Integration of QSARs with risk assessment

CADASTER Work package 4

Tomas Öberg

Linnæus University

Work package 4 - tasks

1. QSAR models in a probabilistic risk assessment framework (LnU)
2. Testing ECETOC's TRA tool (IVL)
3. Economic valuation of impacts (RU)
4. QSAR models in the legal framework (IVL)
5. Policy and management (RIVM)

Focus in this presentation on 1-3, where results are available

Linnæus University

IVL Swedish Environmental Research Institute

Radboud University Nijmegen

Rijksinstituut voor Volksgezondheid en Milieu
Ministerie van Volksgezondheid, Welzijn en Sport

Work package 4 - deliverables

1. Application of QSAR models for probabilistic risk assessment, report (LnU, month 40)
2. Guidance on using QSAR models for probabilistic risk assessment, report (LnU, month 48)
3. Evaluation of ECETOC's TRA tool, report (IVL, month 24)
4. Evaluation of options for economic valuation of chemical impacts, report (RU, month 30)
5. Evaluation of QSAR models in the legal framework, report (IVL, month 36)
6. Synthesis of research findings and recommendations for prioritization, report (RIVM, month 48)

Task 4.1 - QSAR models in a probabilistic risk assessment framework

Approach

Stage 1. State-of-the-art

- Methodology of probabilistic environmental risk assessment (PRA)
- Uncertainties in QSARs
- Methods to characterise uncertainty in QSARs
- OECD principles on uncertainty

Stage 2. Case-studies

- Suggest and evaluate methods to characterise uncertainty related to QSARs

Stage 3. Implementation

- Integrate QSAR into probabilistic risk assessment within REACH
- Evaluate the role of uncertainty from QSARs

Task 4.1 - QSAR models in a probabilistic risk assessment framework cont.

Progress

Stage 1. State-of-the-art

- Literature review completed
- Results presented as a poster at Euro-QSAR 2010
- Method review published in *Molecular Informatics*

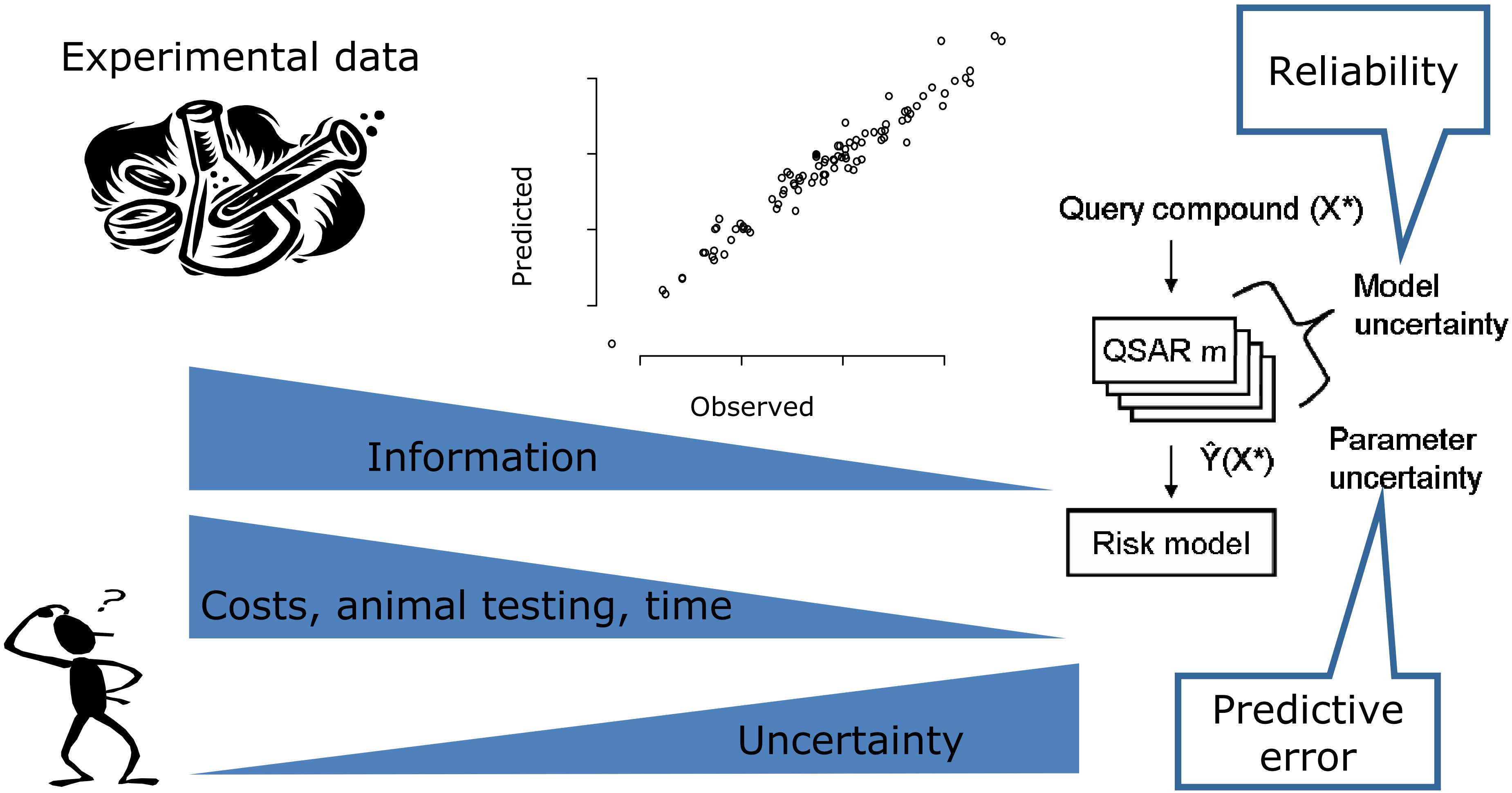
Stage 2. Case-studies

- Compilation of code for multiple method characterisation of prediction uncertainty implemented on data and models
- Preliminary results presented as a poster at OpenTox workshop 2011

Stage 3. Implementation

- Level I, II and III fugacity models set up in RiskCalc
- Evaluated with some ”standard” pollutants, manuscript submitted
- Test case with PBDE being prepared for MC-evaluation with EUSES

Uncertainty in QSARs – a balancing act



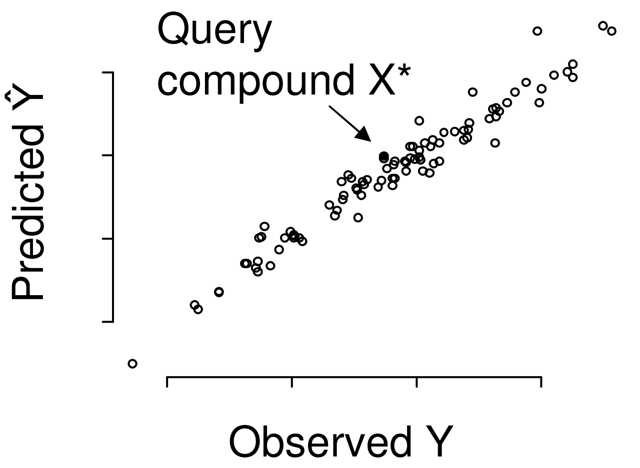
Relevant questions

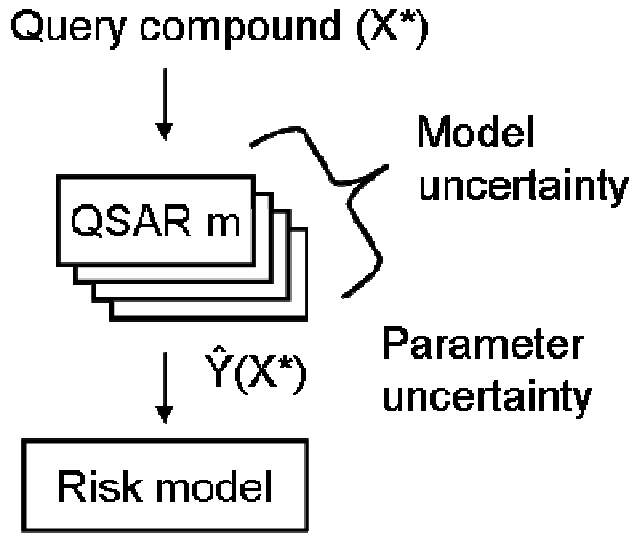
- What kind of characterizations of predictive uncertainty are suitable for risk assessment or weight-of-evidence approaches?
- What measures of reliability are useful?
- Which methods for characterization of predictive uncertainty works best and which are most appealing to potential end-users?
- When does it matter how predictive uncertainty has been characterized?

Predictive uncertainty – a risk assessment perspective

Parameter uncertainty – uncertainty in predicted values of query compound

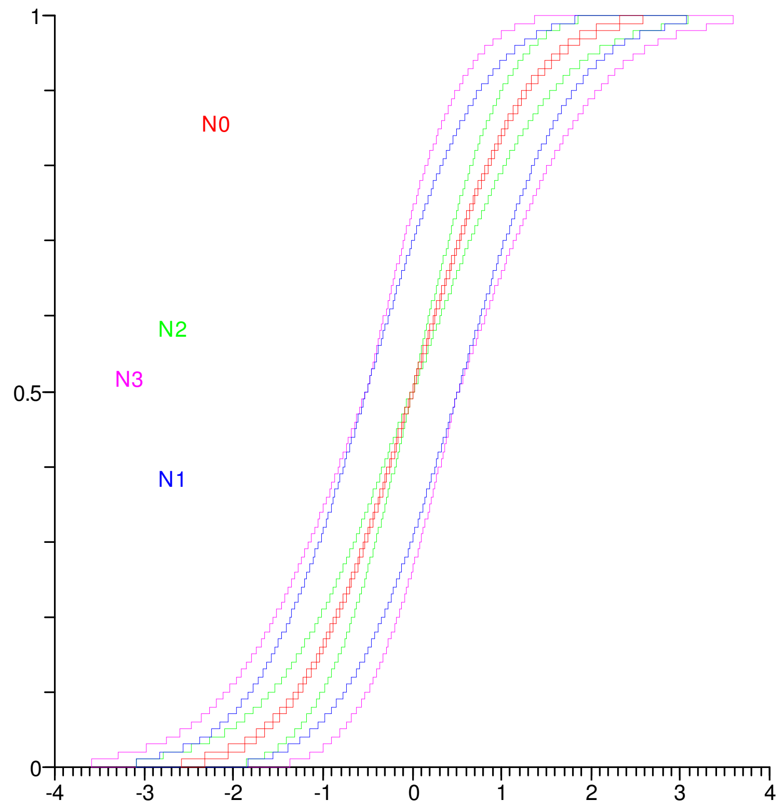
Model uncertainty – uncertainty in using the QSAR to predict the query compound





Predictive uncertainty – characterizing error in a predicted value

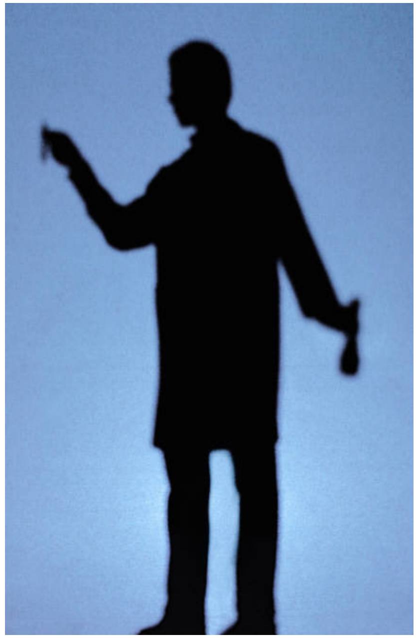
- Predictive probability distribution
- 2-Dimensional predictive probability distribution
- Interval or fuzzy number
- Combination of these – probability box
  - N0=N(0,1)
  - N1=N([-0.5,0.5],1)
  - N2=N(0,[0.8,1.2])
  - N3=N ([-0.5,0.5], [0.8,1.2])



Methods to assess predictive uncertainty

Predictive distribution may be assessed

1. from **estimates of predictive variance** (e.g. by sampling or re-sampling)
2. directly as **probability distributions**
3. based on experimental data – **expert judgment**



Predictive variance depends on the applicability domain

Sahlin, U., Filipsson, M., Öberg, T. A risk assessment perspective of current practice in characterizing uncertainties in QSAR regression predictions. Molecular Informatics 30, 551-564 (2011).

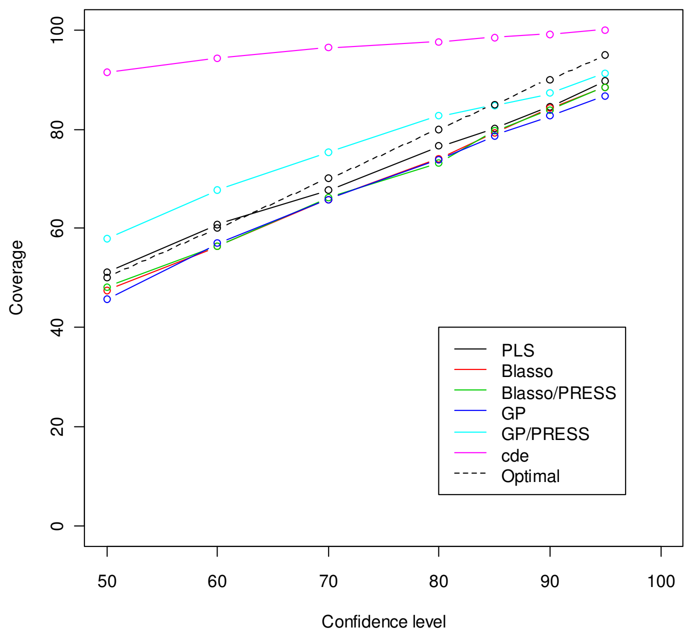
Theoretical and statistical aspects

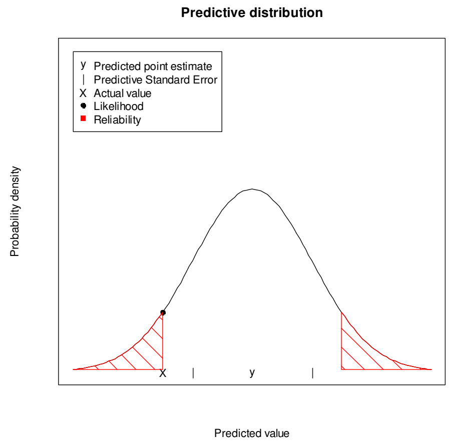
Bayesian probabilistic approaches
Pros: Assess uncertainty directly based on data, and prior knowledge. Can combine empirical data and expert judgement.
Cons. Difficult to implement in practise, requires understanding of difficult mathematical language.

Empirical approaches
Pros: Works with any type of underlying algorithm, Can be straightforward to calculate.
Cons: Sampling sensitive to the availability and choice of test set

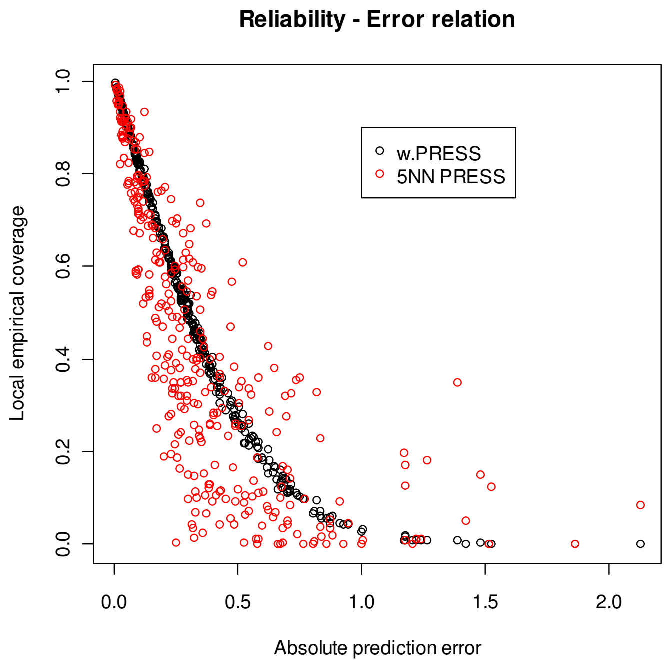
Predictive uncertainty – characterizing reliability in prediction

Global coverage ← **Predictive distribution** → Local coverage





(weighted) PRedicted Error Sum of Squares, Conditional Density Estimator, Gaussian Process, Bayesian Lasso



Empirical evaluation

- Comparison of approaches and methods
  - CADASTER data sets
  - external data sets
- Some preliminary findings
  - given a well trained predictive model, the choice of method was not critical
  - for small datasets, however, the probabilistic approach performed better

Any recommended method(s) must be general enough to encompass a range of different model building approaches

How does the QSAR uncertainty affect the risk assessment?

- Will be evaluated by:
  - Discrete uncertainty calculations (fuzzy/PBA) using level I-II(-III) fugacity models
  - Monte Carlo simulation in the EUSES spreadsheet (level III model)
- Data input (PBDEs, PFCs):
  - CADASTER QSAR models (predictive distributions)
  - Other QSAR models (e.g., EPI Suite, predictive distributions or p-boxes)
  - (Experimental data)
- Quantification:
  - Regional risk characterization ratios (PEC/PNEC)
  - Sensitivity to uncertainty in the QSAR estimates

Task 4.2 Testing ECETOC's TRA tool

- An Excel-based application for targeted risk assessment
  - Worker exposure (also in ECHA's Chesar plug-in to IUCLID)
  - Consumer exposure
- Evaluation of environmental part in the tool
  - Validated against a level III fugacity model for brominated phenols
- Report (2010)
  - Discuss also usability and needs for improvement
  - Available for download at CADASTER web site

Task 4.3 Economic valuation of impacts

- Aim
  - To exploring the possibilities for assessing economic costs of chemical impacts
- Case study
  - Impacts of PBDEs on the peregrine falcon population of California
- Why this species/population?
  - High PBDE concentrations in eggs
  - Increasing trend in PBDE concentrations
  - Data availability

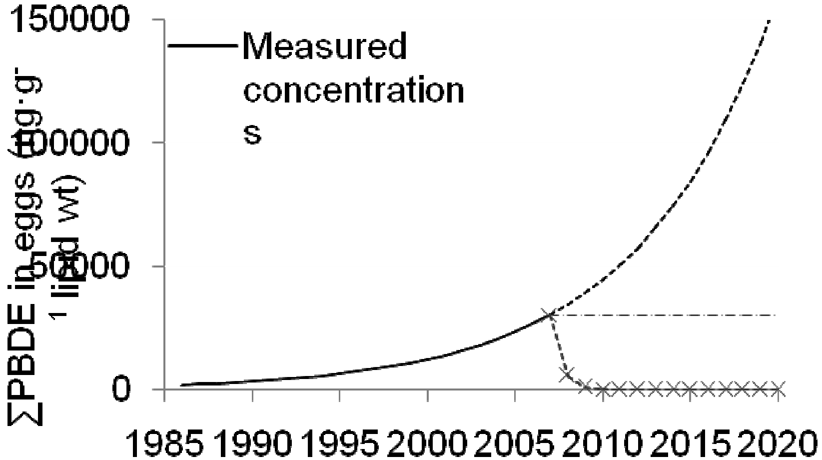
Falco peregrinus

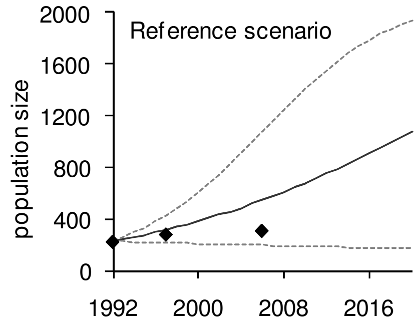
Population model

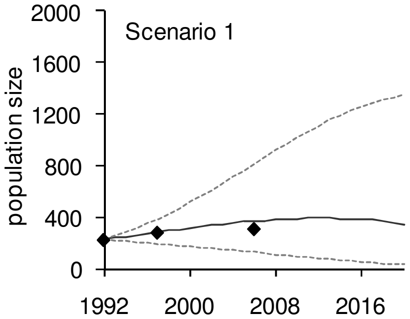
$$N_{t,C} = N_{t-1,C} + \frac{r(C)}{r(0)} \cdot r(0) \cdot N_{t-1,C} \cdot (1 - \frac{N_{t-1,C}}{N_\infty})$$

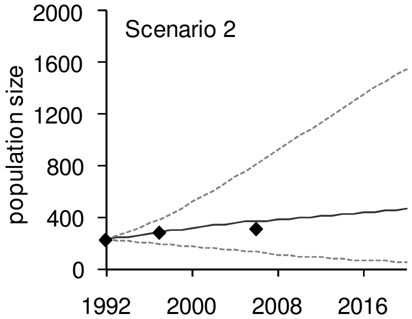
- $N_{t,C}$, number of falcons under contaminated conditions
- R(C), rate of increase under contaminated conditions
- N_∞, carrying capacity of the area
- Probabilistic approach to account for uncertainties in ecological and toxicological model parameters

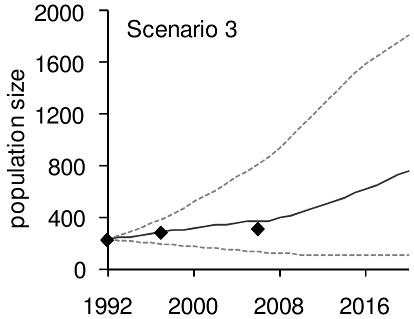
Three exposure scenarios



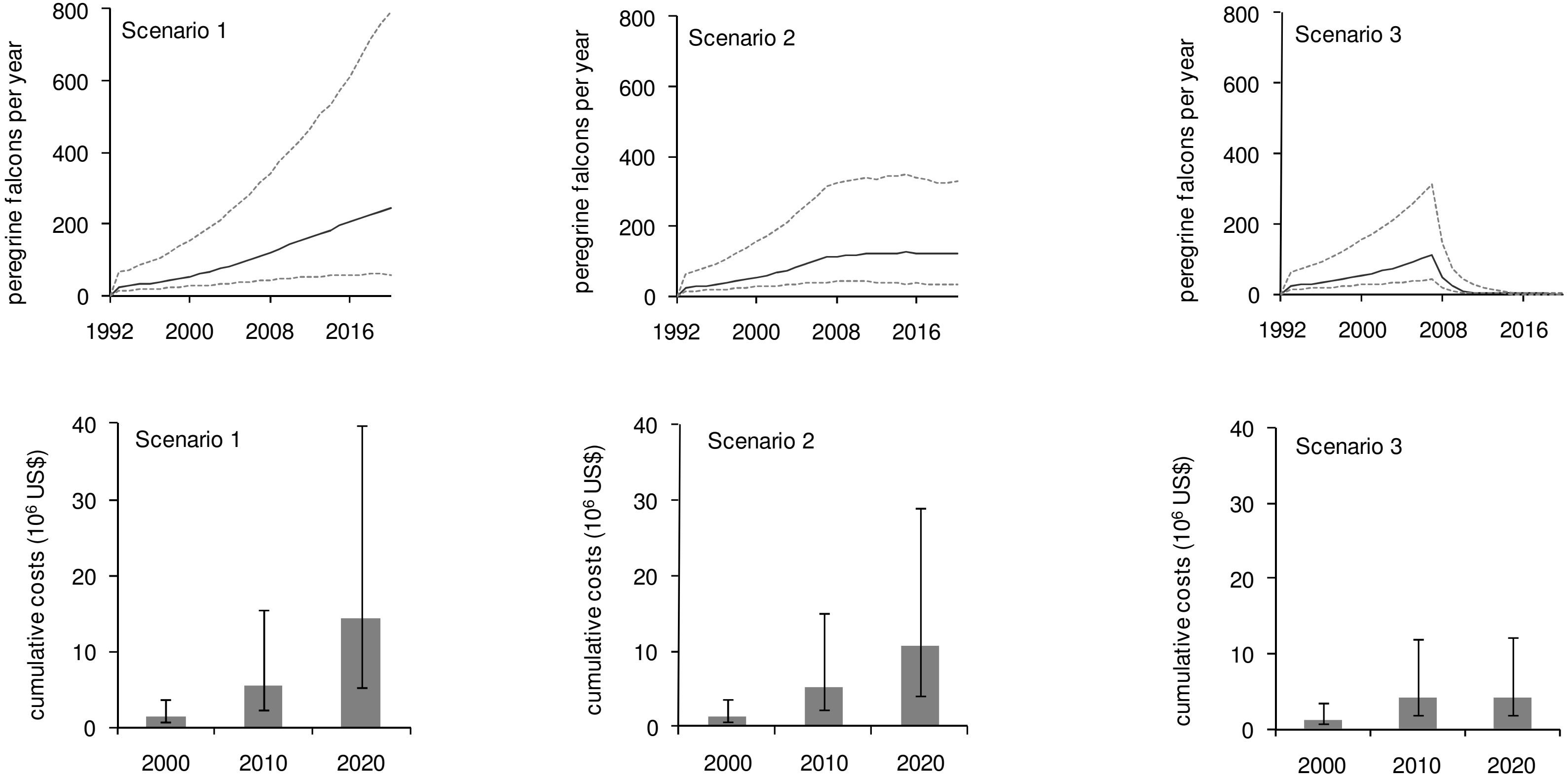








Results – replacement with captive-bred birds and associated costs



Conclusions – economic valuation

- Model provides a relatively straightforward approach to put economic value on chemical impacts on animal populations
- Uncertainties in the input data may considerably influence the outcomes
- Reliable data are important to obtain reliable cost estimates

Report (2011) submitted and soon available at the CADASTER web site

Work in progress and remaining deliverables

- Report on the Application of QSAR models for probabilistic risk assessment, report (LnU, April 2012)
- Report on Guidance on using QSAR models for probabilistic risk assessment, report (LnU, December 2012)
- Report on the Evaluation of QSAR models in the legal framework, report (IVL, December 2011)
- Report on the Synthesis of research findings and recommendations for prioritization, report (RIVM, December 2012)

Acknowledgements

- All CADASTER groups contribute to WP4
- In the preparation of this presentation, inputs and slides were received from:
  - Aafke Schipper and Mark Huijbregts at Radboud University
  - Magnus Rahmberg at IVL – Swedish environmental Research Institute
  - Ullrika Sahlin at Linnaeus University