Hydrol. Earth Syst. Sci. Discuss., 12, C6931–C6948, 2016
www.hydrol-earth-syst-sci-discuss.net/12/C6931/2016/

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Hierarchy of climate and hydrological uncertainties in transient low flow projections" *by* J.-P. Vidal et al.

J.-P. Vidal et al.

jean-philippe.vidal@irstea.fr

Received and published: 4 March 2016

The referee comments are recalled in italics and followed by the authors' responses.

The article presents a transient decomposition of uncertainties in low flow changes in two Alpine catchments. The decomposition is done for 30-year as well as for yearly statistics. As a method for the decomposition, the quasi-ergodic ANOVA method proposed by Hingray and Saïd (2014) is used. It is shown that in the ensemble mean, the low flows generally decrease. The largest fraction of uncertainty comes from internal variability. Hydrological models also contribute substantially to the total uncertainty, which is discussed to be due to differences in snow and evapotranspiration routines between the different hydrological models. Also, a comparison to a standard

ANOVA method is presented. It shows that the quasi-ergodic ANOVA method results in a smoother transient uncertainty decomposition than the standard ANOVA.

General comments

The discussion paper studies a relevant topic and applies state-of-the art methods to look at low-flows in a climate-impact study. It is well written and has a high scientific quality including a sound literature discussion. Also, the mathematical details of the applied method are given in a conclusive way. I recommend publication after my comments below have been taken into account.

The authors would like to thank the referee for this positive evaluation of the manuscript and for the insightful comments on the QE-ANOVA method.

Major comments

The paper very much relies on the statistical method of the quasi-ergodic ANOVA introduced by Hingray and Saïd (2014) and consequently, other aspects in the impact modeling chain are less well detailed. I do not mind that and in fact would like to see even more discussion of the QE-ANOVA. There are many assumptions made in the QE-ANOVA and some of them could be verified. For example, the stationarity of the variance has not been proved, something which is even more relevant since the authors use yearly anomalies with a higher degree of variability. If there is a considerable degree of non-stationarity in the variability, I would like this to be included in the discussion of the QE-ANOVA results.

Assuming a constant coefficient of variation of the variable studied with respect to the inter-realization dispersion (for SSIV) and with respect to the inter-run dispersion (for LSIV) is indeed a central hypothesis in the QE-ANOVA method. We will try to avoid here using the term of stationarity which is subject to much discussion in the recent literature.

It is indeed possible to relax this hypothesis and compute yearly empirical values of the variance terms in Equation (A1) for SSIV. In Equation (A1), the empirical variance Vark[Y(m,r,k,t)/ŷ(m,t)] may be calculated for any (m,r) at any time horizon t of the simulation period. SSIV(t) may then be calculated without the quasi-ergodic assumption. Figure 1 below compares the temporal evolution of SSIV with the quasi-ergodic (QE) assumption – as in the manuscript – and without it, as defined above. It shows that the hypothesis is quite reasonable and allows removing some noise without altering the overall temporal evolution.

Similarly, it is also possible to relax this hypothesis for LSIV, even if in a degraded mode because of the different numbers of runs from each GCM, and because of the fact that 2 of them (out of 4) only have one run. In equation (A3), the temporal evolution of Varr[Y(m,r,•,t)/ŷ(m,t)] may therefore be calculated for any m where r>1 (i.e. for model chains that include either IPCM4 and ECHAM5). Figure 2 below shows here again that the quasi-ergodic hypothesis is quite reasonable. The above comments will be included in the discussion section and figures provided as a supplementary material or in annex.

Over all, I would also like to see a more critical discussion of the results, not only highlighting the advantages but also the limitations of the QE-ANOVA results. It has to be clear for a Non-ANOVA specialist what they can expect from the method, since many impact modelers would probably like to use the QE-ANOVA approach. For example, the QE-ANOVA approach would not be suitable to study changes of extreme precipitation for which other studies have shown that the variability can increase even in case of decreasing mean.

In addition to the response to the above comment, the revised manuscript will also include a dedicated subsection of the discussion on the advantages and limitations of the QE-ANOVA method, including the central issue of extracting the signal with a relevant shape. This issue necessarily leads to overestimate LSIV (see Hingray and Saïd, 2014). As a complement to the response of the previous comment, the hypothesis of a constant coefficient of variation over the whole period may actually be thus relaxed by

using only a limited number of time steps around the target time horizon for calculating LSIV. In such a local QE-ANOVA approach, the estimation can next be applied in turn for each target time horizon, allowing the LSIV to depend on the target time. This approach leads to quite interesting results for synthetic data with various response to uncertainty ratios, as shown by Hingray et al. (submitted).

Also, although the literature review is generally good, it would be good to include a part about other ANOVA methods. In particular, the study by Northrop and Chandler (2014) could be cited to refer to another method that is able to deal with an unbalanced design.

We agree that a number of other ANOVA methods have been presented and applied in the recent years. The main focus on our work was however not to present a new ANOVA method but rather to explore the hierarchy of uncertainty sources in a specific hydrological issue using (and adapting in a way) an already existing ANOVA method. Of course the results of the uncertainty analysis may significantly depend on the method. We recently explored this issue from synthetic experiments (Hingray et al., submitted). We will acknowledge this point in the discussion and mention the possibility / interest to apply other methods such as the one of Northrop and Chandler (2014) pointed out here, which seems indeed interesting to explore especially the robustness of uncertainty estimates.

I have tried to give as detailed comments as possible below. I am looking forward to the author's response and would also be happy to discuss certain aspects if necessary.

Detailed comments

- *Title: The term "Hierarchy" is a bit misleading, as there is no dominating hierarchy but the contributions of the different uncertainty sources are changing over time. To me, hierarchy is something structurally inherent. Also, the term might lead to confusions with the use of hierarchical ANOVA models, which are not used in this* *study.*

We would prefer to keep the current title as the work indeed attempts at finding the hierarchy of climate and hydrological uncertainties in hydrological projections. The fact that such a hierarchy depends on the target time horizon will be specified in the revised abstract and conclusions in order to remove any ambiguity. Moreover, we do not think confusions may arise with hierarchical ANOVA models as we never use the term in the manuscript (abstract included).

- *Section 2.2.1: Which variables were used from the GCMs?*

  The GCM variables used as statistical downscaling predictors depend on the SDM. More information on predictors can be found in Lafaysse et al. (2014), but SDM versions used here slightly differ. A complete description of versions used here may be found in Hingray et al. (2013, p. 24). With their notations, the versions used here are: analog20, d2gen22 and dsclim11. In short, all 3 SDMs use some sort of geopotential fields (either sea level pressure or geopotential height at 700 or 1000 hPa), and for d2gen and dsclim, some large-scale indicator of temperature (either at the surface or at 700 hPa), and additional predictors like for example humidity (relative, specific or flux at 700hPa) or geostrophic wind components at 700 hPa.

- *And what do the different runs in Tab. 1 stand for? Of course, the introduction gives some hints, but it should be clearly stated in this section, too.*

  The different runs correspond to different free simulations of a GCM differing only by their initial conditions (here in 1850), and thus provide an estimate of the GCM internal variability. This will be reminded in this section.

- *Also, has the data since the end of ENSEMBLES been published publicly? If so, please indicate the data source.*

  ENSEMBLES data are publicly available through the project website http://ensembles-eu.metoffice.com. This data source will be added to the revised manuscript.

- *Page 12656, lines 18-29: It is unclear which parameters that are used for the subsampling. Was it changes in mean annual temperature and precipitation or anything else? Please specify.*

  The conditioning variables used for the subsampling have indeed not been specified in the manuscript. They are changes between 2 periods (1980-2009 and 2036-2065), on summer and winter precipitation and temperature, and on inter-annual variability of annual precipitation and temperature, all of them on basin-average (whole Durance basin) variables. They have been carefully chosen based on their relevance for water management. This will be added to the revised manuscript.

- *Also, what are the properties of LHS regarding the joint properties of the subsampled distributions?*

  Conditioned LHS critically seeks to preserve the joint properties of the multivariate distributions (see e.g., Christierson et al., 2012).

- *Section 2.2.3 and Table 3: A list of required input variables for each hydrological model should be given. Furthermore, since the evapotranspiration process description is mentioned later on to be a potential reason for differences in low flow projections, a short description of the evapotranspiration routines should be included in Table 3 in a similar manner as the snow routines have been listed.*

  We agree with the referee. We opted in the first place not to mention these descriptions, but we now understand it may bring some relevant information. Roughly, two types of evapotranspiration modelling approaches may be identified: computation of actual evapotranspiration from energy balance models in

CLSM and ORCHIDEE, and use of Penman-Monteith potential evapotranspiration (Allen et al., 1998) for the other HMs. This information will be added to Table 3.

- *Section 3.1: Please indicate in the text and caption of Fig. 2 that the regimes were estimated based on reconstructed streamflows and not observations.*

  We will specify that naturalised streamflow time series have been used for estimating natural regimes.

- *Page 12660, lines 19-20: This is only true if the trend model is correctly separating the LSIV from the NFS for any given lead time. In general and given the linear trend model, it is likely that LSIV and the SSIV are overestimated (see also discussion in Raisänen 2001 and Hingray and Saïd (2014). Please discuss this limitation here and at other text passages where the partitioning between NFS and variability is presented.*

  As mentioned in a response above, this will be commented in the "QE-ANOVA advantages and limitations" subsection of the Discussion.

- *Page 12660, line 20-22: I understand that the SSIV is generated using the stochastic SDM realizations which in turn use the GCM data as input. Thus, there might be some sort of interaction between the LSIV and SSIV. For e.g., the SDM might generate a different variability for a GCM that is at the high end of the range with respect to one that is at the low end of the range of projected changes. It would be good if the authors could comment on that and discuss this either here or later in the article. Is there a reason why not to construct a 2-way-ANOVA for the variability part of the data? Such an ANOVA could take interactions into account. The design is unbalanced, but this should not affect your sum of squares estimation in a more severe way as what you do in Eq. A3 and A4 where all available runs of a particular GCM are taken thereby giving more weight to the GCMs with more runs.*

  Unfortunately, it is not possible to use a standard 2-way ANOVA to disentangle the two internal variability components. Indeed, for a given chain GCM-SDM, SDM realization#1 of, say, GCM run#1, does not correspond to realization#1 of run#2, due to the stochastic nature of SDM realizations. No ANOVA with only fixed effects can therefore be applied here. Mixed-effects ANOVA models may be tested, but they cannot provide an assessment of possible interactions. Considering the comment on the number of runs, Equations (A3) and (A4) use an average over all available runs for a given model chain. There is therefore no larger weight given to GCMs with more runs.

  On the other hand, we agree that the SSIV value could in principle depend on the GCM. Figure 3 below shows the QE-ANOVA estimates of the SSIV for groups of hydrometeorological chains associated with each GCM. No clear dependence of SSIV to GCMs emerges from this figure: discrepancies between GCMs do exist but they may vary over time.

- *Page 12662, line 6: It is unclear how the time slice averages are calculated. Do you use some fixed time slices or a moving 30 year approach? Please clarify at some stage in the manuscript. I have noticed that this comes later in 4.2 but I would have expected it to be defined earlier.*

  Thank you for pointing out this possible source of confusion. We actually use a moving 30-yr approach. This will be clarified in the revised manuscript.

- *Section 4.2, subsection title: Same comment as for the title. Hierarchy is arguably not the best term here.*

  As mentioned above, we would prefer to keep this term here for supporting the message that hydrological uncertainty is qualitatively and quantitatively as important as climate uncertainty.

- *Page 12663, line 21: Although I agree that internal variability often is larger than* *other sources of uncertainty, the manuscript has up to this point not given a reason why this had to be expected.*

We will add some supporting information earlier in the manuscript to support the fact that this results was expected, notably by including results from other studies (Hawkins and Sutton, 2011, etc.), but also by referring to Figure 3 where internal variability components are both very high compared to the change signal from this particular hydrometeorological modeling chain, but also from the grand ensemble change signal.

- *Page 12663, line 27: "previous studies". References are needed to point the reader to the previous studies.*

  They actually refer to the studies referenced in the previous sentence (Lafaysse et al. 2014; Hingray et al, 2014). This will be clarified.

- *Page 12664, lines 2-4: Interesting to see that the same set of SDMs leads to different degrees of uncertainty distributions when mean streamflow or low flows are analysed. I would ask the authors to also include a short discussion on the relation to other uncertainty sources. Without knowing the details about the employed SDMs, it seems to me that all make use of a similar concept (analogues) and represent only a small part of all available SDMs. If more diverse SDMs would have been used, the SDMs might have contributed more to the total uncertainty.*

  The representativity of the set of SDMs applied here within the large superpopulation of possible SDMs (with reasonable skill) is an interesting question, which could be also posed for GCMs and HMs. It has to be noted that even if the three SDMs rely on the same basic idea of analogue resampling, the concepts for selecting analogue situations are quite different (see Table 2 and detailed description in Lafaysse et al., 2014). Moreover, Lafaysse et al. (2014) found large differences between different versions of a given SDM using slightly different sets

of predictors. It is therefore unclear whether more diverse SDMs (or a larger number of versions from the SDMs used) would contribute more to the total uncertainty. To come back to the different effects on different streamflow indicators, it has to be noted that low flows are much more dependent on catchment processes than annual streamflow, therefore reducing the impact of different SDMs. These comments will be added to Section 5.3.

- *Page 12664, lines 25-27: The authors use a lognormal distribution to transfer the estimated variances into confidence bounds. I think this cannot be done straightforwardly since the variance parameter in QE-ANOVA is estimated based on non-logarithmized data. In other words, from the QE-ANOVA you get an effect with is normally distributed with zero mean and some variance, but those parameters are not directly portable to a lognormal distribution, which can be seen by, for e.g., the fact that a lognormal distribution never has zero mean. Anyway, judging from the results in Fig. 10 that look fairly ok, I assume that the authors have taken this into account and there is just a need for more clarification in the text on how the estimated variance and mean parameters are transferred so that a lognormal distribution can be used.*

  The relevant transformations of mean and variance from a normal distribution to the ones of a log-normal distribution have indeed been used. This will be clarified in the revised manuscript.

- *Page 12665, lines 9: The authors should state that also here, the decrease in internal variability for the 30-year time averages is due to the decrease in the ensemble mean.*

  We agree. See also responses to previous comments on this topic.

- *Page 12665, lines 10-14: A discussion of the decrease in the internal variability is necessary. A link to the relevant equations in the appendix might be helpful* *for the interested reader. It should also be stated that this decrease is a direct consequence of the quasi-ergodic assumption and could be an artefact.*

This is right, and this issue will be added to the new "QE-ANOVA advantages and limitations" subsection of the Discussion, as detailed in the response to previous comments.

- *Page 12665, line 21: "...in 2033-2039, that is for 30 year time-slices starting before 2015." Unclear as both 2033-15 and 2039-15 are not less or equal than 2015.*

  The sentence indeed contains a typo. It should read: "... starting around 2020".

- *Page 12666, lines 13-17: Unclear sentence. I understand it in a way that you are discussing the time of emergence for the results based on yearly anomalies, however, I cannot see that the lines in Fig. 11 exceed the 95*

  The sentence is indeed unclear. It actually comments the differences between actual modelling chains and perfect ones. It notably states that with a perfect modelling chain, one may be able to detect the change signal one decade earlier for both catchments in summer. We will rephrase it accordingly in the revised manuscript.

- *Section 5.1: I would suggest including a discussion of the relation between hydrological model uncertainty and the performance of the hydrological models with respect to the analyzed variable - here low flows. The two LSMs (ORCHIDEE and CLSM) used are behaving quite differently from the rest of the hydrological models. If those two were excluded from the analysis, the hydrological model uncertainty would probably be quite a bit smaller. And I would also expect those two HMs to have a worse performance in the reference period than the remaining ensemble - of course due to their main goal to be a LSM rather than a catchment model.*

  The relation between a present-day performance and a climate change signal is being highly discussed in the literature for GCMs. And indeed, a similar reasoning may of course be followed for HMs. However, this study uses anomalies from a reference period as a target variable in order to remove any present-day bias, and to hopefully reduce such a relationship if it actually exists. See also the response to comments to referee #1 on this particular topic. About the specific HMs ORCHIDEE and CLSM now: while present-day performance of ORCHIDEE on the interannual variability of low flows is indeed low, CLSM performance is actually generally in the middle range of other models. This result thus does not corroborate a simple relation between present-day performance and future changes. A sensitivity test of the uncertainty decomposition results on the subset of HMs retained would therefore be interesting (and similarly on subsets of GCMs or SDMs) but out of the scope of the present study. However, the above comments will be added to the discussion in Section 5.3, together with corresponding answers to comments from referee #1. It has to be noted that a similar experiment on HM uncertainty evolution following removal of CLSM has been performed on another multimodel study on the Seine catchment by Habets et al. (2013).

- *Section 5.1, first paragraph: I would argue that also the common way how HMs are calibrated leads to larger uncertainties for low flows. If, for e.g., NSE is maximized, the model is fitted better to high values than low values as the squared deviations give more weight to high values.*

  Actually, each HM has been calibrated in a specific way, as this has been left to the discretion of each R2D2-2050 project partner. The following table summarizes the objective function used for each model and provides some additional comments.

  This table will be added to the Table 3 of the manuscript. The general guidelines for the R2D2-2050 project were aiming at having HMs able to correctly simulate the whole range of flows, and not specifically high flows. However, the ref-

Table 1. Hydrological model calibration details. KGE refers to the Kling-Gupta Efficiency (Gupta et al., 2009).

Acronym	Calibration approach	Objective function
GR5J	Optimisation	KGE on $\sqrt{Q}$
MORDOR	Optimisation	KGE on Q
CEQUEAU	Semi-distributed optimisation	multicriteria
J2000	Manual sensitivity analysis	–
CLSM	Manual calibration	KGE on Q plus bias
ORCHIDEE	–	–

eree is right when stipulating that different calibration approaches may lead to an increase in HM uncertainty in low flow changes. This also goes along the lines discussed in the manuscript P12670L25-P12671L12 about the uncertainty in hydrological parameters. Some comments on the different possible objective functions across HMs will be added to Section 5.3.

- *Page 12667, lines 17-19: Isn't this to be expected since HM's fraction of variance is estimated based on the linear trend fit as well as the internal variability is very much smoothed due to the quasi-ergodic assumption, therefore removing a large part of the variability in time? The authors should discuss that the smoothness comes at the cost that one relies on the assumptions made.*

  This is right of course, and it indeed directly follows the assumptions made in the QE-ANOVA method. The underlying hypotheses behind this method are that both the Noise-Free Signal (NFS) and the internal variability evolve in a smooth way over time. These hypotheses appear quite reasonable as they ensue from a global gradual phenomenon – the increase in greenhouse gas concentrations – whose consequences are themselves gradual, at least within the time frame considered in this study. What can be discussed is how these hypotheses are implemented here in the uncertainty decomposition method, i.e. through (1) the

choice of a linear trend for estimating NFSs and (2) the choice of a constant coefficient of variation of internal variability over time. The above comments will be added to the discussion on the advantages and limitations of the QE-ANOVA method.

- *Figure 1: The coordinate system is not defined here. Preferably, the coordinates should be converted to Lat/Lon or at least the projection specifications for the lambert projection should be indicated.*

  We prefer sticking to the projection most commonly used in France. This is a Lambert conformal conic projection called "Lambert II étendu" with parameters specified in this document for example: http://www.ign.fr/sites/all/files/geodesie_projections.pdf. This reference will be added to the figure caption.

Technical comments

- *Page 12653, line 3: Should be "water manager's"*

  This will be corrected.

- *Page 12666, line 9: "an unchanged ..." instead of "a unchanged..."*

  This will be corrected.

References

Allen, R. G., Pereira, L. S., Raes, D. & Smith, M.: Crop Evapotranspiration – Guidelines for computing crop water requirements, FAO Irrigation and Drainage Paper 56, FAO, 1998.

Christierson, B. v., Vidal, J.-P. & Wade, S. D.: Using UKCP09 probabilistic climate information for UK water resource planning, J. Hydrol., 424-425, 48-67, doi: 10.1016/j.jhydrol.2011.12.020, 2012.

Gupta, H. V., Kling, H., Yilmaz, K. K. & Martinez, G. F. : Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 90-91, doi:10.1016/j.jhydrol.2009.08.003 ,2009.

Habets, F., Boé, J., Déqué, M., Ducharne, A., Gascoin, S., Hachour, A., Martin, E., Pagé, C., Sauquet, E., Terray, L., Thiéry, D., Oudin, L. & Viennot, P.: Impact of climate change on the hydrogeology of two basins in northern France, Climatic Change, 121, 771-785, doi: 10.1007/s10584-013-0934-x, 2013.

Hawkins, E. & Sutton, R.: The potential to narrow uncertainty in projections of regional precipitation change, Clim. Dyn., 37, 407-418, doi:10.1007/s00382-010-0810-6, 2011.

Hingray, B., Saïd, M.: Partitioning internal variability and model uncertainty components in a multimember multimodel ensemble of climate projections. J. Climate, 27, 6779-6797. doi:10.1175/JCLI-D-13-00629.1, 2014

Hingray, B., Blanchet, J. & Vidal, J.-P.: Uncertainty components estimates in transient climate projections. 2. Robustness of unbiased estimators in the single time and time series approaches, J. Climate, submitted.

Hingray, B., Hendrickx, F., Bourqui, M., Creutin, J.-D., François, B., Gailhard, J., Lafaysse, M., Lemoine, N., Mathevet, T., Mezghani, A. & Monteil, C. RIWER2030. Climat Régionaux et Incertitudes, Ressource en Eau et Gestion associée de 1860 à 2100, Final Report, ANR, Grenoble, France, 2013.

Northrop, P. J., Chandler, R. E.: Quantifying Sources of Uncertainty in Projections of Future Climate. J. Climate, 27, 8793-8808. doi:10.1175/JCLI-D-14-00265.1, 2014

Räisänen,J., 2001: CO2-induced climate change in CMIP2 experiments: Quantification of agreement and role of internal variability. J. Climate, 14, 2088-2104, doi:10.1175/1520-0442(2001)014,2088:CICCIC.2.0.CO;2

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 12, 12649, 2015.

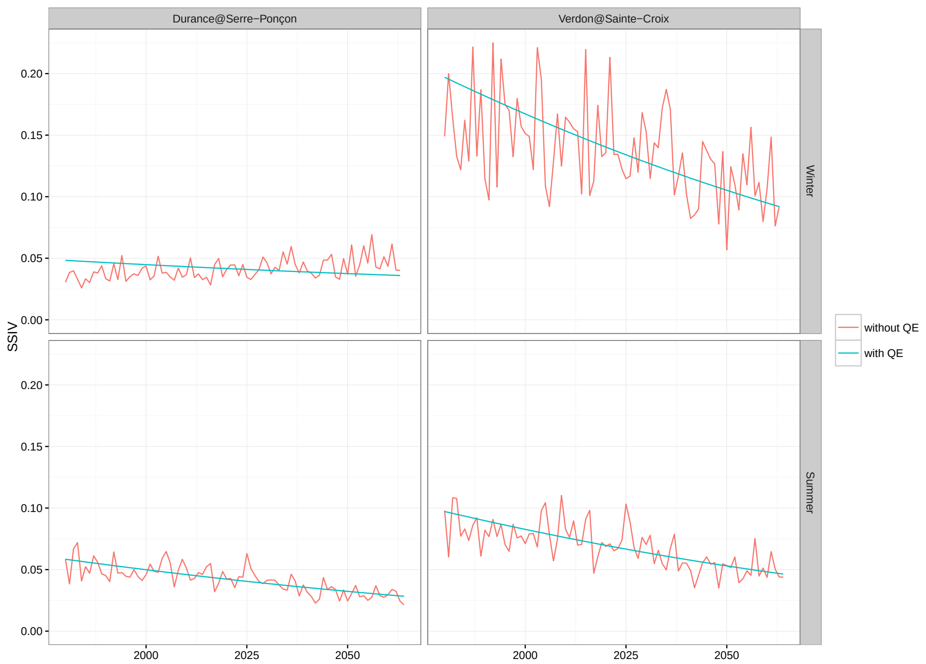

Fig. 1. Temporal evolution of SSIV with and without the quasi-ergodic (QE) assumption.

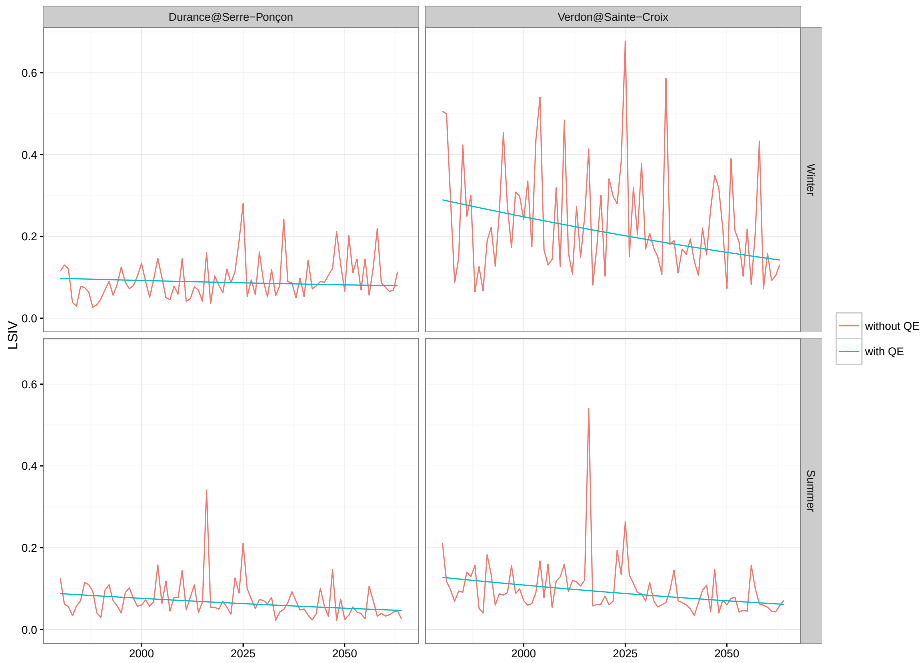

Fig. 2. Temporal evolution of LSIV with and without the quasi-ergodic (QE) assumption.

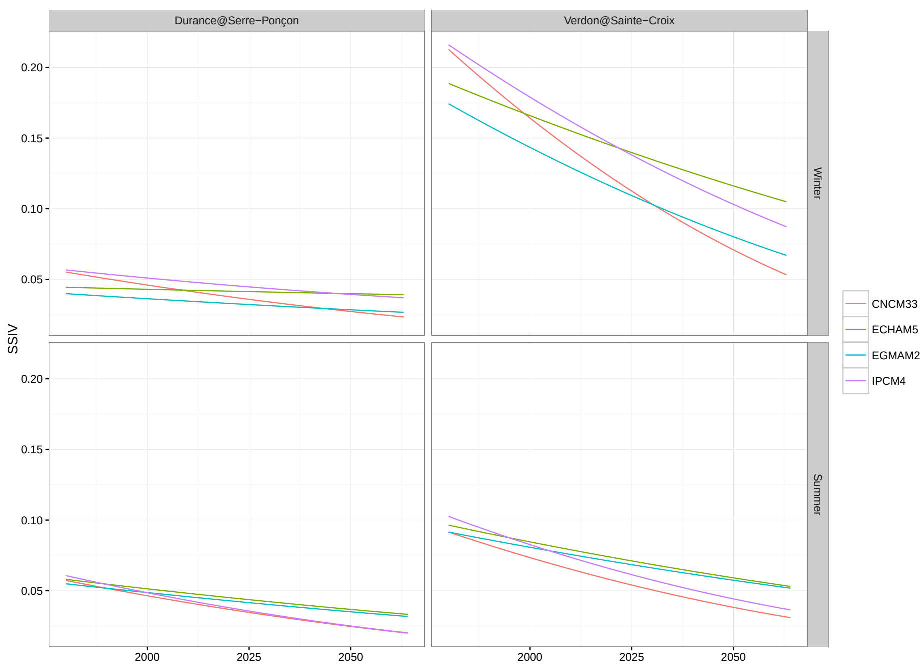

Fig. 3. Temporal evolution of SSIV computed for groups of hydrometeorological chains associated with each GCM.