Inverse modelling and its potential to fill knowledge gaps concerning the emission of environmental pollutants.

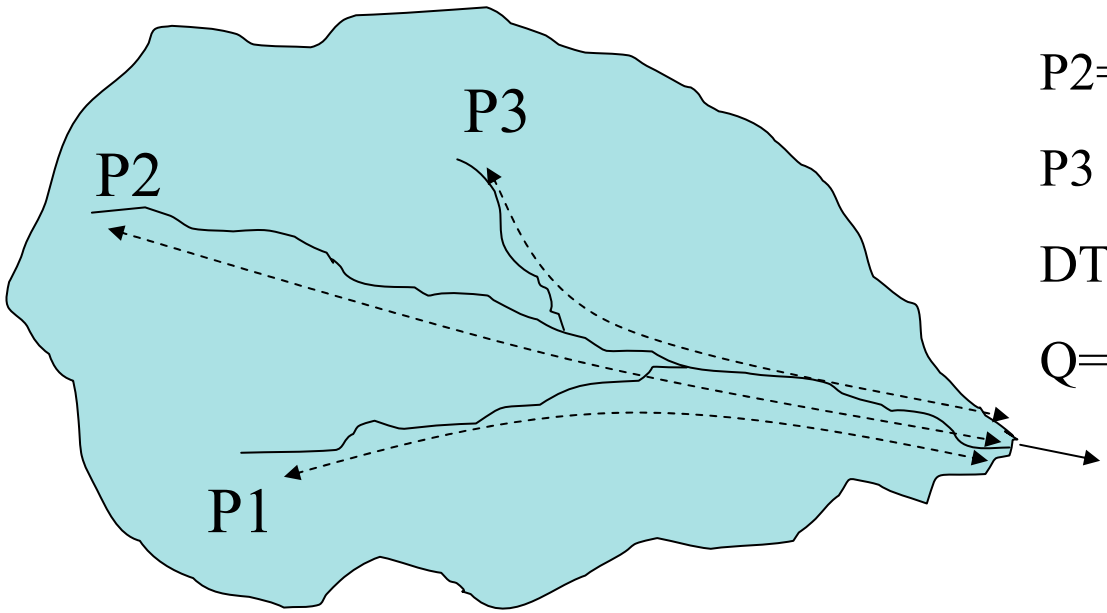
A.Pistocchi

GECOsistema srl, Italy

www.gecosistema.eu

(Formerly EC JRC – IES, Ispra, Italy)

Prediction of water concentration from known emissions: example



P1= 700 ug/s, T1=10 d

P2= 400 ug/s, T2=3 d

P3 = 600 ug/s, T3=4 d

DT50= 5 days

Q=100 m^3/s

$$C = \frac{700\, e^{-\frac{\ln 2}{5}10} + 400\, e^{-\frac{\ln 2}{5}3} + 600\, e^{-\frac{\ln 2}{5}4}}{100} = 7.8\, ug/m^3$$

This is how direct models work

- Models can be more complicated, but the basic idea is as above
- Direct models **believe in prior emission estimates which are an input**
- They fine-tune parameters (e.g. DT50, residence time) to match observed C

GECOsistema - Geographic_Environmental_COnsulting
Viale Carducci, 15 - 47023 - Cesena (Italy)
Tel (+39) 0547 22619 - Fax (+39) 0547 22619
Sede legale Piazza Malatesta 21, 47900 Rimini (P.IVA 03236780403)

Direct models

- The most commonly thought-of type of models
- Usually one assumes emissions and calibrates removal rates
- A plethora of examples, some with excellent results in applications (e.g. GREAT-ER)
- Require accurate emission inventories
- May achieve high accuracy in predictions when properly calibrated (but a lot of efforts! + data demand)

Knowledge of emissions

- EPER/E-PRTR or other inventories
- Expert estimates (e.g. models of emission processes)
- Industry data (e.g. REACH reporting, pre-REACH Risk assessment reports)
- EUROSTAT/national stat. data
- …

GECOsistema - Geographic_Environmental_COnsulting
Viale Carducci, 15 - 47023 - Cesena (Italy)
Tel (+39) 0547 22619 - Fax (+39) 0547 22619
Sede legale Piazza Malatesta 21, 47900 Rimini (P.IVA 03236780403)

Limitations of inventories

- For instance, E-PRTR covers only 91 substances, or roughly 0.5% of the chemicals of potential concern
  - Greenhouse gases
  - Other gases
  - Heavy metals
  - Pesticides
  - Chlorinated organic substances
  - Other organic substances
  - Inorganic substances.
- The diffuse (*) part of emissions is not known

(*) diffuse = “having no apparent legal responsible”

Unknown emissions→ Estimation required!

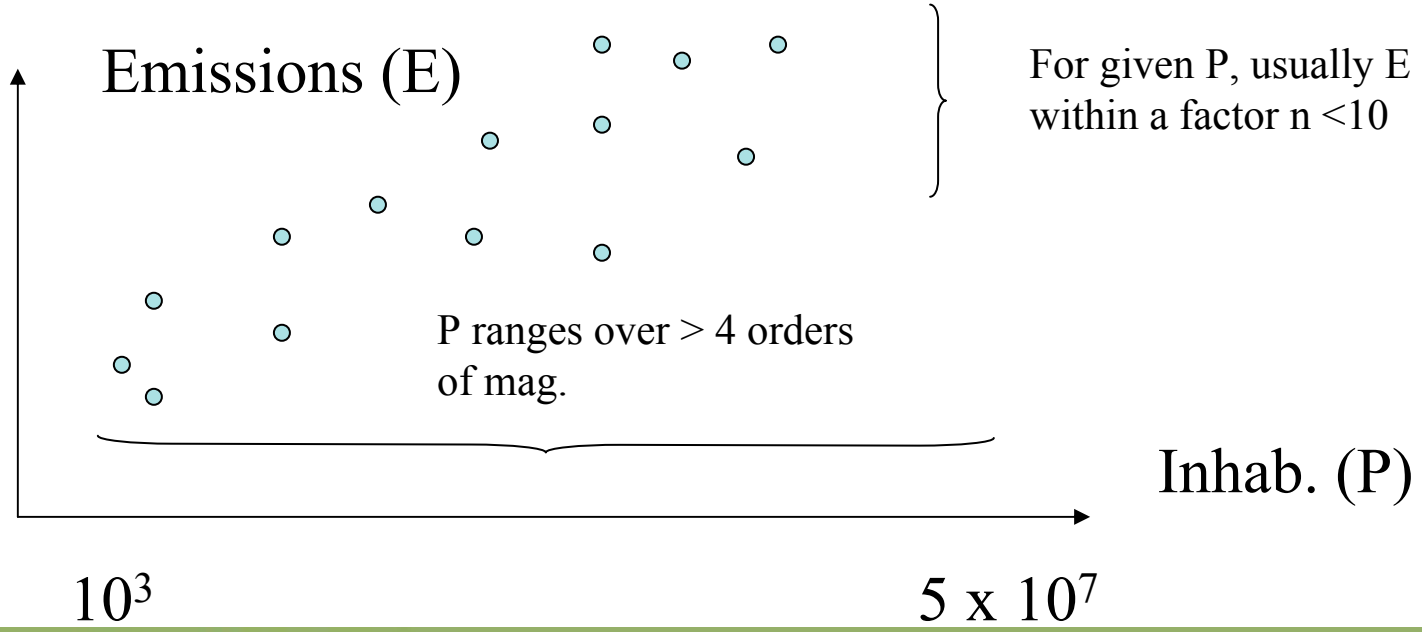
- Bottom-up: modeling of emission factors indirectly from statistics such as sales or production, or from theoretical emission models.
- Top-down: use information on observed concentrations
- (a midway “data assimilation” might be using a blend of prior estimates and observations)

- „Direct“ models introduce prior knowledge on emissions and removal rates and predict concentrations; prior knowledge is adjusted to match observed concentrations (model calibration)

- „inverse “ models assume emissions and removal rates depend on a limited number of parameters, and back-calculate these parameters from observations

Now, let's assume we don't know emissions but we know C

- This is quite a common situation…
- We can still assume emissions are proportional to population



Theoretically, we can compute emissions <u>for a given k or DT50</u>

P1= 1000 inhab, T1=10 d

P2= 2500 inhab, T2=3 d

P3 = 700 inhab, T3=4 d

DT50= 5 days

Q=100 m^3/s

η=emission factor (ug/inhab/s, unknown)

$$C = \frac{700\, e^{-\frac{\ln 2}{5}10} + 400\, e^{-\frac{\ln 2}{5}3} + 600\, e^{-\frac{\ln 2}{5}4}}{100} = 7.8\, ug/m^3$$

$$C = \frac{1000\eta\, e^{-\frac{\ln 2}{5}10} + 2500\eta\, e^{-\frac{\ln 2}{5}3} + 700\eta\, e^{-\frac{\ln 2}{5}4}}{100} = 7.8\, ug/m^3$$

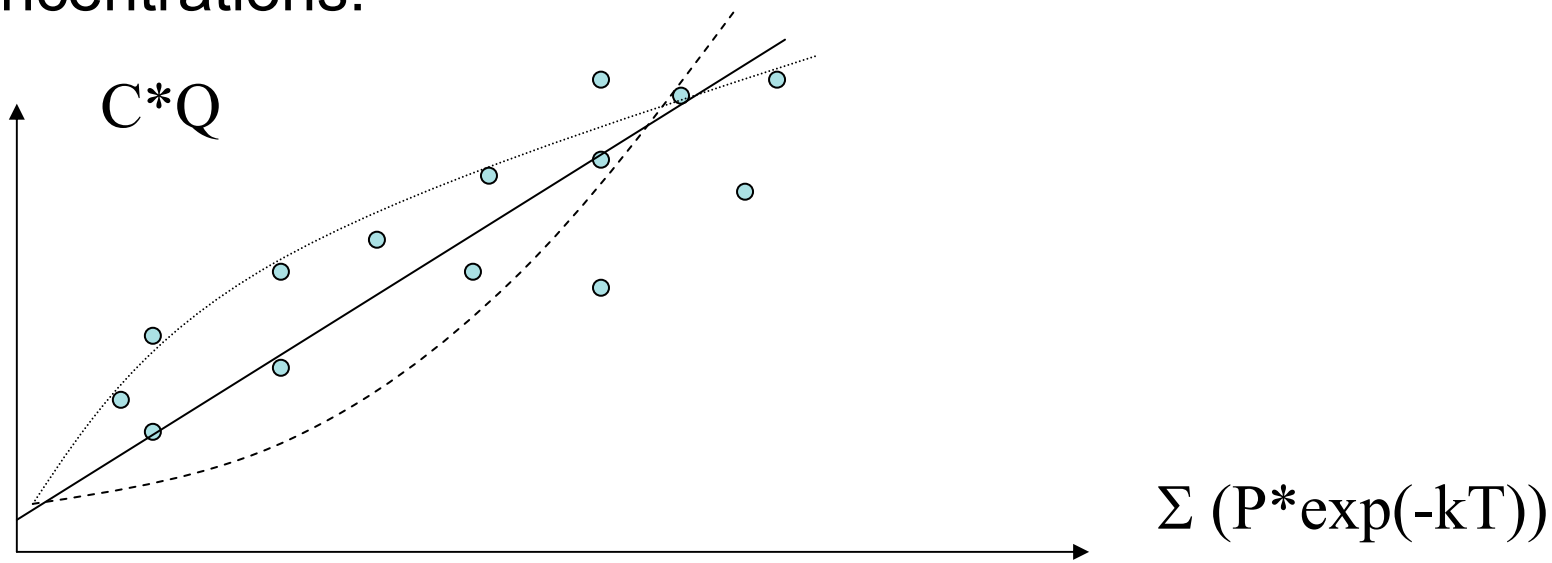
$$\eta = \frac{7.8}{\left(\frac{1000\, e^{-\frac{\ln 2}{5}10} + 2500\, e^{-\frac{\ln 2}{5}3} + 700\, e^{-\frac{\ln 2}{5}4}}{100}\right)} = 0.34 ug/inhab./s$$

In reality…

- If we repeat the same calculation for different points where C is monitored, we obtain different η
- This is due to the fact that emissions per capita are not always constant
- Moreover, many other sources of error introduce variations (sampling and measurement, variations in half life, errors on water travel time…)

…however

- We generally expect that the sum of populations upstream of a sampling section, each multiplied by its decay, and divided by water discharge, be related to concentrations:



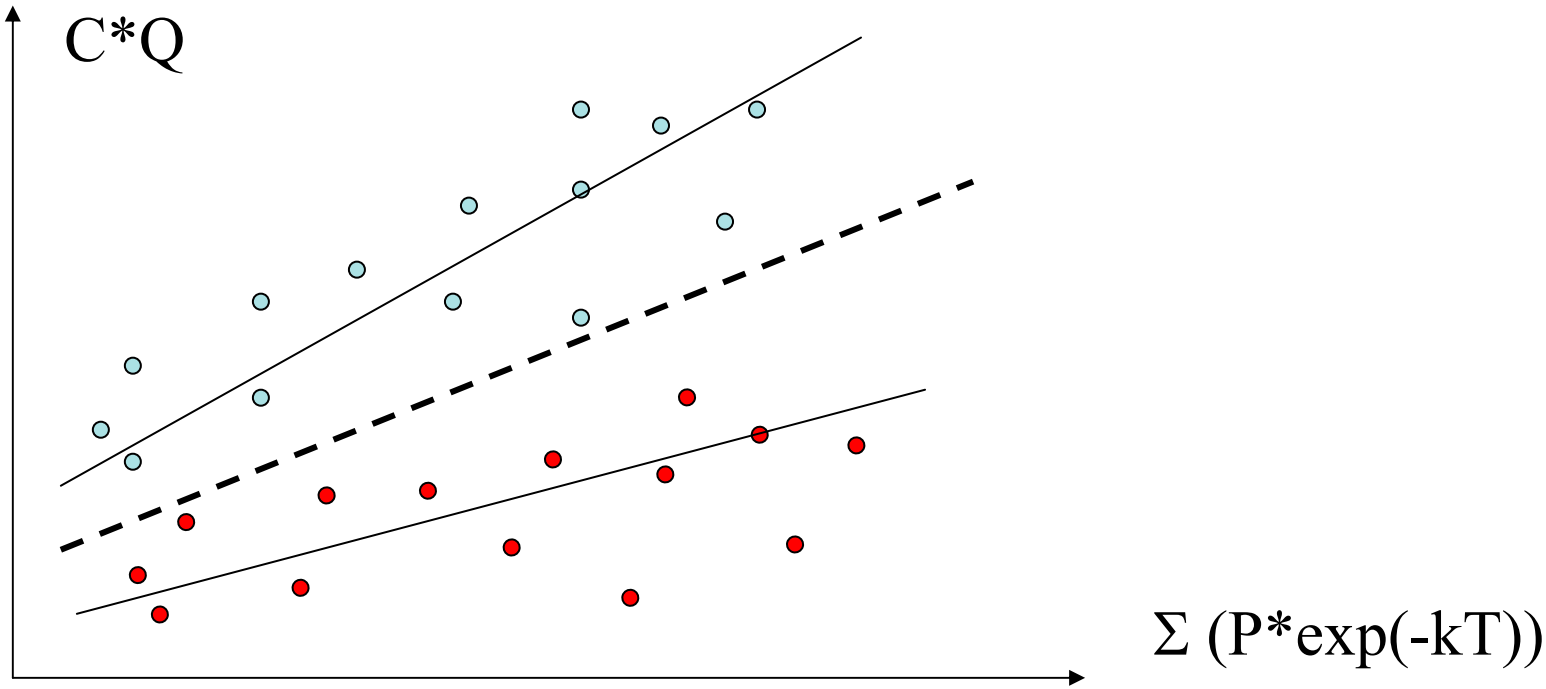
Just emission factors?

- In principle, the exercise should be repeated for any DT50; usually, a few values are enough to understand the general behavior
- Other model parameters such as water residence or travel time are usually less uncertain and can be assumed to be known as a first approximation

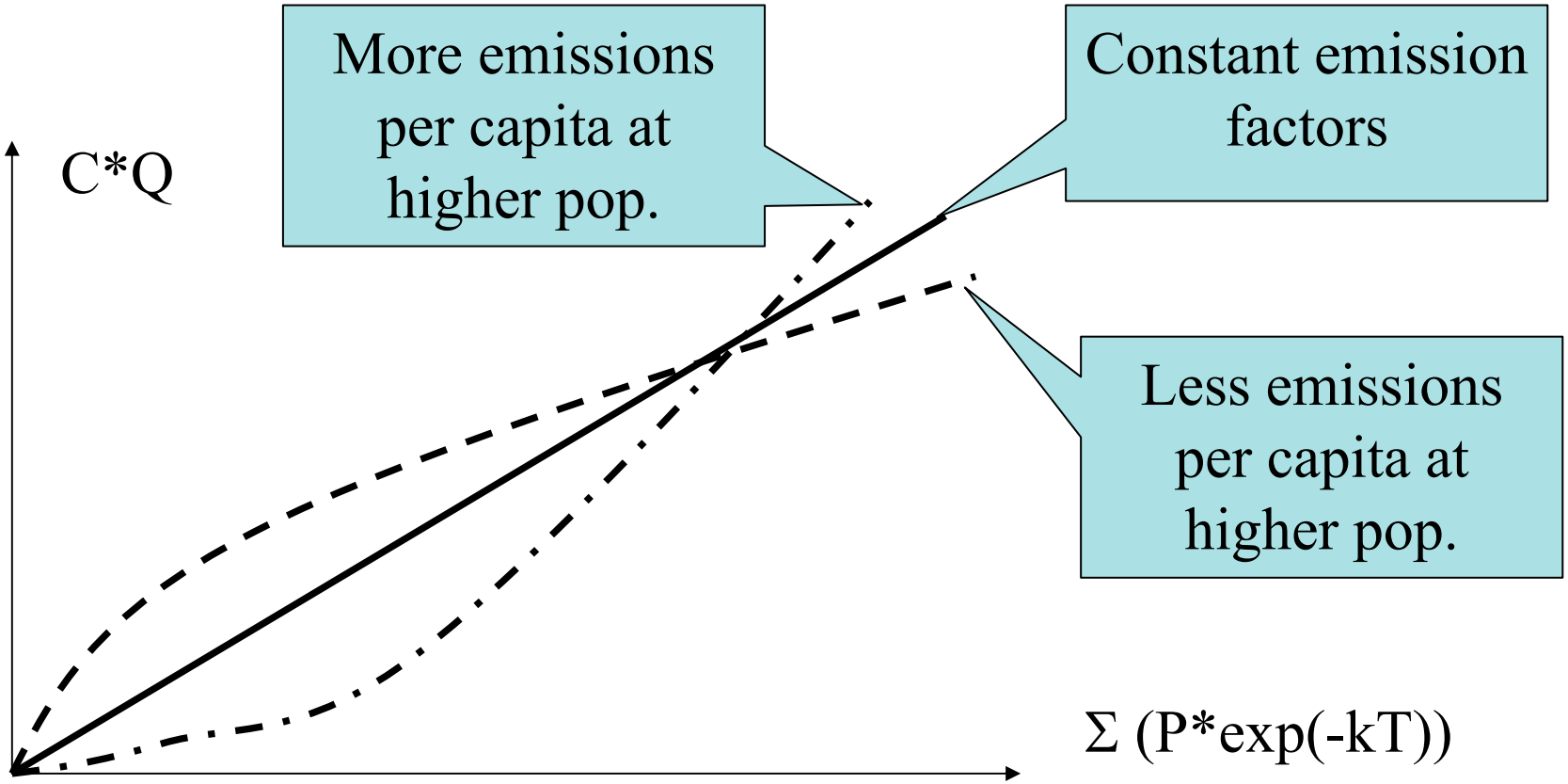
Just population?

- In principle, the most reasonable proxy of emissions should be used
  - Land use
  - Sludge application
  - Agriculture/crops/pesticide use rates
  - Lights at night…
- The point is to identify a variable to which we assume emissions are proportional
- It all depends on the information we have, and the intended use of the chemical

Case of significant variations in emission factor distributions



The meaning of the shape of the trend



Direct models

Where do chemicals go?

Estimate emissions E

Estimate environmental removal rates K

Estimate environmental concentrations C = E/K

Requires Emission estimates

Can Be Used To Simulate Scenarios

Uses Monitoring Data For “Validation”

Deals With “Effects” from a Given “Cause”

> **Indirect Application For Policy Support**

Inverse models

Where do chemicals come from?

Take observed concentrations C

Estimate environmental removal rates K

From observed concentrations, E = C/K

Requires Monitored Concentrations

Helps Identifying Relevant Scenarios By Detecting Emission Sources

Uses indirect evidence of emissions for validation (e.g. known chemical production volumes, sales of pesticides, pharmaceuticals…)

Emissions Are A **Direct Policy Variable**: If A Chemical Is A Problem, Areas Of Action Are The Ones Of Emission

Common applications of inverse models

- Calculation of atmospheric emissions
- Parameter estimation for groundwater flow and chemical transport
  - In these applications, use of high complexity numerical models is made and parameters are estimated from observations through complex (e.g.variational) error minimization methods
- The JRC GREEN model (Grizzetti et al., 2008)

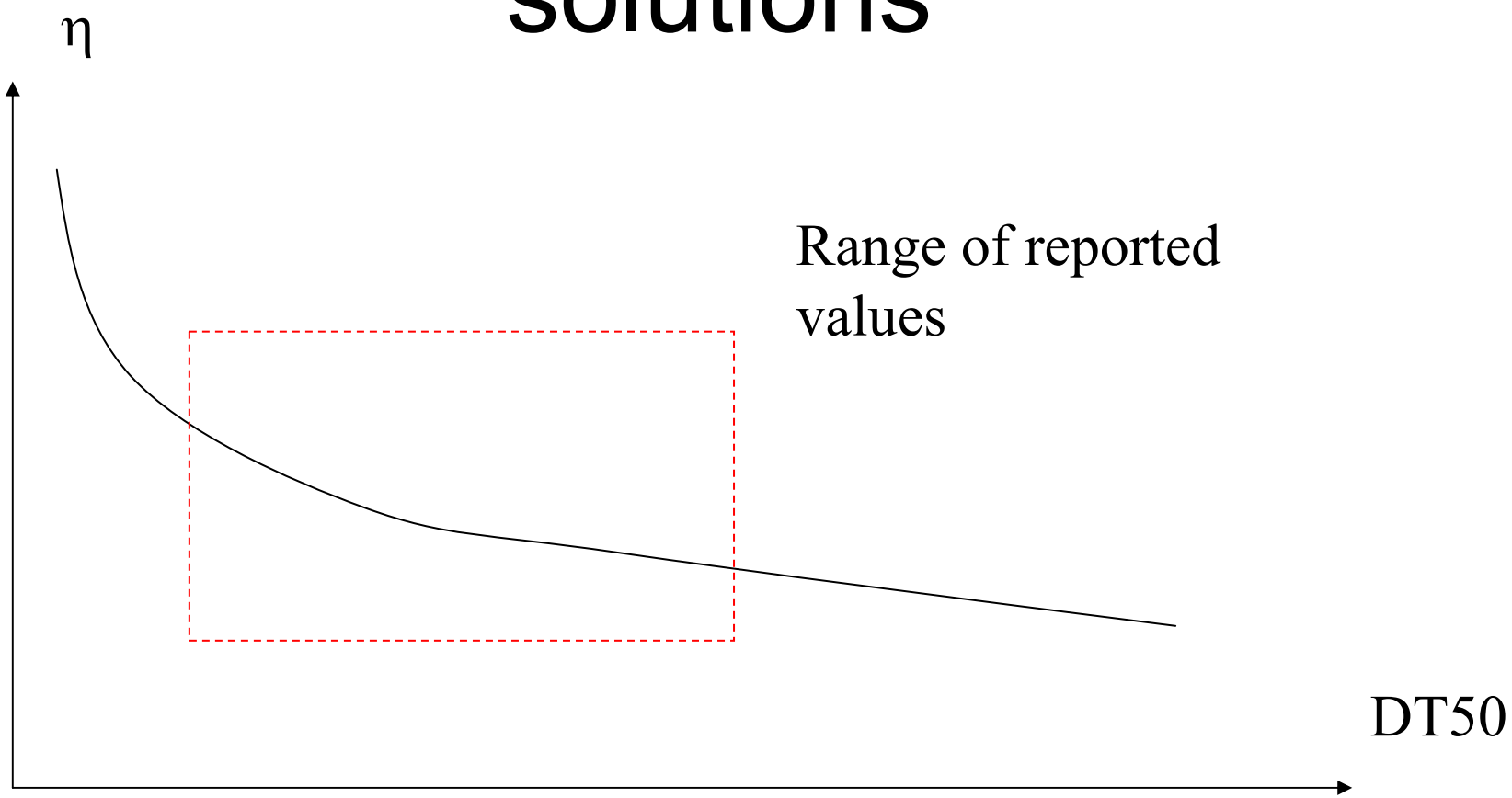
Methods

- Parameters are estimated by minimizing some error indicator, such as the sum of squared differences between observed C and a functional combination of input variables
- A particularly simple case is the ordinary least squares procedure (linear or linearizable models only)
- More complex procedures derived from the data assimilation practice are used for complex models (e.g. for atmospheric emissions or groundwater)

An inverse problem is ill-posed

- For instance, an infinite number of combinations of DT50 and emission factor may yield the same results, i.e. they are non-dominated (Pareto-optimal)
- To narrow the ensemble of solutions, one has to apply ***prior knowledge***, e.g. on DT50 or total production volume

Narrowing the ensemble of solutions



Inverse models and monitoring

- For water concentrations, a simple model can be set up
- The inversion of the model to obtain a non-dominated set of emission factors and half lives is quick, rather simple and easy to interpret
- A feasible standard practice in any evaluation of monitoring data:
  - Compare the observed load with the suspected sources of emission upstream
  - Essential to get not just C, but also Q!