Reply to Referee #1

We would like to thank the reviewer for your constructive comments to the manuscript. Here are our explanations for the unclear points that the reviewer concerned. But at first, we would like to reproduce the reviewer's comments in a black font:

The paper deals with a novel interpretation of the NSE score measure starting from the observation that it is mostly used in hydrology and poorly exploited in other sciences.

This interpretation is based on a signal processing viewpoint.

While the paper is interesting and useful the following concerns are at the hand.

The NSE interpretation provided is based on a model error for forecast which is given in eq 5 and is basically driven by a Gaussian random error being it equal to noise in the signal processing viewpoint. This basic foundation provides by itself a lack of generality with respect to application of NSE to other sciences included hydrology. In hydrology the NSE is intended as a model performance metric where the difference between model and observation is not limited to noise. Differences between model and observation but also differences between observation and reality and also differences between different models can be analyzed by means of NSE or KGE whose meaning is quite clear and leaves no doubts in my personal opinion.

In a hydrologic model, but also in other earth sciences, different models may arise because different processes are modelled in a stochastic or determinist way and/or because some processes are described or neglected following the fact that they can be more or less important according to the time-space scale of application and modeling purpose. Hence the difference between model output and observations may be very different from what given in eq 5, It can be deterministic or stochastic, and affected by deterministic or stochastic (or both) variability.

As a consequence, it seems that the proposed analysis, while interesting and well founded in the context of the signal processing field (or any other fields where only noise provides difference between model output and observation). In the same light one may not accept the "general case" version of NSE which is obtained by considering the multiplicative error, beside the additive error, defined in eqs (32). Even in this case the "general case" should be addressed as relative to which field of application, besides the field of signal processes or affine methods.

I believe the authors should strongly address this issue in a revised version of the manuscript.

Reply: First of all, we totally agree with the reviewer on the point that the additive error model that we consider in Section 2 is relatively limited and cannot cover all complicated relationships between models and observations in reality. Our purpose when using this additive error model is to show that the scientific meaning of NSE is truly revealed under this error model, which turns out to be the well-known signal-to-noise ratio in signal processing.

Due to the limit of the additive error model, in Section 3 we consider more general cases which can be described by joint probability distributions of model simulations and observations. Although the reviewer mentioned that general cases have not been dealt, we believe that our mixed multiplicative-additive error model addressed a wide range of cases of model-observation fitting. We would like to emphasize here that any relationship between model simulations and observations can always be analyzed through their joint probability distributions. Then we assume that such distributions are bi-variate normal distributions, which is usually observed in reality. Again, we would like to emphasize that we do not assume the mixed multiplicative-additive error model for the general cases. Rather than that, this multiplicative-additive error model is a consequence of the bi-variate normal distribution of model simulations and observations.

Of course, if this joint distribution is not Gaussian, we cannot expect to see the mixed multiplicative-additive error model and the relationship between model simulations and observations may be even more complicated as the reviewer commented. However, we believe that there are some suitable transformations to make observations and forecasts follow Gaussian distribution in most cases including hydrological applications. The generality of our error model described above has already been discussed in the current manuscript:
Since all information on forecasts and observations is encapsulated in their joint probability distribution, we can seek the general form of this conditional distribution from their joint distribution in the general cases. For this purpose, we will assume that this joint probability distribution is a bivariate normal distribution. If the joint distribution is not Gaussian, we need to apply some suitable transformations to f, o such as the root squared transformation $(f, o) \rightarrow (\sqrt{f}, \sqrt{o})$, the log transformation $(f, o) \rightarrow (\log(f), \log(o))$, the inverse transformation $(f, o) \rightarrow (1/f, 1/o)$, … (Pushpalatha et al., 2012).

We have shared the same opinion with the reviewer that the meanings of NSE and KGE "is quite clear and leaves no doubts" until we have come up with a counterexample based on the mixed multiplicative-additive error model. Let us consider a model simulation with a random error

$s_1 = o + \varepsilon,$ (1)

where we assume $\mu_{s1} = \mu_o = 0$, and $\sigma_o = \sigma_e$. This simulation in deed gives us the boundary between "good" and "bad" simulations from the viewpoint of NSE as we have examined in Section 2. Then it is easy to calculate its NSE and KGE: $NSE_1 = 0, KGE_1 \approx 0.5$. It is very clear that we cannot improve this simulation since the power of random noise is equal to the power of observations. But this is not true if we measure model performance by NSE or KGE. Constructing a new simulation which is half of s_1

$s_2 = 0.5 s_1 = 0.5 o + 0.5 \varepsilon,$ (2)

and calculating its NSE and KGE we obtain

$$NSE_2 = 1 - \frac{\overline{(o-s_2)^2}}{\sigma_o^2} = 1 - \frac{\overline{(0.5o-0.5\varepsilon)^2}}{\sigma_o^2} = 1 - \frac{0.5\sigma_o^2}{\sigma_o^2} = 0.5, \qquad (3)$$

$$KGE_2 = 1 - \sqrt{(\rho-1)^2 + (\sigma_2/\sigma_o - 1)^2} = 1 - \sqrt{\left(1/\sqrt{2}-1\right)^2 + \left(1/\sqrt{2}-1\right)^2} = 2 - \sqrt{2} \approx 0.6. \qquad (4)$$

Suddenly, both NSE and KGE indicate that s_2 is better than s_1 considerably, although all we do is just halving s_1. However, Eq. (2) in nature is equivalent to Eq. (1), and we should not improve any simulation by just scaling the observations and the random error.

We have tried to find a remedy for this problem in Section 3 when we explore extended versions of NSE and KGE in the general cases. If the extension is applied, we will see that $NSE_2 = 0, KGE_2 \approx 0.5$, which show consistency with the evaluation of s_1. Our motivation to build the new scores is somehow unclear in the original version of the paper. Therefore, in the revised version of the paper, we intend to add this example into the revised manuscript to make our approach more rigorous.

Sincerely,
On behalf of the authors,
Le Duc.