

Application of different bootstrap techniques to estimate the statistical significance of systematic model differences



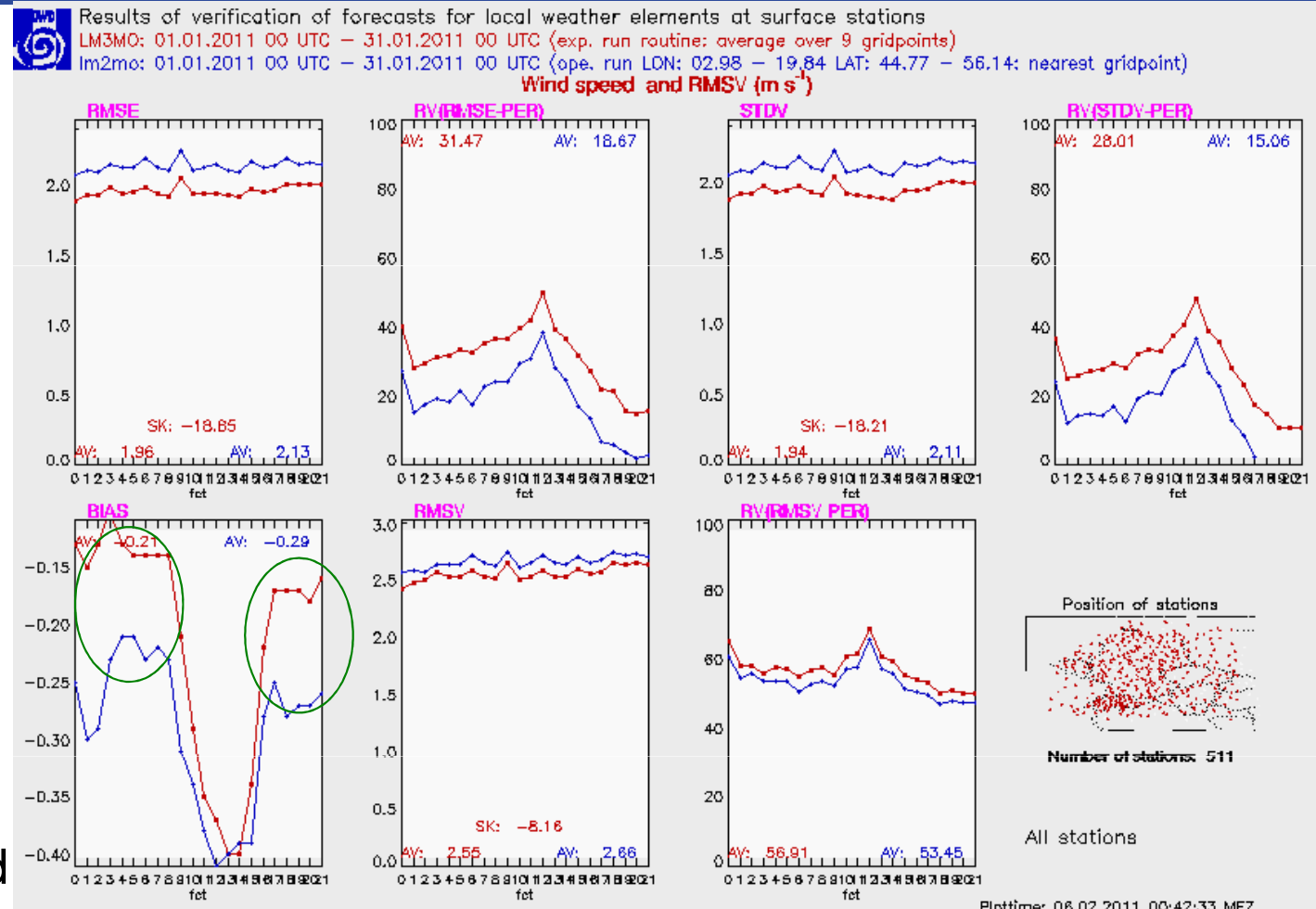
SA typical situation occurring during operational verification:

The problem:

- Are those differences between errors of different models or for different weather situations due to chance or are they statistically significant?

A possible solution:

- Looking at the samples using the bootstrap method



Problems while using bootstrap methods:

1. How many replications are necessary to get stable statistical results?
2. How should the sample data be grouped in order to avoid autocorrelation effect?

