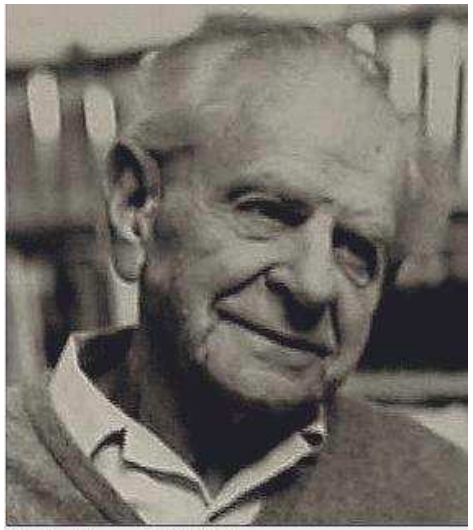
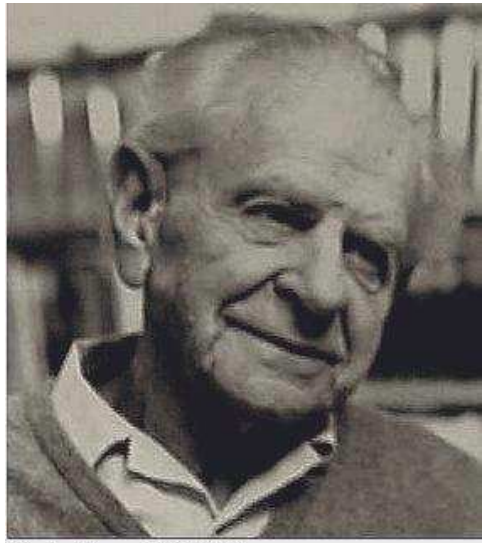
Is it possible to falsify complex models?

Karl Popper (1902-1994)



Sir Karl Popper (1902-1994)

For Popper, a theory is scientific only if it is refutable by a conceivable event



Sir Karl Popper (1902-1994)

According to Popper, scientific progress involves the **abandonment of partially true, but falsified, theories**, for theories with a higher level of verisimilitude, i.e., which **approach more closely to the truth**.

From Stanford Encyclopedia of phylosophy

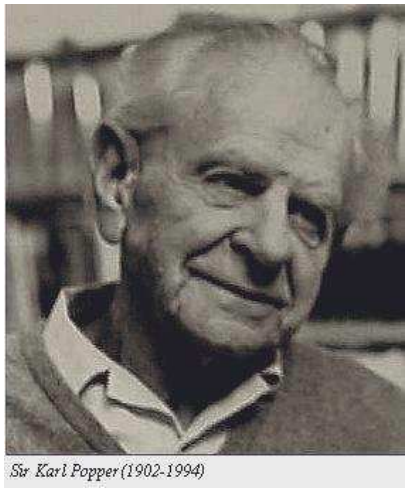
Thus, if a complex model is a scienfic theory, it should be possible:

1. to falsify this model,
2. to measure how close this model is from the truth

Can we do that?

Do we need to do that?

Georges Box's view:
Essentially, all models are wrong, but some are useful



Sir Karl Popper (1902-1994)

Popper's questions	Box's answers
Possible to falsify models?	We do not need to answer this question: we already know they are all wrong
Possible to measure how close the model is from truth?	Not the right question: we need to measure usefulness

New questions:

Is it possible to measure the usefulness of complex models?

How to do that?

Many difficulties

- Define criteria to measure usefulness
- Collect data
- Run models and Estimate criteria from data
- Select a model

Example: Relative RMSE of grain yield prediction (%) for several crop models

AZODYN (Makowski *et al.*, 2008)	**20.5**
Static model (Makowski *et al.*, 2008)	**18.2**
STICS (Brisson *et al.*, 2002)	**18.2**
SUCROS (Jamieson *et al.*, 1998)	**14.8**
CERES (Jamieson *et al.*, 1998)	**13.5**

Not easy to choose

RMSE values are only estimated values

- RMSE values are sample dependent
- Models not used in the same conditions (cultivars, soils, years, cropping systems)
- Not the same data used to compute RMSE

No unique method to run a complex model

- No unique method for measuring input variables → Not the same levels of measurement error
- No unique method for parameter estimation → Not the same levels of errors in parameter estimates

RMSE differences do not necessarily reflect differences in model quality

- RMSE differences may be due to chance
- RMSE differences may be due to differences in quality of data used to compute RMSE
- RMSE differences may be due to differences in quality of input data
- RMSE differences may be due to differences in quality of parameter values
- RMSE differences may be due to differences in the skills of the model users

Debate

Do we need to measure how close complex models are from the truth?

Do we need to assess model usefulness?

How to do that?

Can we really assess complex models?