Thorpex Montreal 2004

How can we best combine forecasts for added value?

J. Broecker[1], L. Clarke[1],
D. Kilminster[2] and L.A. Smith[1,2]

1. Department of Statistics, London School of Economics, UK
2. Pembroke College, Oxford, UK

Thorpex Montreal 2004

How can we best combine forecasts for added value?

How might we use TIGGE?

J. Broecker[1], L. Clarke[1],
D. Kilminster[2] and L.A. Smith[1,2]

1. Department of Statistics, London School of Economics, UK
2. Pembroke College, Oxford, UK

Motivation

Forecast improvement can be achieved in two ways:

Motivation

Forecast improvement can be achieved in two

Motivation

Forecast improvement can be achieved in two ways:

- improving the models (strategic)
- using the available information more effectively (tactical)

Motivation

Forecast improvement can be achieved in two ways:

- improving the models (strategic)
- using the available information more effectively (tactical)

THORPEX:

> “THORPEX will develop, demonstrate and evaluate a multi-model, multi-analysis and multi-national ensemble prediction system, referred to as TIGGE.”

Overview

- Combining Simulations
- Evaluation
  - skill scores $\longrightarrow$ Broecker
  - bootstrapping and meaningful skill comparison
- Example - combining ECMWF and NCEP

Inputs

Motivation

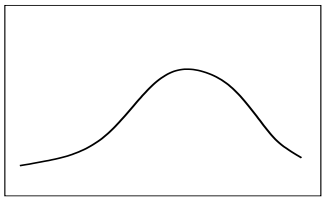
Overview

Combining

Evaluation

Example

Climatological Distribution



Inputs

Motivation

Overview

Combining

Evaluation

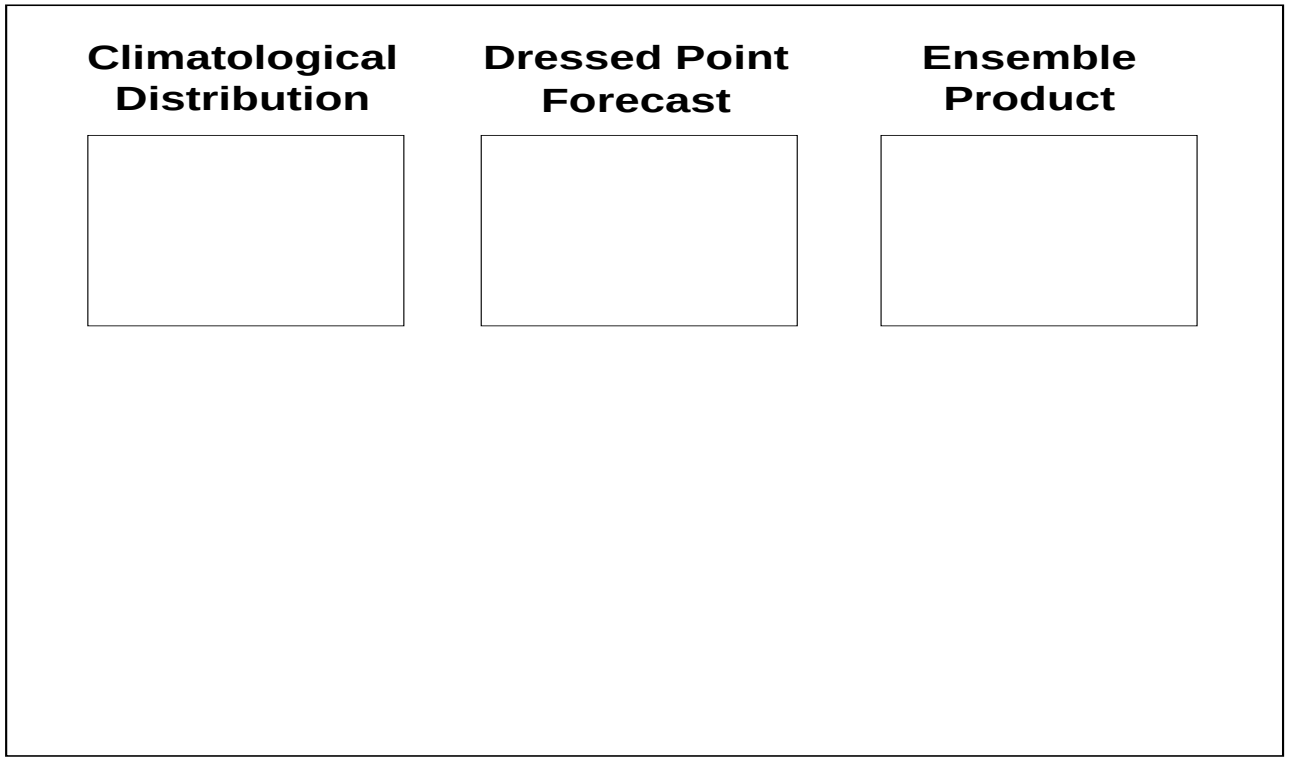
Example

Climatological Distribution

Dressed Point Forecast

Inputs

- Motivation
- Overview
- Combining
- Evaluation
- Example



Combining: Skill Scores

The combination is based on the skill of the final forecast

$$s = \mathcal{S}(f, o)$$

f forecast distribution

o verth7ed4042 0 Td (er)Tj 0.9031vj 1.79917.89T82 (

Combining

One combination method is to take a weighted sum of the component distributions

$$f = \sum \alpha_i f_i$$

Choose α_i that maximise (sum)Tj m

The Right Comparison

We do not want to compare the uncertainty in the average performance of two models.

We want the uncertainty in the comparative performance of the models to each other.

Bootstrap the difference: $\langle s_A - s_B \rangle_{BS}$

Not the difference of the bootstraps: $\langle s_A \rangle_{BS} - \langle s_B \rangle_{BS}$

Comparative Skill

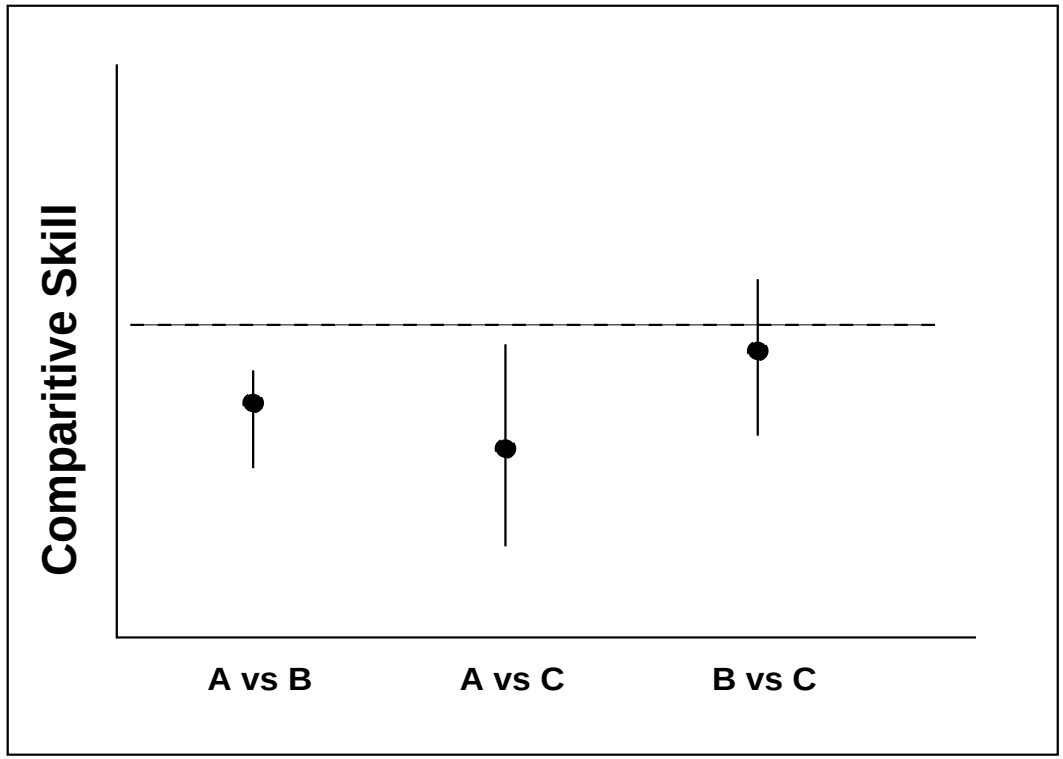
Motivation

Overview

Combining

Evaluation

Example



Example

Predicting temperature at Heathrow.

Using

- NCEP high resolution
- NCEP ensemble
- ECMWF high resolution
- ECMWF ensemble

Evaluating using Ignorance - out of sample

Example - one lead time

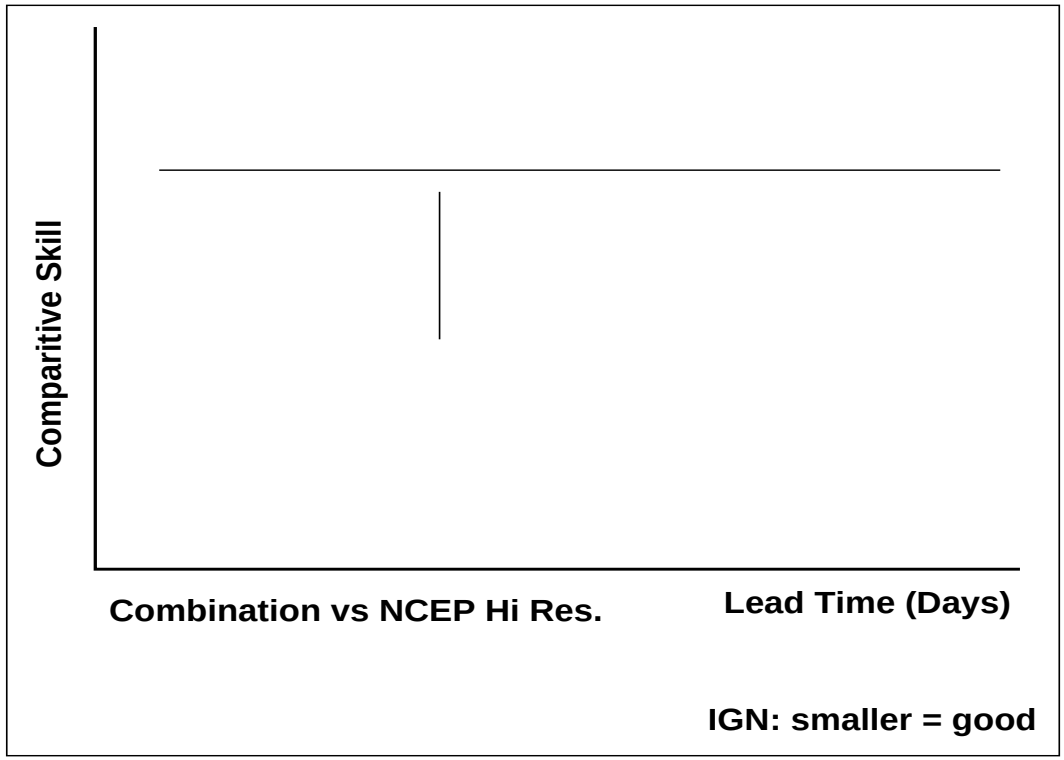
Motivation

Overview

Combining

Evaluation

Example



Example - one lead time

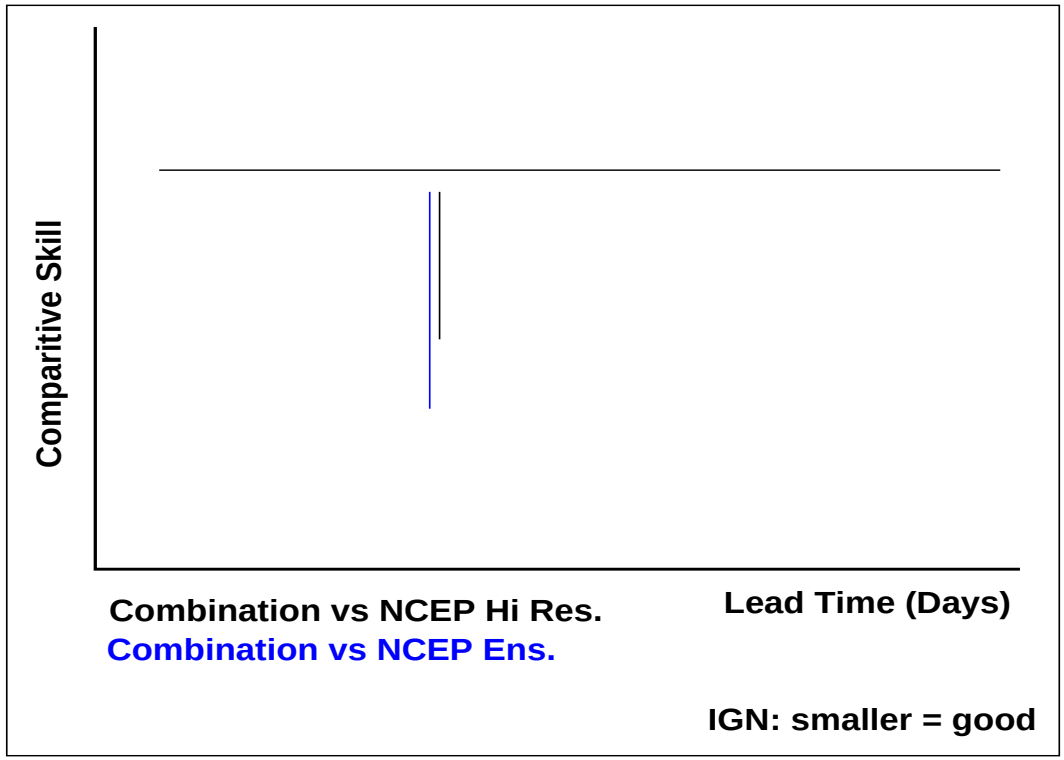
Motivation

Overview

Combining

Evaluation

Example



Example - one lead time

Motivation

Overview

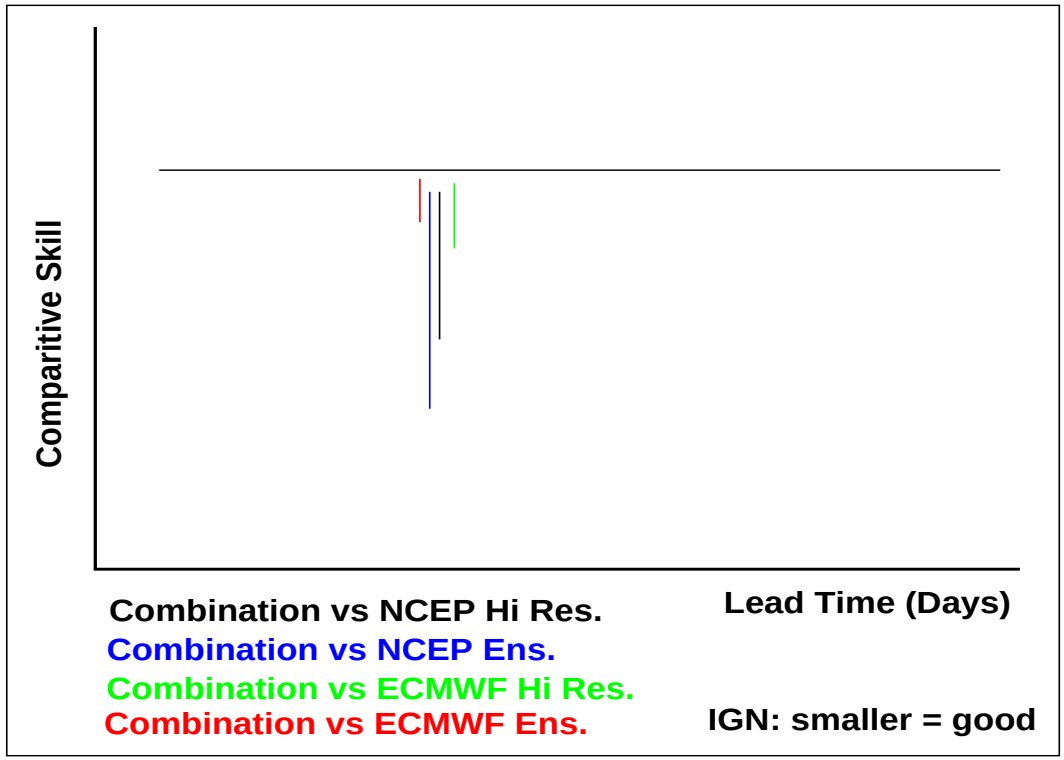
Combining

Evaluation

Example

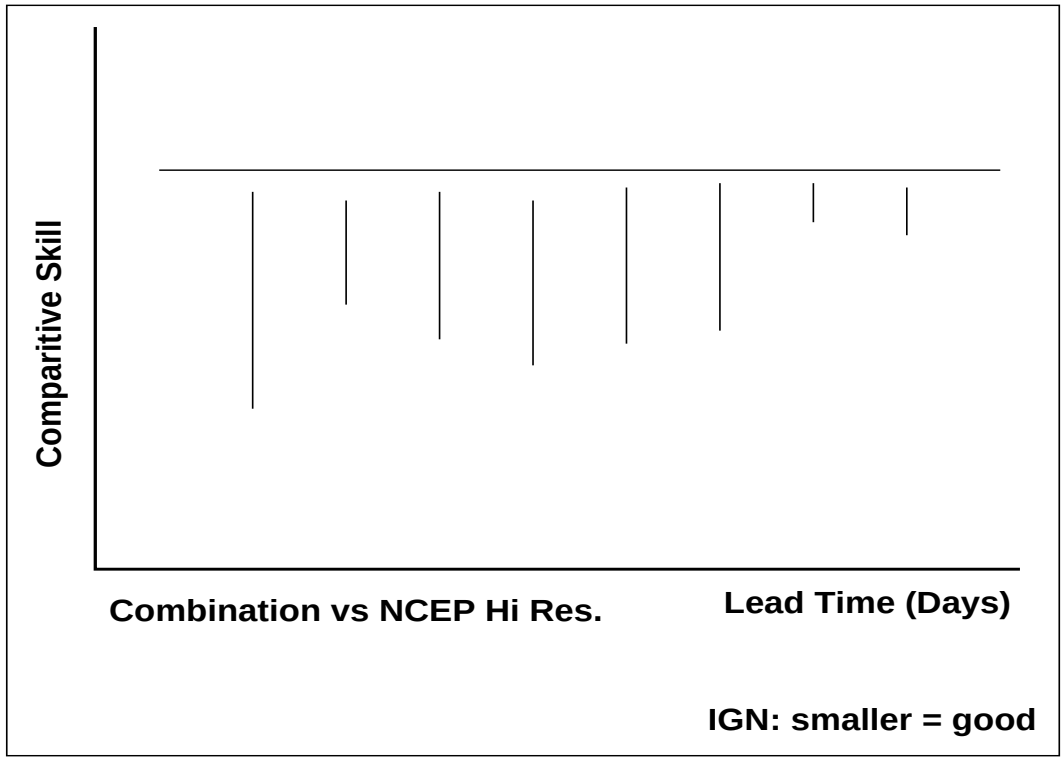
Example - one lead time

- Motivation
- Overview
- Combining
- Evaluation
- Example



Example - all lead times

- Motivation
- Overview
- Combining
- Evaluation
- Example



Example - all lead times

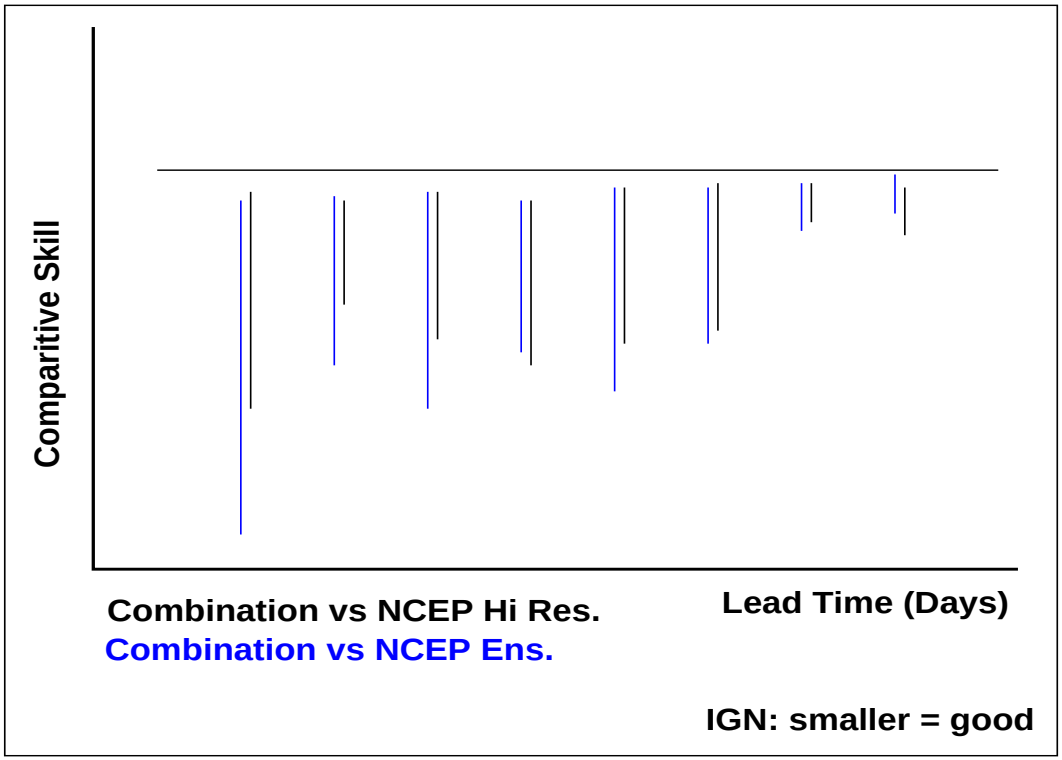
Motivation

Overview

Combining

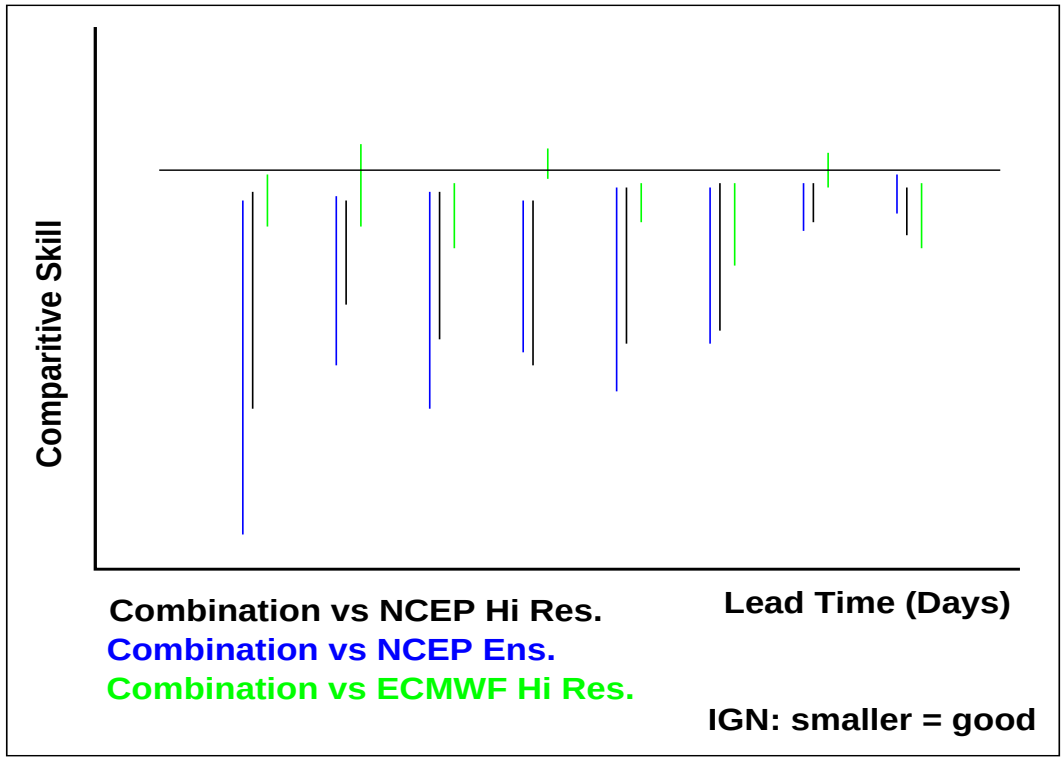
Evaluation

Example



Example - all lead times

- Motivation
- Overview
- Combining
- Evaluation
- Example



Example - all lead times

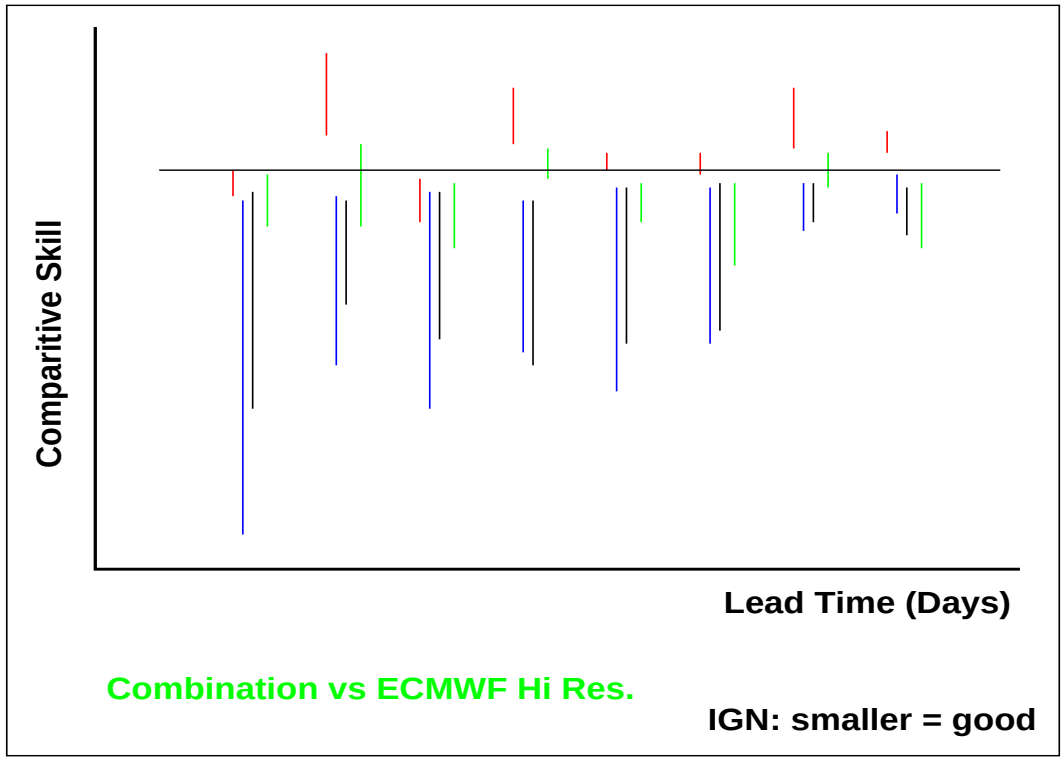
Motivation

Overview

Combining

Evaluation

Example



Summary

We have:

- presented a user-orientated methodology for combining simulations
- whatever combination method, the evaluation must be robust
- dressing method, combination method and size of forecast-verifi cation archive affects performance
- potential relevance to TIGGE, provides a framework for allowing users to extract the forecast information most relevant to them