We are grateful to András Bardossy for his patience and in-depth assessment of our manuscript. We formulated our previous response with great care. We take this opportunity to clarify our response with even greater precision based on this commentary.

Comment: In my opinion the Markov inequality is not a stronger inequality than the Chebysev – it is based on the same very simple …

Response: Thank you for the opportunity to further clarify our response. Our reference to Pande et al (2012) (in the response that the referee is referring to) used Hoeffding's inequality which is an extension of Chebysev (hence referring to Hoeffding as Chebysev like). The authors used it for a class of linear reservoir models. Further, Chebysev is an application of Markov's inequality. It was for these reasons that we suggested that the bounds used in Pande et al (2012) were deemed looser than Markov's inequality (atleast they cannot be tighter than Chebysev).

We agree that Markov and Chebysev are based on the same principle since the latter is an application of the former. The referee is correct in suggesting that we have used Chebysev inequality in our current algorithm (please see Arkesteijn and Pande, 2013 for further details). But it is used as an *upper bound* on a possibly tighter probability bound. Arkesteijn and Pande (2013) used Chebysev inequality to suggest that the probability of error is no larger than $O(N^2)$. Hence they argued that a tighter probability bound may be discovered from a class of functions quadratic in N. The Algorithm 2 estimates the parameters of an appropriately tight bound based on the data that the weather resampler generates.

We would be happy to provide further details on why we think the bounds computed by Algorithm 2 are tighter than Chebysev's. It would be linked to the complexity measure that the bounds (estimated by Algorithm 2) contain. Such arguments have also been provided in Arkesteijn and Pande (2013).

Nonetheless, we agree with the referee that tighter bounds are highly desirable while acknowledging that these bounds are independent of any assumption on the underlying probability distribution. In our opinion, the latter is a property that is equally desirable.

We will incorporate the above discussion in our revised manuscript.

Comment: Unfortunately your arguments are not fully precise…

Response: We appreciate the patience of the referee. We take this opportunity to further clarify the doubts that have been raised in this comment.

The indices appear to be consistent with the arguments presented, except $|E\xi_1 - \xi_1| + |E\xi_1 - \xi_1| \leq t$. It should be $|E\xi_1 - \xi_1| + |E\xi_2 - \xi_2| \leq 2t$. The referee is correct in pointing out the inaccuracy of our indices and a missing multiplier, though it should not be a strict inequality. We have now corrected the affected terms that appeared around and after this inequality in our previous response. The following presents the updated set of arguments (by substituting t by 2t). As this referee would agree, the arguments and the final proposition, nonetheless, remain the same and stand valid.

"In other words one can say that $|E\xi_1 - \xi_1| + |E\xi_2 - \xi_2| \le 2t$ *with a probability of atleast 1-*$[\delta_1(t) + \delta_2(t)]$*. Finally we know* $E\xi_1 - \xi_1 - (E\xi_2 - \xi_2) \le |E\xi_1 - \xi_1| + |E\xi_1 - \xi_1|$*.*
From the above sequence of valid arguments, we note that the following holds with a probability of atleast 1-$[\delta_1(t) + \delta_2(t)]$ *that*
$E\xi_1 - E\xi_2 \le (\xi_1 - \xi_2) + 2t$*.*
We can now obtain a valid answer to the question of which model is better in the sense of its (expected) risk. What is the probability that model1 is better than model 2 in expected sense, i.e. what is the probability that $E\xi_1 - E\xi_2 \le 0$*?*
Answer: with a probability of atleast 1-$[\delta_1(\frac{\xi_2-\xi_1}{2}) + \delta_2(\frac{\xi_2-\xi_1}{2})]$*. We obtain this by letting* $(\xi_1 - \xi_2) + 2t = 0$*. Thus better models can be truthfully revealed with a confidence level that is resolved by the upper limit on risk."*

Indeed the bounds on *1*-$[\delta_1(\frac{\xi_2-\xi_1}{2}) + \delta_2(\frac{\xi_2-\xi_1}{2})]$ are artificially respected (that it they lie between 0 and 1) in case $\frac{\xi_2-\xi_1}{2}$ are such that lead *1*-$[\delta_1(\frac{\xi_2-\xi_1}{2}) + \delta_2(\frac{\xi_2-\xi_1}{2})]$ to go out of bounds.

Arkesteijn and Pande (2013) provide an in-depth interpretation of such an upper bound. For e.g. $\delta_i(t)$ is a deterministic function of t. The above inequalities have been simplified compared to those provided in the paper and in Arkesteijn and Pande (2013). But the essence of the arguments remains the same. In Arkesteijn and Pande (2013) and this paper, the upper bound on the probability is estimated from Algorithm 2 (please see also our response to the first comment above) and using equation 2 of the paper. Hoeffding bounds can as well be used but as Pande et al (2012) have suggested, these bounds tend to be weak.

Let us assume that we have estimated functions $\delta_1(.)$ and $\delta_2(.)$. For a given data set, we have empirical errors for both the models ξ_2 and ξ_1. By plugging these in 1-$[\delta_1(\frac{\xi_2-\xi_1}{2}) + \delta_2(\frac{\xi_2-\xi_1}{2})]$, we obtain the least probability with which model 1 is better than model 2 in expected sense. In context of the current paper, the same can be said by using inequality (2) of the paper. Then $\delta_i(t)$ would be a function of type $F(h,N)\eta^2/t^2$ if we let t to represent γ in inequality 2 of the paper, where $\eta > 0$, $F(h,N)$ is a function of N defined in the paper and the parameters of the quadratic functions represent the corresponding model complexity.

We will add the above set of arguments in the discussion section of the revised manuscript.