	Bayes ($\mathcal{LN}$)	Bayes ($\mathcal{G}$)
TSN	*.177*	**.217**	**.231**
TSD	*.115*	*.102*	*.104*
TLN	*.131*	*.158*	*.143*
TLD	**.178**	*.092*	*.094*
WSN	.072	.074	.075
WSD	.094	.082	.081
WLN	.066	.069	.066
WLD	.060	.049	.051

Highest value; 2^{nd}-4^{th} *values*

- Location is the most relevant covariate
- TLD: 3 failures along 2.8 Km but quite unlikely to fail according to the experts
- $\mathcal{LN}$ and $\mathcal{G}$ $\Rightarrow$ similar answers

GAS ESCAPES: MATERIALS

- traditional cast iron (CI)
  - resistent to corrosion and usage
  - 70-80 years of useful life
  - very fragile with respect to random shocks
- steel (ST)
  - subject to corrosion, despite of cover, and affected by usage
  - 30-40 years of useful life
  - very robust with respect to random shocks
- polyethylene (PE)
  - cheap, currently used in most replacements
  - resistent to corrosion
  - very fragile with respect to ground digging

GAS ESCAPES: MATERIALS

- Evolution over time from CI to ST to PE
- Decision based upon conflicting aspects
  - costs (material, placement, replacement, etc.)
  - reliability (corrosion, frailty, etc.)
  - external conditions (stray currents, traffic, digging, etc.)
- Small example of *Multicriteria Decision Making*

FAILURES IN STEEL PIPES

- Steel pipelines are subject to corrosion, leading to reduction of wall thickness; it can be reduced using
  - bitumen cover
  - cathodic protection (via electric current), working especially when bitumen cover is imperfect

- Most of the low pressure network is without cathodic protection to avoid electrical interference with other metal structures

- Cathodic protected areas can be tested (using electricity) once or twice a year to check for cover status; other areas cannot $\Rightarrow$ need to identify riskier cases for their preventive inspection $\Rightarrow$ *Statistics*

- Different causes (e.g. digging) destroy bitumen cover and start corrosion process

FAILURES IN STEEL PIPES

- Data: 53 failures in 30 years on an expanding net, ≈ 380 Km long (year 2000)
- Three major factors related to failures
  - Age
    - continuous electrolytic process reducing wall thickness
  - Type of corrosion
    - natural corrosion
    - galvanic corrosion
    - corrosion by interference (or stray currents)
  - Lay location
    - near streetcar substations or train stations
    - o.w.

FAILURES IN STEEL PIPES: AGE

Physical aspects

- Different installation dates of different sections
  - same physical properties?
  - same installation procedure?
- Unknown date for start of corrosion process
- Different operating conditions
  - diameter, location, depth
  - electricity in the ground
  - kind of pipes (e.g. junction)

FAILURES IN STEEL PIPES: AGE

Mathematical aspects

- Different installation dates of different sections
  - A unique process or as many as (say) the installation years?
  - If many processes, one for each installation date or one starting (say) on July 1st every year?
  - Parameters for each process?
    - equal
    - similar (exchangeable)
    - completely independent
- Unknown date for start of corrosion process
  - model or ignore it?
- Different operating conditions
  - reasonable or feasible discriminating among them?

FAILURES IN STEEL PIPES

Bad data quality

- Diameter, location, depth and installation data of broken pipes sometimes unavailable
  $\Rightarrow$ *impossible to perform analysis similar to the one on cast-iron pipes (there data collected in 1991-96 and here on a larger period because of their scarceness and age dependence)*

- evolution of the net exactly known for the last 10 years and approximately since first installation (1930)
  $\Rightarrow$ *we assume it known, after interviewing company's experts, and performing linear interpolation, adjusted for the years of WWII*

- unknown installation date of few failed pipes
  $\Rightarrow$ *imputation with statistical methods or looking at nearby failed pipes*

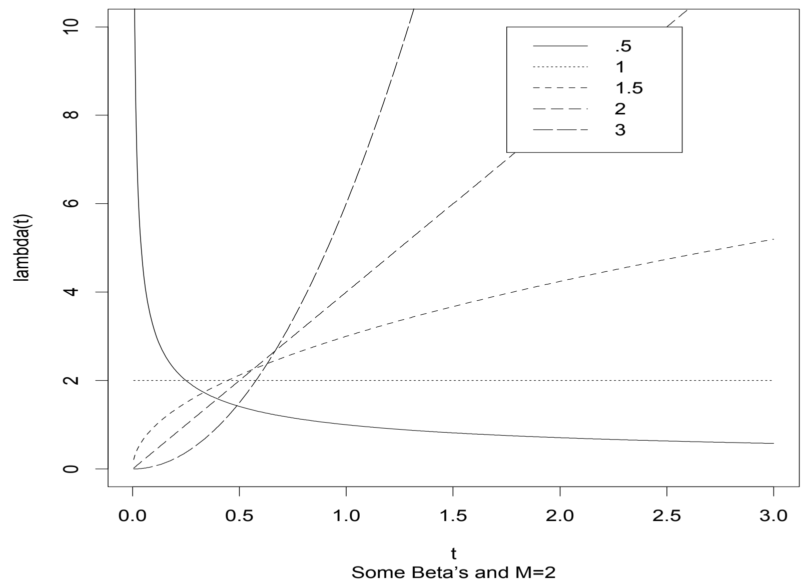
- (probably) improperly recorded escapes (e.g. six in 24 hours in different parts of the city, without any physical explanation, e.g. earthquake)
  $\Rightarrow$ *impossible to examine what happened so that we kept them as they were and statistical analysis had been badly affected*

FAILURES IN STEEL PIPES

- Network split into subnetworks based upon year of installation, *as if* pipes were installed on July, 1st each year
- Removals and replacements *not relevant* for network reliability $\Rightarrow$ *repairable system*
- Independent NHPP's for each subnetwork
- *Superposition Theorem*: Sum of independent NHPPs with intensity functions $\lambda_i(t)$ is still a NHPP with intensity function $\lambda(t) = \sum \lambda_i(t)$
- The same parameters vs. exchangeable ones in each NHPP
- Statistical analysis
  - parameter estimation
  - prediction of future escape times
  - computation of reliability measures

POWER LAW PROCESS

- $\lambda(t) = M\beta t^{\beta-1},\ M, \beta, t > 0$
- $\beta > 1 \Rightarrow$ reliability decay
  $\beta < 1 \Rightarrow$ reliability growth
  $\beta = 1 \Rightarrow$ constant reliability

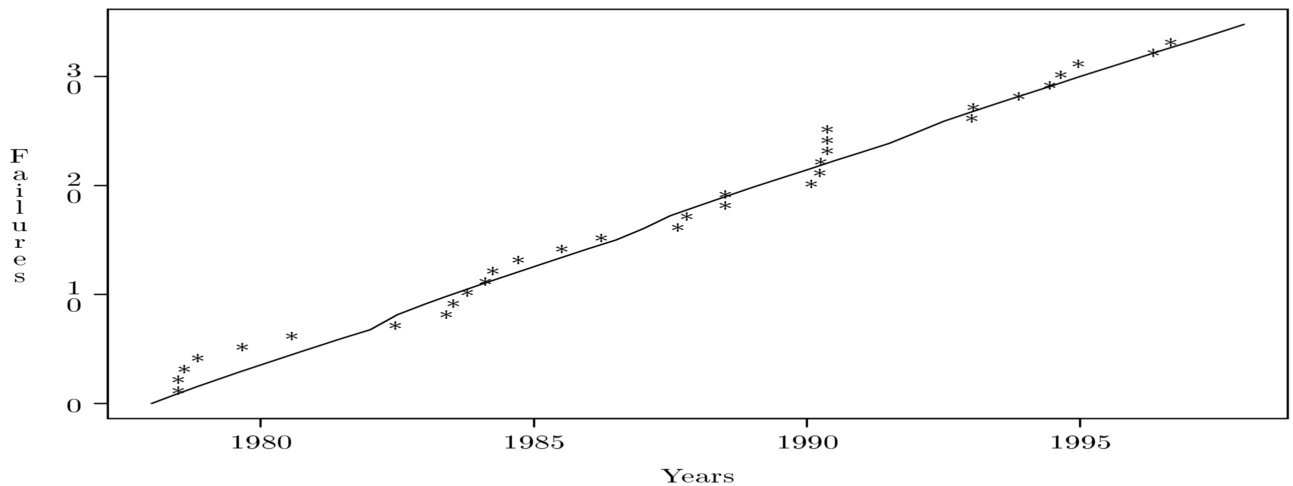


MODELS FOR STEEL PIPES

Exchangeable M and β; known installation dates

95% credible intervals for reliability measures:

- System reliability over 5 years: $P\{N(1998, 2002) = 0\} \Rightarrow [0.0000964, 0.01]$
- Expected number of failures in 5 years: $EN(1998, 2002) \Rightarrow [4.59, 9.25]$
- Mean value function (solid) vs. cumulative # failures (points)



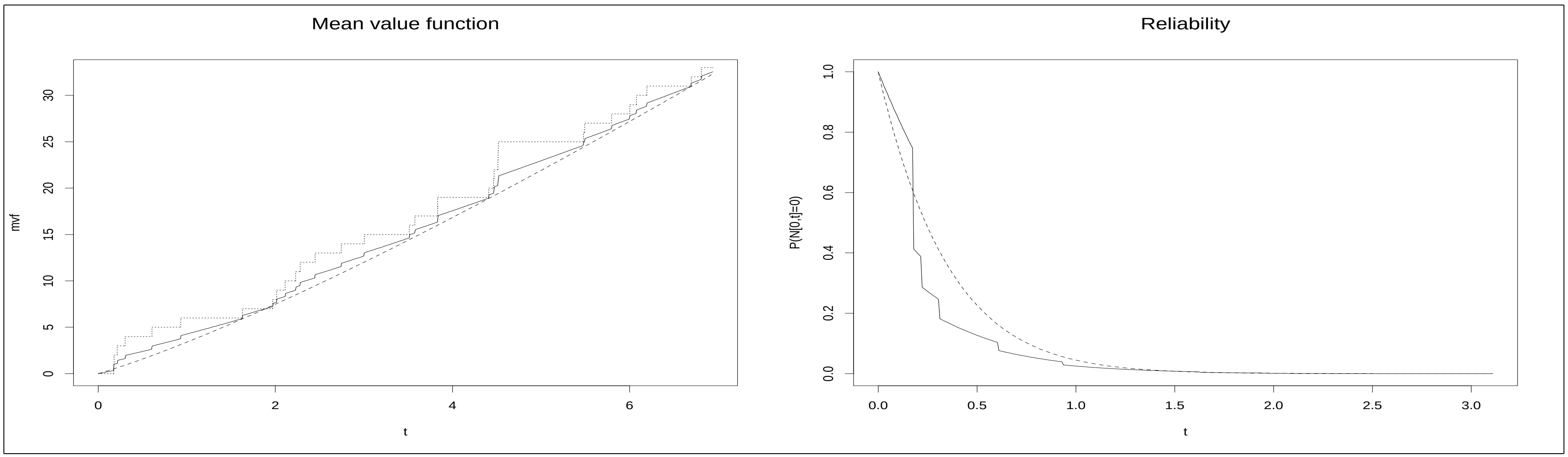
STEEL PIPES

Parametric NHPP: $\widetilde{\Lambda_\theta}(t) = \int_0^t [\tilde{a} \log(1 + \tilde{b}t)]dt + \hat{c}t$

Nonparametric model: $M \sim \mathcal{P}_{\alpha,\beta} : \alpha(ds) := \widetilde{\Lambda_\theta}(s)/\sigma\, ds,\ \beta(s) := \sigma$

$\Rightarrow \mathcal{E}MS = \widetilde{\Lambda_\theta}S$ and $VarMS = \sigma\widetilde{\Lambda_\theta}S$

$\Rightarrow MS$ “centered” at parametric estimator $\widetilde{\Lambda_\theta}S$ and closeness given by σ



Nonparametric (solid) and parametric (dashed) estimators and cumulative $N[0,t]$ *(dotted).*