Some mathematical results on percent model affinity (PMA) index

Johanna Ärje[1], **Kwok-Pui Choi**[2], **Fabio Divino**[3] and **Salme Kärkkäinen**[1]

[1] University of Jyväskylä, Finland

[2] National University of Singapore, Republic of Singapore

[3] University of Molise, Italy

Percent Model Affinity index (Novak & Bode 1992) is used in biomonitoring to assess the ecological state of rivers and lakes based on benthic macroinvertebrate samples. In PMA index, the taxa proportions found in a sample are compared to those of a reference population and any divergence from the reference is considered harmful. The PMA index can be defined as

$$PMA = 1 - \frac{1}{2}\sum_{h=1}^{C} |p_h - q_h|, \tag{1}$$

where p_h is the unknown proportion of taxonomical group h, $h = 1, \ldots, C$, in the population and q_h the unknown proportion of taxonomical group h in the reference population. In practise, both parameter sets are estimated: the first one from the sample in question and the latter one from samples from a reference population. The formula above is similar to several indices, such as the percentage similarity (Renkonen 1938) and Bray-Curtis similarity (Bray & Curtis 1957) with the key difference that PMA index is used to compare a sample to a reference population, not to another sample.

The properties of the index have not been thoroughly studied when q's are estimated. Assuming the multinomial distribution, Smith (1982) calculated the expected value and variance of PMA when the reference parameters $\mathbf{q}$ are known. For the case when $\mathbf{q}$ is estimated, Smith & Zaret (1982) calculated the expected value when assuming normal distribution and sample being from the reference population. Further developments are needed in order to study which factors affect the values of the index and thus the decision making of environmental administration.

We present the expected value and variance of the PMA index in several set-up's. We assume that the samples follow multinomial distribution, $\mathbf{X} \sim MN(n; \mathbf{p})$. For comparing the results in Smith (1982), we show our results for the expected value and variance of PMA when the reference parameters $\mathbf{q}$ are known. Further, we consider the case when the parameters of the reference population are also estimated from a sample $\mathbf{Y} \sim MN(m; \mathbf{q})$ and show the theoretical and/or simulation results. Third, we study the case, when parameters $\mathbf{p}$ and $\mathbf{q}$ are estimated from several samples.

Keywords: biomonitoring, expectation, multinomial distribution, PMA index, variance

References:

Bray, J.R. and Curtis, J.T. (1957). An ordination of the upland forest communities of Southern Wisconsin. *Ecological Monographs 27(4)*, 325-349.

Novak, M.A. and Bode, R.W. (1992). Percent model affinity: a new measure of macroinvertebrate community composition. *Journal of North American Benthological Society 11(1)*, 80-85.

Renkonen, O. (1938). Statisch-kologische Untersuchungen ber die terrestrische Kferwelt der finnischen Bruchmoore. *Ann. Zool. Soc. Bot. Fenn. Vanamo 6*, 1-231.

Smith, E.P. (1982). Niche breadth, resource availability and inference. *Ecology 63(6)*, 1675-1681.

Smith, E.P. and Zaret, T.M. (1982). Bias in estimating niche overlap. *Ecology 63(5)*, 1248-1253.