The effect of selecting different training sets in the spectral and colorimetric reconstruction accuracy

Carlos Sáenz Gamasa, Begoña Hernández Salueña, Coro Alberdi Odriozola, Santiago Alfonso Ábrego and José Manuel Diñeiro Rubial

Departamento de Física. Universidad Pública de Navarra. Campus Arrosadía. 31006 Pamplona (SPAIN)

Abstract

The effect of selection of particular training sets to form the set of basis vectors for spectral and colorimetric reconstruction is analyzed. A number of training sets is randomly obtained from a spectral data base and the basis vectors obtained using Principal Component Analysis (PCA). The performance of each set of basis vectors is tested in the training and test sets in terms of Root Mean Square Error (RMSE) and CIE ΔE color difference. The effect of the number of vectors in the training set and the number of basis vectors is studied. Results show how particular sets of vectors are more suitable than others as training sets although this advantage can not be inferred from the RMSE values associated to them.

Introduction

Dimensionality reduction in spectral color science is a desired objective in many applications including spectral image databases, spectral capture devices and displays, spectral image encoding and others. Different approaches are possible but at the end, we must pay the price of some reduction in the accuracy in the reconstruction of the original data as a consequence of the reduction of the number of degrees of freedom to describe it.

Accuracy in spectral reconstruction is often associated with the need of accurate color reproduction. Both aspects are intimately related but have distinctive characteristics (consider metamerism for instance) that result in difficulties to obtain precise spectral and color fidelity between original and reconstructed data at the same time. In fact, it has been recognized the importance of the metrics used in the dimensionality reduction and several metrics have been tested trying to improve the results.[1, 2]

In a typical situation a spectral database, the 'Training Set', is used to obtain a reduced set of (metric dependent) basis vectors. If spectra in the training database is specified using nw wavelengths the number of basis vectors will be nb < nw to ensure the desired reduction in the dimensionality. The decision on how many basis vectors nb must be retained may depend on the required accuracy in the spectral or color reproduction or the need to have a physically realizable spectral system in order to capture or to display the data.

Since the selection of basis vectors is done in a particular data set, the Training Set, the performance of the reconstruction capabilities of the selected basis vectors is then tested in other spectral data sets, the 'Test Set'. Spectral and color reproduction in test sets are inevitably worse than in the training set. Several Spectral databases can be used for both training and test sets.[3]

This is an expected situation, since the basis vectors optimize the spectral and/or colorimetric reconstruction within the Training Set and there is no a priori reason guarantying the same optimal behavior within the Test Set. Although expected, this can not be considered a satisfying situation since in general we want a set of basis vectors capable to describe, to some degree of accuracy, all possible spectral (and color) data within the context of a particular application.

It is clear that the set of basis vectors depends on the metric but it is also biased by the particular data set used as Training Set. In this work we explore the influence of the selection of a particular Training Set in the accuracy of the spectral and colorimetric reconstruction of both the Training Set and a Test Set.

Experimental

A particular spectral database can be considered as a sample of the complete set of all existing spectra. Since this complete set is not available we have simulated the situation described in the introduction reducing the spectral space to a single database and obtaining Training Sets by random sampling subsets of the database.

We have used the spectral reflectance factors of the color patches in the Macbeth Color Checker DC chart (CCDC) to simulate the entire space of existing spectra. Reflectance factors were obtained with a Photo Research PR650 spectroradiometer in the range 380 – 780 nm at 4 nm intervals. Spectral vector are specified by 101 data points.

The CCDC database is split in two, the training and test sets. The training set consists in a predefined number nt of vectors randomly obtained from the CCDC database. All the vectors do not entering the training set form the test set. Then PCA is performed in the training set, keeping only a predefined number nb of basis vectors. The training and test sets are reconstructed using the nb basis vectors and mean RMSE and ΔE color difference between original and reconstructed spectra are computed.

The number of vectors nt in the training set has been varied from 10 to 80 and, for each one, the process has been repeated for a number of basis vectors nb (3 to 8). For each combination of nt and nb 1000 different training sets have been generated.

Results

Figure 1 and figure 2 show the reconstruction performance, mean RMSE and ΔE values over all training and test set combinations for a given nt and nb pair. As expected, for a given nb, mean RMSE and mean ΔE decrease as the number of basis vectors increase for both the training and test sets. In the case of the training set, for a given nb reconstruction accuracy decreases with nt. In the test set the opposite behavior is observed and RMSE and ΔE values improve (decrease) with nt. It is noticeable that for a given nt, the difference in mean RMSE values between the test and training sets are approximately constant regardless de number of basis vectors nb. On the other hand, ΔE differences between test and training sets decrease with nb.

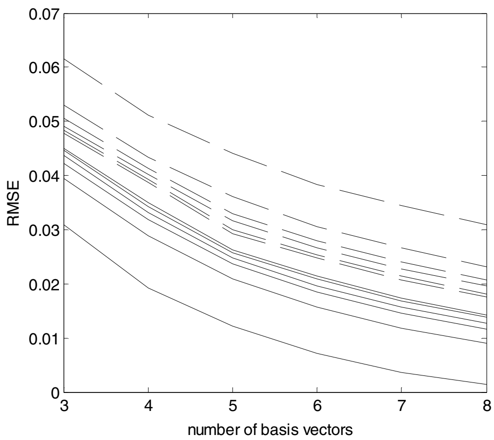

Figure 1. Mean RMSE values as a function of the number of basis vectors nb used in the reconstruction process. The Training Set (solid lines) with nt=10, 20, 30, 40, 60 and 80 (from bottom to top) and Test Set (dashed lines) at the same nt values (from top to bottom).

The mirror like behavior observed in figure 1 between training and test sets suggests a strong correlation between mean RMSE values that is confirmed in figure 3 where lines approach the diagonal representing equal mean RMSE values as the number nt of vectors in the training set increase.

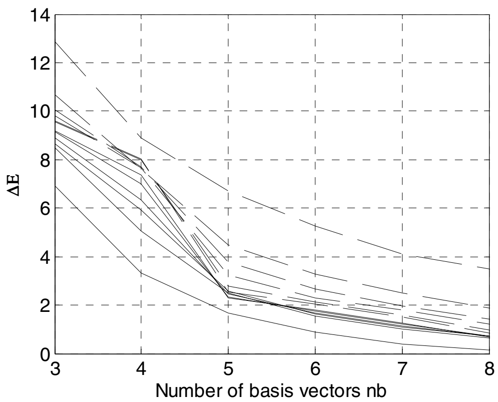

Figure 2. Mean ΔE values as a function of the number of basis vectors nb used in the reconstruction process. The Training Set (solid lines) with nt=10, 20, 30, 40, 60 and 80 (from bottom to top) and Test Set (dashed lines) at the same nt values (from top to bottom).

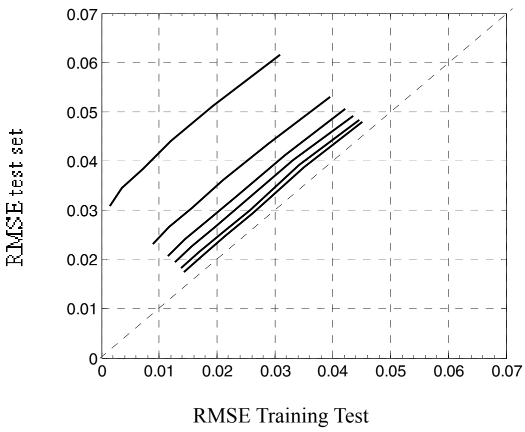

Figure 3. Mean RMSE in the Test Set versus Mean RMSE in Training Set. Each line represent a fixed nt value and varying nb values. Lines approach the diagonal as nt increases. Values of nt and nb the same as figures 1 and 2..

These results refer to mean RMSE and ΔE, averaging over the entire random sampling performed for each nt and nb pair. However for fixed nt and nb reconstruction accuracy strongly depends on the particular set of vectors used to derive the basis vectors.

As an example, we show in figure 4 the RMSE on the test sets versus the RMSE value in the training set (similar figures are obtained for other nt and nb values). In general there exists a slightly negative correlation between them but there is a considerable amount of dispersion in both axes. The particular selection of a training set has great influence on the RMSE on that set and, what is more important in our discussion, does not guarantee good results in the RMSE values for the test set. In fact, training sets that have similar RMSE values may produce very different RMSE values in the test set. Minimizing RMSE in the test set does not imply good performance in the test set.

A similar conclusion can be drawn from ΔE data in figures 5 and 6. Training sets that exhibit small RMSE values can produce ΔE values in either the training or the test sets (or both) which are higher that those obtained by training sets possessing ‘a priori’ worse RMSE values. It can be even possible to find one training set that performs better that another even with less vectors (nt) and less basis vectors (nb).

These results point out the importance of what vectors constitute the training set. Their selection can not be made on terms of RMSE alone. Intrinsic properties (multicolinearity, correlation..) may play an important role in improving the performance of databases as training sets when testing different spectral or colorimetric metrics.

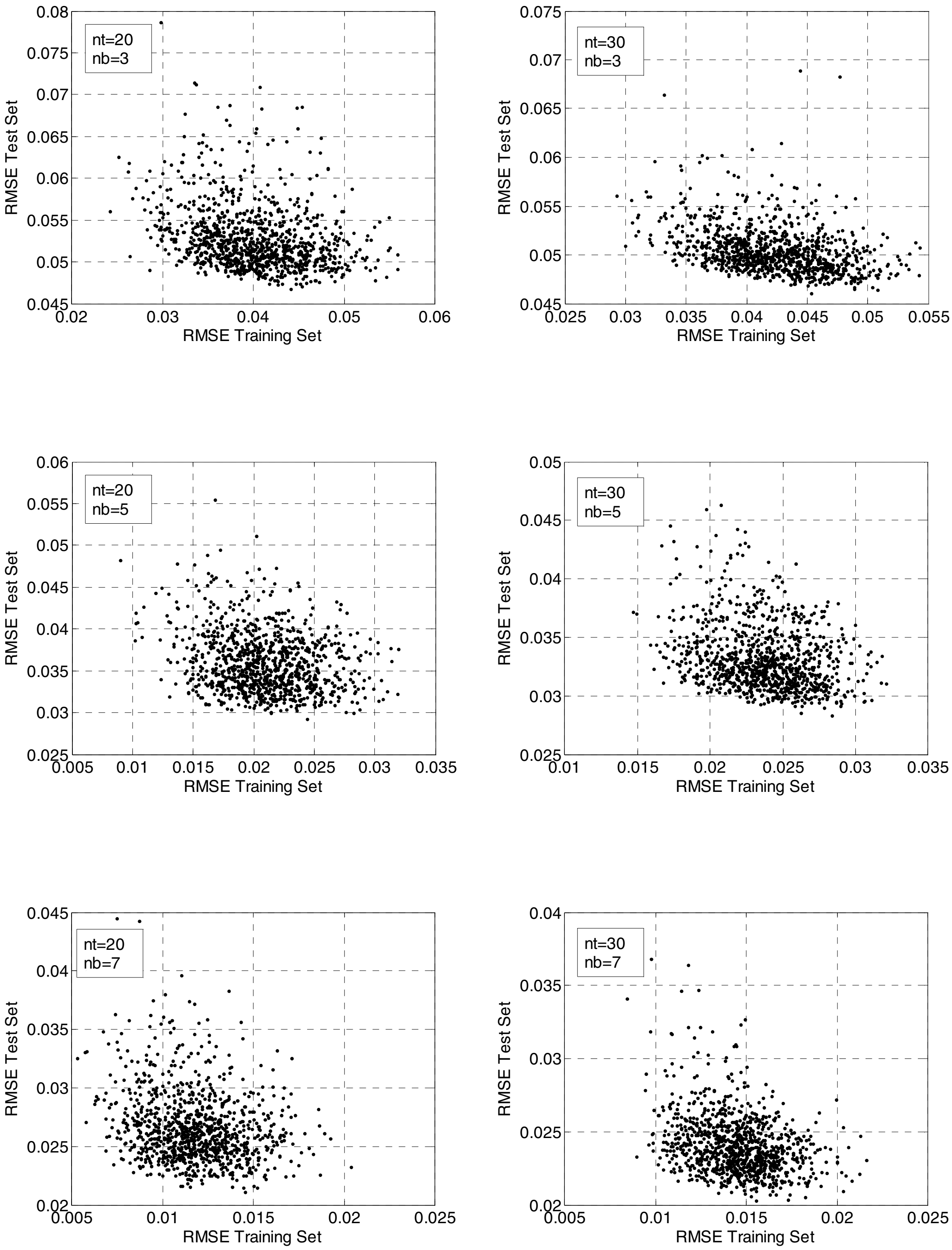

Figure 4. RMSE in the test set versus RMSE in the training set for nt=20 (left) and nt=30 (right) in the case of nb=3,5 and 7 basis vectors.

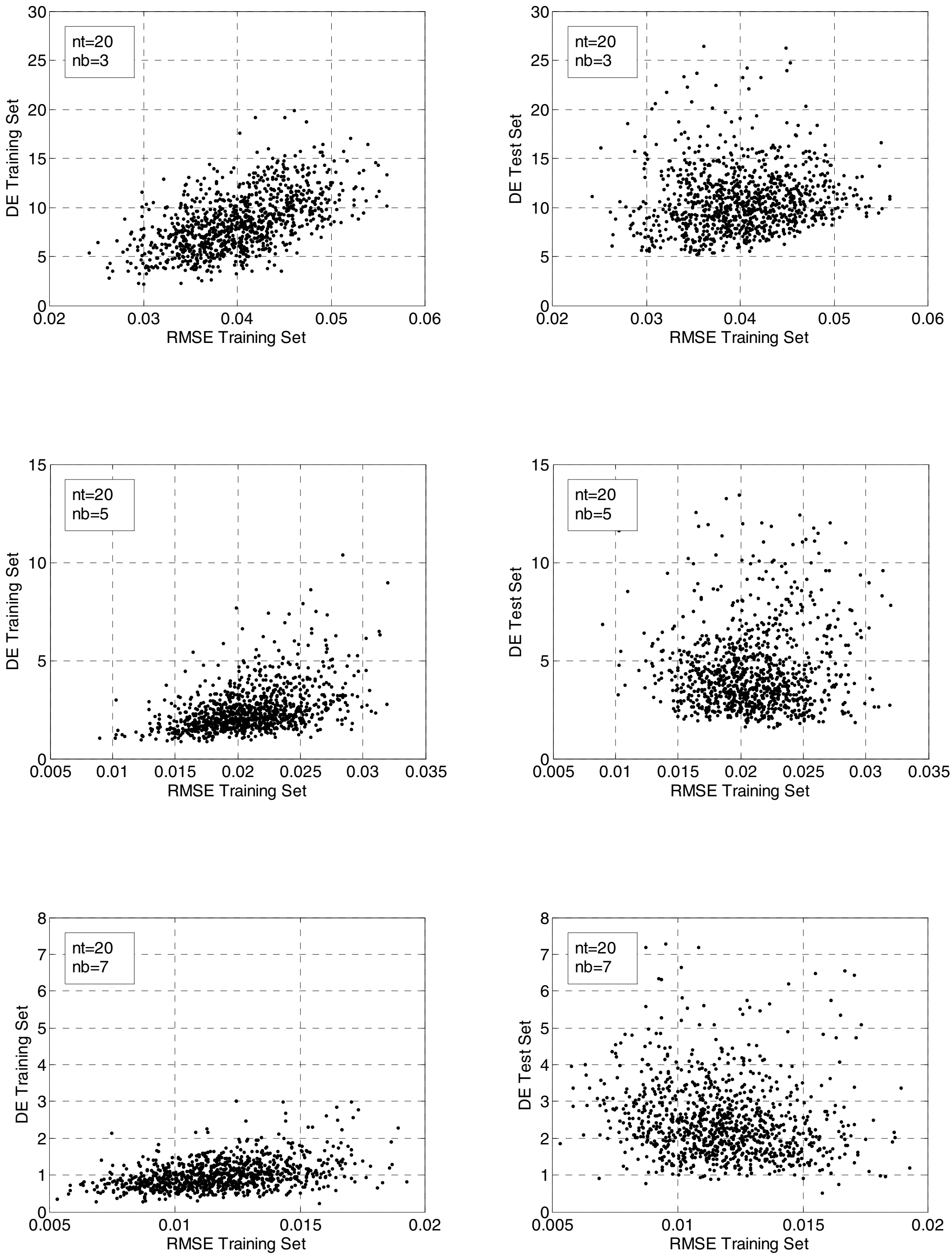

Figure 5. ΔE in training set (left) and test set (right) as a function of RMSE in the training set for nt=20 and nb=3, 5 and 7

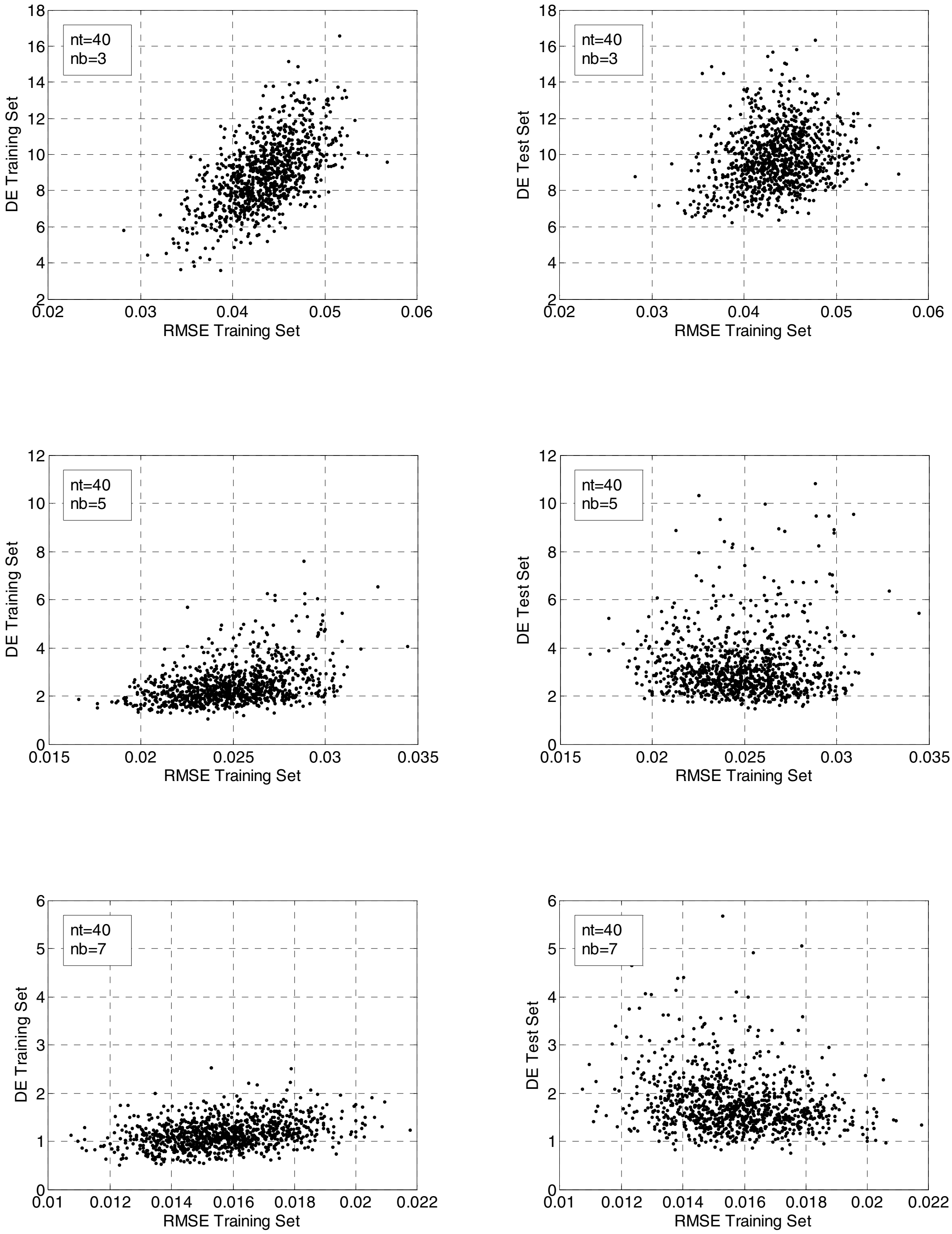

Figure 6. ΔE in training set (left) and test set (right) as a function of RMSE in the training set for nt=40 and nb=3, 5 and 7

Conclusion

The analysis corroborates that RMSE alone is not a good indicator of the performance of a particular training set in reconstructing spectral and colorimetric data on a test set (and colorimetric data in the training set itself).

Which vectors are used in the training set is of grate importance, since better results can be obtained sometimes with fewer vectors in the training set. The results indicate that intrinsic (spectral) properties exist that make one set of vectors more suitable than another set to serve as training set. Knowing these properties and making a pre selection of vectors may improve the performance of a particular spectral database.

Results have been obtained for a particular metric. As mentioned in the introduction, different metrics can be used, maybe depending on the final application, but if the observed behavior relays on the selection of vectors in the training set, as it seems to be, similar results should be obtained.

It could be argued that these results depend on the particular selection of the CCDC as a database to represent the spectral population. We are planning to repeat this analysis with other databases having more spectra. We do not think however that the results will differ significantly from a qualitative point of view although quantitative differences may be of practical importance.

References

1. Francisco H. Imai, Mitchell R. Rosen and Roy S. Berns. Comparative Study of Metrics for Spectral Match Quality. Proceedings of te First European conference on colour graphics, imaging and vision (CGIV), 492-496 (2002)
2. J. A. Stephen Viggiano. Metrics for Evaluating Spectral Matches: A Quantitative Comparison. Proceedings of CGIV-2004: The second European conference on colour graphics, imaging and vision, 286-291 (2004)
3. Oili Kohonen, Jussi Parkkinen, TImo Jääskeläinen. Databases for Spectral Color Science. Color Research and Application 31 5, 381-388 (2006)