

Purpose: Much research has been done on prediction of respiratory motion traces for motion compensation in radiotherapy. Unfortunately, the results of different groups cannot be compared easily due to different standards in pre-processing and analysis of the results. Furthermore, it has been speculated that the typically used measure for prediction quality, the RMS error, is not sufficient alone.

Methods: We propose a set of guidelines for signal pre-processing (i.e., for scaling, detrending, resampling, and denoising) as well as measures for the analysis of the prediction results. The latter complement the RMS error with confidence intervals, the signal's smoothness (called jitter) and a measure for the periodicity of the error (called frequency content). Additionally, we have developed an extendable cross-platform prediction toolkit for easy analysis of prediction algorithms.

Results: We found that very different signals (corrupted by noise, scaled by a constant factor, delayed in time, and scaled by random factors for each respiratory period) feature the exact same RMS error when compared to the original signal. The fundamental difference in the error signals can only be determined when using spectral measures, like the frequency content.

Conclusion: Using the guidelines developed, the proposed evaluation measures, as well as the publicly available prediction toolkit, should help the community in establishing a better understanding for the capabilities and shortcomings of individual prediction methods. Additionally, it should allow others to more readily compare newly developed methods to already published algorithms. In the future, it would be desirable to also create a database of motion traces from various sources. If these signals would represent the characteristics of motion traces observed in the clinic, it could serve as a general benchmark for the quality of algorithms for motion prediction.