

Since information theory plays an important role in this book, I sent it to P. Vitányi, co-author of an excellent book on Kolmogorov Complexity. (M. Li & P. Vitányi: An Introduction to Kolmogorov Complexity and Its Applications. Springer: Texts in Computer Science, 3rd ed. 2009.)

He wrote a nice e-mail back. Being an expert in the field, he also gave some pointers to the literature which I think are worth mentioning:

It struck me that you mentioned
P. Gacs, J. Tromp, P. Vitanyi, Algorithmic statistics, IEEE Trans. Inform. Theory, 47:6(2001), 2443-2463.

However, you didn't mention
N.K. Vereshchagin and P.M.B. Vitanyi, Kolmogorov's Structure functions and model selection, IEEE Trans. Inform. Theory, 50:12(2004), 3265-3290

although this is far more suited to your interest, nor its (extreme) generalization

N.K. Vereshchagin and P.M.B. Vitanyi, Rate distortion and denoising of individual data using Kolmogorov complexity, IEEE Trans. Information Theory, 56:7(2010), 3438-3454.

Also interesting for you may be:
P. Adriaans, P.M.B. Vitanyi, Approximation of the two-part MDL code, IEEE Trans. Inform. Theory, 55:1(2009), 444-457.

P.M.B. Vitanyi and M. Li, Minimum Description Length Induction, Bayesianism, and Kolmogorov Complexity, IEEE Trans. Inform. Theory, IT-46:2(2000), 446-464.

As well as:

R. Cilibrasi, P.M.B. Vitanyi, Clustering by compression, IEEE Trans. Inform. Theory, 51:12(2005), 1523-1545.

etc.

In short, most of the papers in <http://homepages.cwi.nl/~paulv/learning.html> where they can also be viewed or downloaded.