Reviewer 1

General analysis: First of all, I want to apologize because the Buendia et al. (2016) paper that states "Overall, results have indicated that increased forest areas are the major driver of reduced streamflows and the magnitude of peak floods" is not the Buendia et al. (2016) paper quoted by the authors but another one (http://dx.doi.org/10.1016/j.scitotenv.2015.07.005) published two months later studying another basin. Indeed, the paper quoted by the authors found that for the embracing studied catchment (Talarn), somewhat less than 50% of the runoff reduction (37%) could be attributed to forest cover encroachment, but these authors state that "Neglectig re-vegetation could lead to erroneous projections resulting in an underestimation of the runoff future trends; thus, evolution of forested cover should not be ignored when designing land and river basin management plans at the light of global change scenarios". Therefore, this paper should not be fairly cited as a reason to omit the role of land cover change in streamflow temporal trend studies.

Response: Thank you again for all your feedback and recommendations to improve our manuscript. We cited Buendia et al. (2016) as an example of study where land use changes, while important, are not the main driver of runoff reduction. Indeed, for the Talarn catchment 37% of runoff reduction is attributed to forest cover encroachment, and while this is a major impact and omitting this land use changes for this particular catchment could lead to erroneous results, the main driver is still climate variability. Moreover, runoff reduction attributed to forest encroachment in the other two basins in Buendia et al. (2016) is 6% and 16% respectively, showcasing how widely the impact of land use changes can vary from basin to basin.

We have reformulated this section so it is more clear that we are not omitting land use changes in our study based on Buendia et al. (2016). We are omitting them based on the analysis of trends in model residuals, as recommended in the first revision, where it is shown that there are no clear trends that could be attributed to their omission, and therefore we can proceed with our analysis. However, we are not negating the importance of land use changes and their impact on streamflow, and that is why we included this discussion in our manuscript, but for the purpose of this study we have considered that their inclusion is not necessary.

Lines 445-450:

However, climate variability rather than afforestation is usually the main driver of streamflow reduction (Buendia et al., 2016). To confirm whether in our study this assumption is valid, and it is reasonable to exclude these land use changes without compromising model results, we performed an analysis of trends in model residuals (see section 1 of the Supplement) using the forested area in 2018 for the whole simulation period. This analysis does not evidence the presence of a factor other than the ones already included in the model affecting the hydrological response, and thus justifies the exclusion of land use changes in our study.

I want to acknowledge the effort made by the authors to follow my recommendations. The new information provided is noteworthy but not easy to understand, so I am trying to analyse it and to provide updated recommendations to the authors.

Response: Thank you very much, we have done our best to incorporate these new recommendations and further improve our manuscript.

Trends: The data shown in the table S1 are really striking as they point to relevant internal inconsistencies in the model results. First, the fact that 70% of the gauging stations show positive or negative significant residual trends, with a coefficient of variation of 822%, demonstrates a high uncertainty of the model results. Second, the spatial distribution of these trends in figure S1 shows disordered patterns; some successive gauging stations without any significant tributary between them show opposite trends, such as Castellbell and Abrera stations on the Llobregat River and the two Les Masies de Roda stations on the Ter River. Third, some of the plots in figure S2 show very asymmetric abnormal shapes, either positive (Balsareny, Fogars de la Selva (Pont Eiffel), Sallent) or negative (Guixers, Sant Feliu de Buixalleu).

Response: Thank you for your comment. However, we disagree that these results point to an inconsistent model performance. Despite most trends being statistically significant, their small R^2 and Figure S2 evidence that most stations do not present any clear tendency. Moreover, the ones that do – for example, Guixers (Cardener – Monegal) and Vilanova de Sau – have a limited number of observations, and thus the trend is not representative of the entire simulated period. The unequal distribution of observed streamflow can also explain the opposite residual trends between successive gauging stations, as coincidentally the Abrera and Les Masies de Roda (Ter) stations have observations for only half the simulation period, while Castellbell and Les Masies de Roda (Ter i Gurri) have a more complete record. Also, in relation to this and as per the suggestion of Reviewer 2, we have also computed the increase in forested area in the actual drainage area of each of the 50 gauging stations (Table S1), and added to the discussion.

Lastly, the "asymmetric abnormal shapes" in Figure S2 are due to the daily model residuals in m^3/s being shown, and therefore the plot is skewed towards the larger residuals during peak flows. This only indicates that some gauging stations tend to overestimate (e.g., Sant Feliu de Buixalleu) or underestimate (e.g., Balsareny) peaks, which does point to worse model performance at these stations, but we do have to consider that there are no perfect models and the fact that we are using so many gauging stations means that it is not possible to find a very good fit for all of them.

Lines 8-15 of the Supplement:

However, we must consider that the increase in forested area varies locally, and not all gauging stations have observed data for the whole period (2001-2022). Therefore, we have used the Corine Land Cover maps from 2000, 2006, 2012 and 2018 (EEA, 2000, 2006, 2012, 2018) to determine for each gauging station the increase in forested area within their drainage area during the closest period to the actual observations (Table S1). Of the 50 gauging stations, most (74%) present an increase in forested area, consistent with the general increase for the whole study area, while 14% present a decrease, and the remaining 12% either do not present any change in forested area or it could not be calculated because the observations start in 2018. Therefore, if the hypothesis that these land use changes are a main driver for streamflow reduction is true, we would expect to observe decreasing trends in model residuals for the gauging stations with an increase in forested area, and vice versa.

Lines 24-34 of the Supplement:

However, of the 37 gauging stations with an increase in forested area, only 54% show a decreasing trend in model residuals (55.6% when only considering statistically significant trends). Similarly, of the 7 gauging stations with a decrease in forested area, 4 show an increasing trend in model residuals, but only 2 are significant. In summary, of the 44 gauging stations where a change in forested area is observed, only 17 (38.6%) present a trend in model residuals consistent with the change. Moreover, we also must consider that R^2 is very small for all trends, with a maximum of 0.098, although 82% of trends present values of $R^2 < 0.01$. Therefore, as we do not observe clear trends in model residuals, we can reasonably assume that land use changes in our study are not a main driver influencing streamflow and their omission for the purpose of our analysis is not incorrect.

Table S1. Trends in the model residuals and change in forested area in the catchment area of each gauging station. Significant trends are marked in bold (p-value < 0.05). LR: Linear Regression. The units for LR slopes are m3/s/day.

Gauging station	Change in forested area (%)	LR slope	LR p-value	R^2	Likelihood
Abrera	+16.13	-7.43E-04	**5.55E-04**	3.21E-03	virtually certain
Balsareny	+11.60	2.74E-04	**3.49E-18**	9.55E-03	virtually certain
Berga	+0.71	2.25E-04	**4.21E-07**	4.91E-03	virtually certain
Cardona	-1.27	2.00E-04	**1.14E-17**	1.08E-02	virtually certain
Castellar de n'Hug	+10.26	1.60E-05	**2.71E-02**	7.93E-04	very likely
Castellbell i el Vilar	+16.05	5.04E-04	**3.69E-16**	8.31E-03	virtually certain
Castellbisbal	-1.59	2.25E-04	1.93E-01	5.58E-04	likely
Castellet i la Gornal	+16.89	-1.26E-05	**3.39E-03**	2.08E-03	virtually certain
El Papiol	+7.26	3.68E-05	**4.20E-03**	2.10E-03	virtually certain
Esponellà	+8.09	-3.20E-04	**5.70E-12**	6.00E-03	virtually certain
Fogars de la Selva (Can Simó)	+4.99	-2.07E-04	**1.64E-15**	8.84E-03	virtually certain
Fogars de la Selva (Pont Eiffel)	+2.05	-1.05E-04	**2.22E-05**	2.42E-03	virtually certain
Girona (Onyar)	+19.66	-1.65E-06	9.35E-01	8.18E-07	very unlikely
Girona (Ter)	+8.64	-1.10E-04	3.96E-01	9.36E-05	about as likely as not
Guardiola de Berguedà	+14.72	1.42E-04	**1.96E-07**	3.47E-03	virtually certain
Guixers (Aigua de Valls)	0	8.39E-04	**7.85E-23**	2.19E-02	virtually certain
Guixers (Cardener - Monegal)	-	-7.24E-04	**1.38E-34**	9.80E-02	virtually certain
Jorba	+20.84	-1.08E-05	6.43E-02	9.39E-04	very likely
La Cellera de Ter	+8.71	2.24E-04	1.35E-01	3.39E-04	likely
La Coma i la Pedra	+32.51	-2.33E-04	**2.05E-11**	1.73E-02	virtually certain
La Garriga	+9.62	3.55E-05	**6.73E-09**	4.19E-03	virtually certain
La Pobla de Claramunt	-0.37	-2.85E-05	**4.83E-09**	8.48E-03	virtually certain
Les Masies de Roda (Ter i Gurri)	+8.05	3.38E-04	**1.42E-05**	2.95E-03	virtually certain
Les Masies de Roda (Ter)	-0.13	-4.37E-04	**4.90E-02**	1.26E-03	very likely
Martorell	-0.26	-9.67E-06	8.40E-01	9.89E-06	unlikely
Montornès del Vallès	+6.65	-1.25E-05	**9.17E-03**	9.57E-04	virtually certain
Montseny	+20.40	-3.41E-06	6.52E-01	3.57E-05	about as likely as not
Navès	+10.24	1.98E-05	7.38E-02	6.49E-04	very likely
Olot	+6.48	-7.09E-05	**1.27E-14**	7.67E-03	virtually certain
Puig-reig	-3.58	2.04E-05	1.76E-01	4.04E-04	likely
Ripoll	+7.43	9.11E-05	6.15E-02	4.35E-04	very likely
Riudellots de la Selva	+11.10	-1.91E-04	**2.43E-34**	2.40E-02	virtually certain
Sallent	+14.88	3.81E-05	**4.14E-11**	6.75E-03	virtually certain
Sant Celoni	+12.71	-2.90E-06	6.22E-01	3.08E-05	about as likely as not
Sant Feliu de Buixalleu	+9.10	4.65E-05	**7.89E-03**	1.91E-03	virtually certain

Sant Gregori	+5.72	-4.30E-05	**3.03E-05**	3.13E-03	virtually certain
Sant Joan de les Abadesses	+9.36	5.34E-05	**2.58E-02**	6.28E-04	very likely
Sant Joan Despí	+14.78	5.75E-04	**9.37E-08**	3.73E-03	virtually certain
Sant Sadurní d'Anoia	+14.53	2.43E-05	9.73E-02	3.54E-04	very likely
Sant Vicenç dels Horts	-1.48	1.56E-03	**7.14E-20**	1.38E-02	virtually certain
Santa Coloma de Gramenet	+5.48	-1.17E-04	**5.37E-06**	2.84E-03	virtually certain
Santa Cristina d'Aro	+36.08	3.14E-06	8.01E-02	5.22E-04	very likely
Santa Perpètua de Mogoda	+2.90	-1.04E-04	**1.10E-24**	2.79E-02	virtually certain
Serra de Daró	+11.39	-8.87E-05	**7.51E-03**	1.09E-03	virtually certain
Torelló	0	1.21E-04	**2.92E-02**	1.62E-03	very likely
Torroella de Montgrí	+9.54	-4.62E-04	**9.68E-03**	8.90E-04	virtually certain
Tortellà	+0.69	-3.08E-04	**1.94E-06**	1.05E-02	virtually certain
Vilada (Merdançol)	0	-3.75E-06	4.59E-01	2.02E-04	about as likely as not
Vilada (Riera Vilada)	0	2.36E-05	4.87E-01	1.41E-04	about as likely as not
Vilanova de Sau	-	1.00E-03	**9.33E-13**	5.01E-02	virtually certain

Negative trends of the residuals may be attributed to the role of increased forest cover in the area, a very likely behaviour already demonstrated by previous works, but here it is more difficult to attribute positive trends to hydrological reasons.

Response: As mentioned above and in section 1 of the Supplement, we disagree that the residual trends, despite many being statistically significant, can be attributed to the increase of forest cover or other hydrological factors. Instead, the clear lack of trends consistent to land use changes evidence that these (which are the only potentially main factor affecting streamflow that we do not already include in our model) are not in fact a main factor in the overall study region. This confirms our hypothesis and allows us to proceed with the analysis of hydrological indicators with the model as it is. We reiterate again that this does not mean that the model is a perfect fit, but that it is an acceptable approximation which allows us to perform a more comprehensive analysis of spatio-temporal trends and patterns that we would with the gauging stations available.

I wonder whether the calibration-validation strategy used (randomly selected for each station independently) may lead to different sensitivities of the model to climate forcing at each station in such a varied climate and induce this scatter. As Buendia et al (2016) stated "precipitation appears to follow a generalised decreasing trend, although the significance of these results depended strongly on the time period considered". In fact, the authors do not provide any evidence of the validity of their consideration (line 371): "We consider that, given the high number of gauging stations in our study, randomly determining the calibration and validation periods for each station effectively captures all spatio-temporal variability in both periods. Thus, a bootstrapping method to repeatedly resample the calibration and validation periods, which is time consuming and would imply running many more iterations, is not necessary".

Response: Thank you for your comment. We do not believe that the calibration-validation strategy leads to different sensitivities to climate forcing at specific gauging stations. While it is true that some station might not capture all its climatic variability within its calibration period (e.g., calibration falls during a prolonged drought), other gauging stations within the basin will compensate, and due to the parameter calibration being basin-wide, for the particular gauging station the model will still be able to simulate climatic conditions not fully captured in the calibration period.

This is the reason for our consideration in line 371 (now line 375), so we have added to it to make it clearer. However, we agree that we do not provide any evidence of that beyond this reasoning. Originally, we started working on a different paper where we tested this calibration-validation strategy among many others, but unfortunately the main researcher working on this exercise left the project before it could be completed. Nonetheless, we already include this discussion in lines 381-388.

Lines 377-379:

In other words, even if the random selection for a specific gauging station leads to a calibration/validation period which does not capture all climatic variability, this variability will be included in other gauging stations, and due to the calibration process being at the basin scale, the model will still account for all variability.

SWAT+: This is not a physically-based model even if it can provide with good results. This qualification of this model contributes to the degradation and loss of usefulness of terminology and concepts. The addition of complementary processes does not modify its essentially empirical character. There is an agreed methodological caveat that just because a model gives good results does not imply that it is for the good reasons (in particular, structure, internal stores and flows). In fact, the results shown by the authors in table S1 may provide a corollary of this principle: despite acceptable flow calibration/validation tests, residual discharge trends show chaotic values difficult to attribute to hydrological reasons.

Response: We have removed the descriptor "physically-based".

Lines 125-127:

SWAT is a semi-distributed ecohydrological model widely used worldwide (Abbaspour et al., 2015; Gassman et al., 2014; Samimi et al., 2020), including many applications in Mediterranean basins (Boithias et al., 2017; Brouziyne et al., 2021; De Girolamo et al., 2022).

Recommendations:

In general terms, the manuscript describes the modelling exercise, shows its results, makes some comparisons with observed data and claims the success of the exercise as it "led to successfully simulating hydrological and anthropogenic processes in water-scarce Mediterranean basins" and "resulted in notable improvements in hydrological modelling and its potential use to support decision-making in the water management sector" without contributing no evidence of these successes and improvements.

Response: Regarding the first claim, we believe that the satisfactory values of the objective functions achieved after calibration and validation are enough evidence of having successfully simulated hydrological and anthropogenic processes in several water-scarce Mediterranean basins. However, regarding the second claim, we agree that we do not properly quantify the improvements in hydrological modelling, so we have reformulated this sentence.

Lines 533-539:

The spatio-temporal analysis of streamflow patterns and trends have provided insights into the evolution of hydrological dynamics under climate change and increasing anthropogenic pressures in basins vulnerable to water scarcity. Moreover, the integration of first-hand expert knowledge from water managers into our modelling framework has resulted in a more realistic simulation of anthropogenic process, highlighting the potential use of this model to support decision-making in the water management sector. Lastly, the introduction of a randomised calibration and validation approach allows us to overcome the limitations and biases arising of conventional approaches when dealing with multiple gauging stations of variable length without the need for complex analysis.

The authors should not claim good modelling results beyond acceptable tests of efficiency and uncertainty, but should be much more analytical by discussing their strengths and weaknesses and suggesting ways to remedy the latter.

Response: We disagree that we should not claim good modelling results beyond acceptable tests of efficiency and uncertainty, because this is in fact the aim of these tests. We agree of course that any model can be subject to improvements so that it can more closely represent the real system, but for the scope of this study, our model results are proven to be sufficient. In the discussion (mainly section 4.1 but also throughout the other sections) we already underline the weaknesses arising from our approach as well as strategies to address them in future studies.

- The authors cannot justify the reason for the omission of the role of land cover in the hydrological changes of this area on the basis of any published work, nor justify the validity of this omission on the modelling results which are largely inconsistent in this respect.

Response: As per a previous response, we have reformulated part of the text, so it is more clear that we are not omitting the role of land use cover in hydrological changes on the basis of a specific published work. However, we disagree that we cannot justify this omission on our modelling results (section 1 of the Supplement).

Lines 445-450:

However, climate variability rather than afforestation is usually the main driver of streamflow reduction (Buendia et al., 2016). To confirm whether in our study this assumption is valid, and it is reasonable to exclude these land use changes without compromising model results, we performed an analysis of trends in model residuals (see section 1 of the Supplement) using the forested area in 2018 for the whole simulation period. This analysis does not evidence the presence of a factor other than the ones already included in the model affecting the hydrological response, and thus justifies the exclusion of land use changes in our study.

- Following Bieger et al (2017), the authors can claim that SWAT is "one of the most widely used hydrologic models in the world" but cannot claim that it is a physically-based model.

Response: We have removed the descriptor "physically-based".

Lines 125-127:

SWAT is a semi-distributed ecohydrological model widely used worldwide (Abbaspour et al., 2015; Gassman et al., 2014; Samimi et al., 2020), including many applications in Mediterranean basins (Boithias et al., 2017; Brouziyne et al., 2021; De Girolamo et al., 2022).

The model parameters were optimized to obtain the best simulation of discharges, but not the various 'hydrological indicators' extensively exposed in the manuscript. Therefore, due to the equifinality problem (Beven, 2006; Kirchner, 2006), various sets of model parameter may give discharge efficiencies very similar to the best one, but may give quite divergent values of these 'hydrological indicators'.

Response: Thank you for your comment, we agree that the equifinality problem inherent to most hydrological models may result in different values for the hydrological indicators. Therefore, we have also calculated the hydrological indicators using the upper and lower limits of the 95PPU uncertainty bands, so that we can propagate this uncertainty to the Sen's slopes. We visualize this uncertainty in Figure S10 (Section 3 of the Supplement), and we have added the following discussion:

Lines 81-84 of the Supplement:

Figure S9 shows the observed and simulated Sen's slopes of each of the 40 indicators, while Fig. S10 shows only the indicators for which both the observed and simulated trend are significant (Mann-Kendall, p-value < 0.05), as well as the uncertainty associated to the simulated Sen's slopes due to model equifinality. While some of the significant pairs present different directions, the majority are both either positive or negative, even considering the uncertainty.

Lines 92-96 of the Supplement:

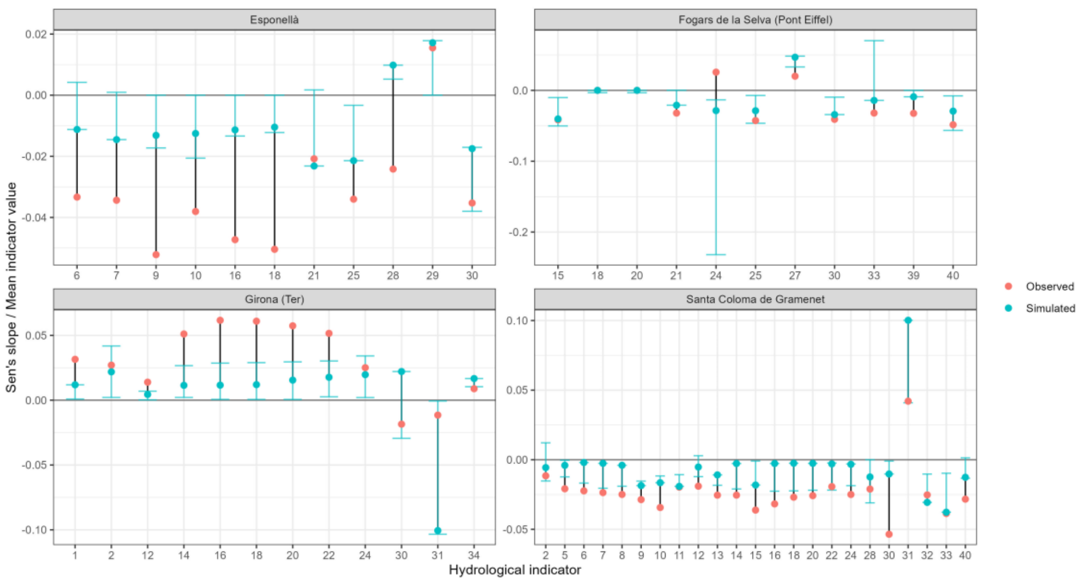

Figure S10: Ratio of Sen's slope to mean indicator value only for significant pairs of observed and simulated trends. Uncertainty associated to simulated Sen's slope is also shown. For indicators 18 and 20 in the Fogars de la Selva plot, the Sen's slope is not divided by the indicator value. This is because due to the Sen's slope being very close to 0, the uncertainty value became too large when standardizing, and so it masked the other indicators. See Table S3 to match each indicator to the number used in the figure.

Consequently, model-simulated 'hydrological indicators' face to two severe uncertainties: the role of land cover changes and the issue of model equifinality. It is not possible to determine whether the differences between modelled and observed trends of these indicators are due to the role of land cover change or modelling equifinality effects. Therefore, the statement in line 76 of the Supplement regarding trend analyses is not acceptable: "Moreover, the fact that streamflow was first calibrated ensures overall the validity of the analysis".

Response: As mentioned above, based on section 1 of the Supplement we disagree that land cover changes during our study period are a main driver of streamflow reduction. Also, the uncertainty due to model equifinality has been quantified (see above), and therefore we consider that the calibrated streamflow can be used to conduct the analysis. However, we have removed this line because it did not quite fit the rest of the text.

- Throughout the manuscript, the trends are shown as "Sen slopes", but the units are not always shown, especially for the time variable, so the value of the rate of change is not clear if it is per day, month or year.

Response: Thank you for your comment. We have added the units of slopes where it wasn't already specified.

Lines 293-295:

Table 5. Analysis of trends for the annual percentage of dry river segments, mean annual temperature and mean annual precipitation. Significant trends are marked in bold (p-value < 0.05). Slope units are %/year, ºC/year, and mm/year respectively. LR: Linear Regression; MK: Mann-Kendall.

Lines 305-306:

Figure 4: Spatial distribution of Sen's slope (a, units hm^3/year) and standardized Sen's slope (b, units $year^{-1}$) for the hydrological indicator total annual flow.

Lines 312-313:

Figure 5: Spatial distribution of standardized Sen's slope (units $year^{-1}$) for the hydrological indicators annual Q50 (a) and Q50 in January (b), in April (c), in July (d), and in October (e), representative of the different seasonal flow patterns.

Lines 328-330:

Figure 6: Spatial distribution of Sen's slope for the hydrological indicators Q90 (a), Q10 (b), number of high and low flow pulses (c-d), and their mean duration (e-f). Sen's slopes for Q90

and Q10 are standardized (a-b, units $year^{-1}$), while the units for the other indicators are number of events/year (c-d) and days/year (e-f).

Lines 351-353:

Figure 8: Spatial distribution of Sen's slope for the hydrological indicators rise rate (a), fall rate (b) and number of flow reversals (c). The units are number of reversals/year for the last indicator, while they are standardized for the rise and fall rates indicators (units $year^{-1}$).

Lines 32-33 of the Supplement:

Table S1. Trends in the model residuals and change in forested area in the catchment area of each gauging station. Significant trends are marked in bold (p-value < 0.05). LR: Linear Regression. The units for LR slopes are m^3/s/day.

Line 38 of the Supplement:

Figure S1: Spatial distribution of trend slopes (units m^3/s/day) for model residuals.

Removing the stronger influence of river segments with higher streamflow in the analysis of temporal trends does not seem to me a sound option when the main objective of the study concerns water resources. On the other hand, the comparison between slopes and mean flows (fractions of runoff gained or loss at annual intervals) are convenient to evaluate their importance in terms of water resources.

Response: We believe that for the purpose of this analysis the standardization of Sen's slopes with volume or flow units is well-grounded, as it allows us to better identify areas of the stream network which follow similar trends regardless of magnitude. This can be observed in Figure 4, where without standardization the interpretation of spatial patterns in temporal trends for Total Annual Flow is skewed due to the larger main reaches of the Llobregat and Ter rivers, but with standardization it can be discerned that smaller tributaries can show in fact more notable trends. However, it is true that the non-standardized Sen's slopes are more relevant in terms of considering water resources, so we have added on to the discussion of Figure 4a.

Lines 297-303:

Significant trends in total annual flow show a general decrease, of up to 6 hm^3/year, except in the headwaters of the Llobregat basin (Fig. 4). However, this region shows a poorer model adjustment (see discussion on section 4.1), which may compromise the reliability of this result. Figure 4a shows that larger negative trends can be observed along the course of the main rivers, when significant, most notably along the lower course of the Llobregat river but also observed upstream of the reservoirs in the Ter river as well as near the mouths of the Besòs and Tordera rivers. Removing the stronger influence of reaches with higher flows, the standardized Sen's slopes show that the larger relative negative trends can be observed in smaller tributaries (Fig. 4b).

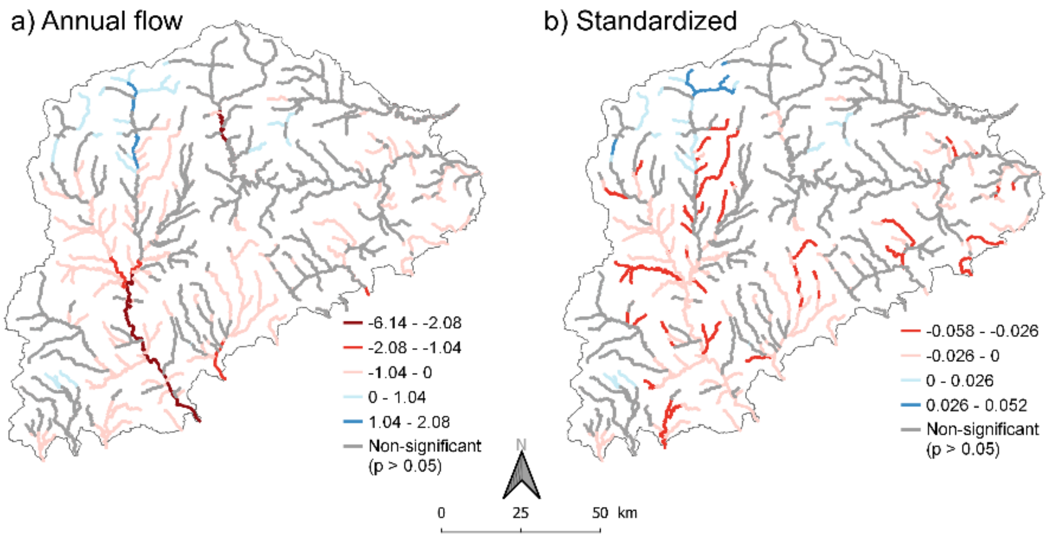

Figure 4: Spatial distribution of Sen's slope (a, units hm^3/year) and standardized Sen's slope (b, units $year^{-1}$) for the hydrological indicator total annual flow.

Lines 417-421:

Significant decreasing trends over the CRBD during the first two decades of the 21st century have been identified for medium and high flows (Fig. 5a & 6a), as well as total annual streamflow (Fig. 4). The larger absolute decreasing trends in total annual streamflow, between -2.08 and up to -6.14 hm^3/year, are found in the lower Llobregat and in the Ter river upstream of the reservoirs (Fig. 4a). Both of these areas are of notable interest from the water management perspective to supply Barcelona's metropolitan area, the largest urban centre within the CRBD.

- The hydrographs shown in figures 2 and S3 to S8 are very difficult to understand because the diverse plots cover each other. A logarithmic scale of the discharge axis might help to better visualize the plots.

Response: Thank you for your comment. We have changed Figure 2 for monthly hydrographs instead of daily, because even with a logarithmic scale, daily hydrographs in a single figure were not easy to visualise. Instead, the daily hydrographs (with a logarithmic y-axis) can be found in the Supplement (Figures S3-S8, including the 95PPU uncertainty bands).

Lines 273-275:

The comparisons between observed and simulated streamflow for representative gauging stations of the main rivers of the CRBD demonstrate good model performance (Fig. 2, also see section 2 of the Supplement for daily streamflow and 95PPU uncertainty bands).

Lines 280-282:

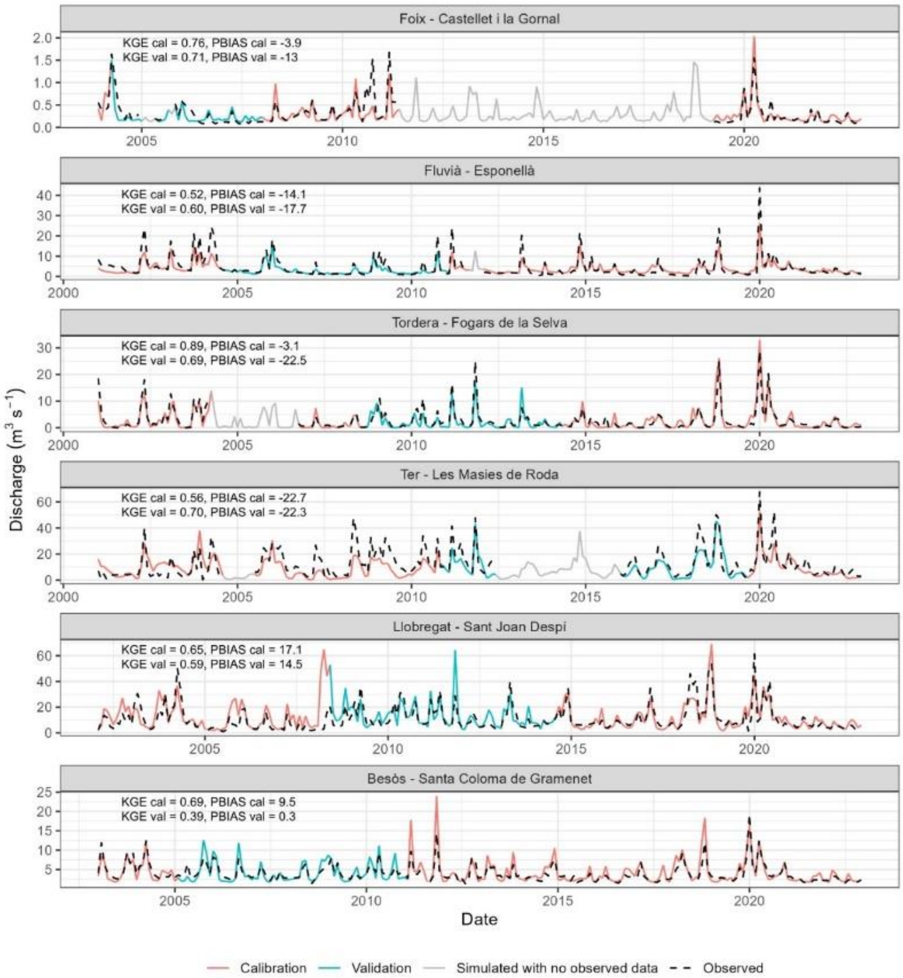

Figure 1: Observed and simulated monthly streamflow for the period 2001-2022 in the six main rivers of SWAT+CRBD. Individual KGE and PBIAS values for both the calibration and validation periods are also shown.

Lines 55-72 of the Supplement:

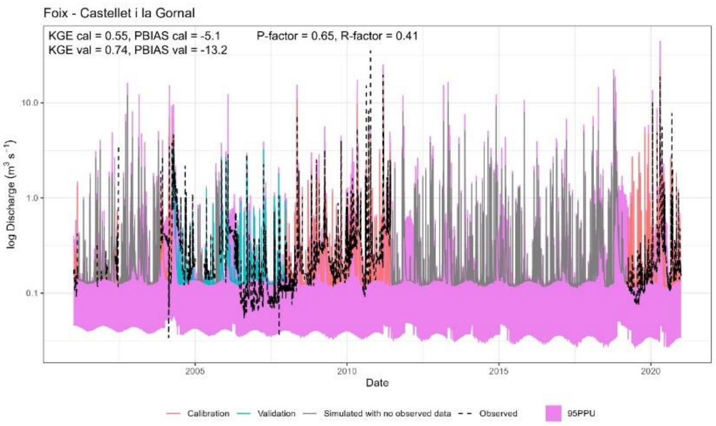

Figure S3: Observed and simulated daily streamflow, and the 95PPU band for the gauging station in Castellet i la Gornal, in the Foix basin. KGE and PBIAS values for the calibration and validation periods and the R-factor and P-factor are shown.

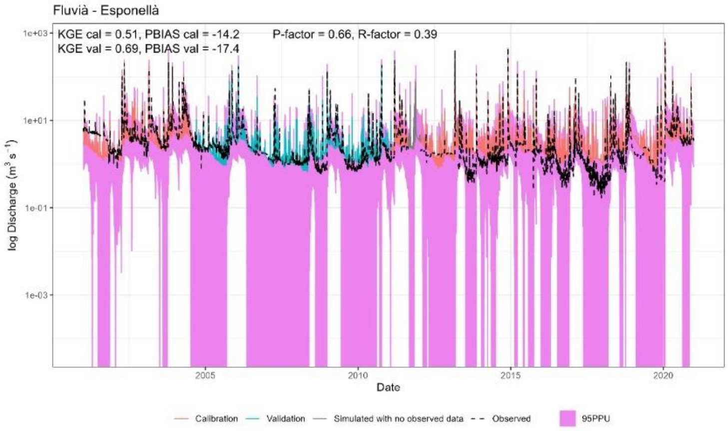

Figure S4: Observed and simulated daily streamflow, and the 95PPU band for the gauging station in Esponellà, in the Fluvià basin. KGE and PBIAS values for the calibration and validation periods and the R-factor and P-factor are shown.

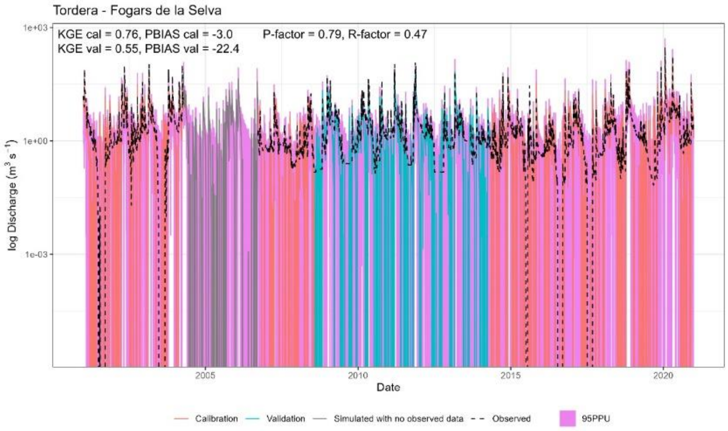

Figure S5: Observed and simulated daily streamflow, and the 95PPU band for the gauging station in Fogars de la Selva, in the Tordera basin. KGE and PBIAS values for the calibration and validation periods and the R-factor and P-factor are shown.

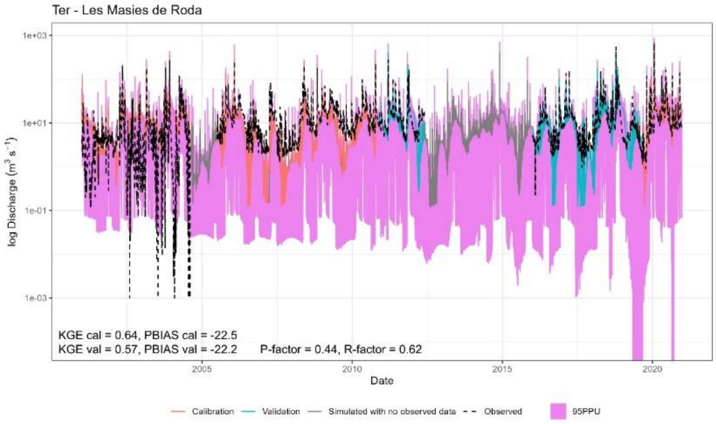

Figure S6: Observed and simulated daily streamflow, and the 95PPU band for the gauging station in Les Masies de Roda, in the Ter basin. KGE and PBIAS values for the calibration and validation periods and the R-factor and P-factor are shown.

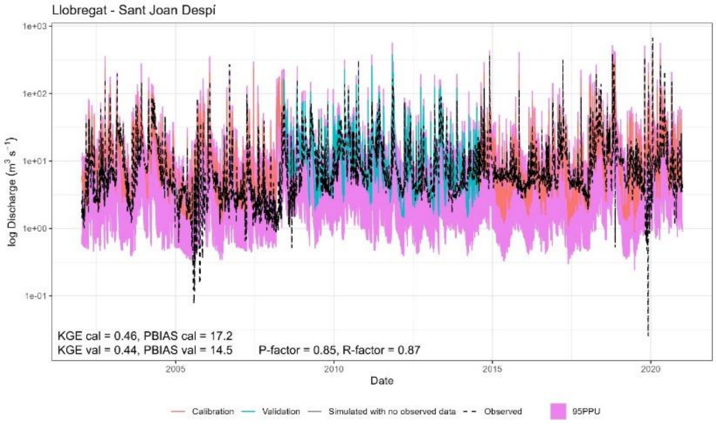

Figure S7: Observed and simulated daily streamflow, and the 95PPU band for the gauging station in Sant Joan Despí, in the Llobregat basin. KGE and PBIAS values for the calibration and validation periods and the R-factor and P-factor are shown.

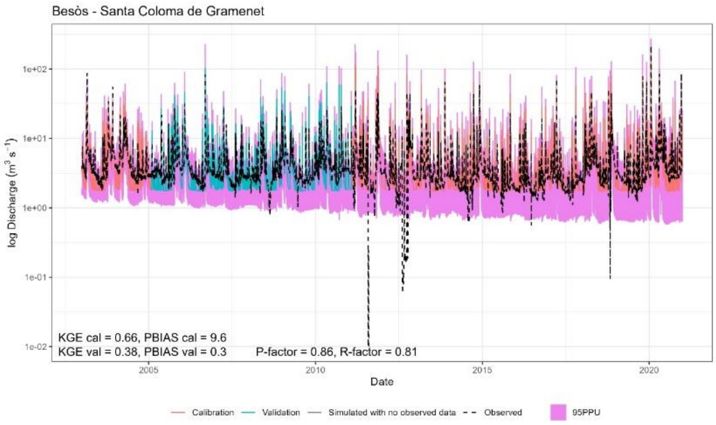

Figure S8: Observed and simulated daily streamflow, and the 95PPU band for the gauging station in Santa Coloma de Gramenet, in the Besòs basin. KGE and PBIAS values for the calibration and validation periods and the R-factor and P-factor are shown.

- The units for the variables x and y in the equation enclosed in Figure 3 are not stated. This graph seems to mix observed and simulated results, which should be explained.

Response: We have removed the equation from Figure 3, as we already show the linear regression slope and R^2 of the percentage of dry river segments in Table 5. We have also clarified what are simulation results (percentage of dry river segments) and what are observations (mean annual temperature and precipitation).

Lines 290-293:

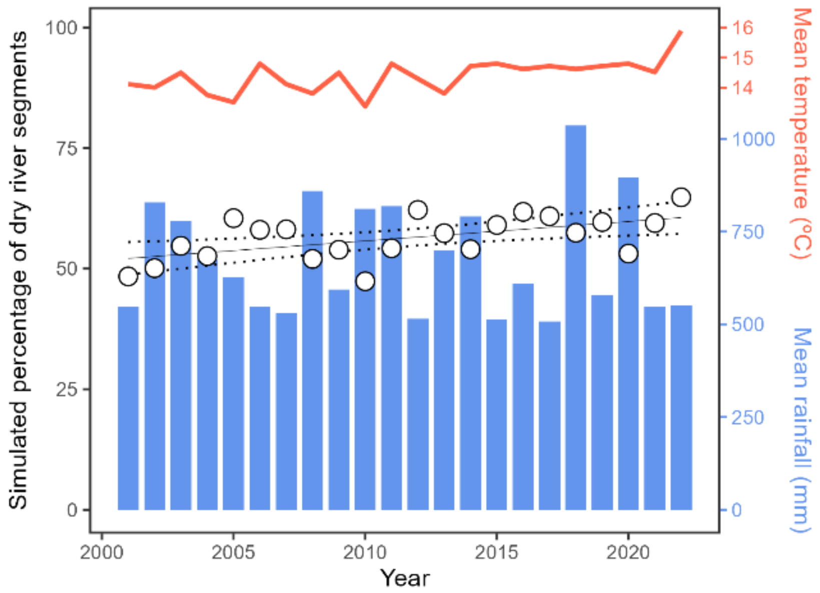

Figure 3: Evolution of simulated annual percentage of river segments that dry at least once a year, as well as observed mean annual rainfall and temperature.

- The analysis of trends in model residuals in section 1 of the Supplement does not justify the validity of excluding land-cover changes in the study, but demonstrates the difficulty of the modelling exercise to provide reliable estimates of the trends.

Response: As mentioned above, we disagree that the analysis of trends in model residuals in section 1 of the Supplement does not justify the omission of land use changes.

- The units of the slopes shown in table S1 are not shown. Both this table and figures S1 and S2 demonstrate the very inconsistent results of the model exercise with respect to these trends. These results must be further discussed and the sentence "we can reasonably assume that land use changes in our study are not a main driver influencing streamflow" should be deleted.

Response: We have added the units of the slopes in Table S1. However, we disagree that Table S1 and Figures S1 and S2 demonstrate inconsistent model results. Section 1 of the Supplement shows that despite seeing statistically significant trends in model residuals for 70% of the gauging stations, this statistically significance is questionable due to small R^2 values and the plots of Figure S2. Moreover, many trends are not consistent with the expected impact of omitting land use changes. Therefore, we can assume that land use changes do not play a factor in residuals trends and thus we can use our calibrated model to proceed with the analysis of trends in hydrological indicators, which was the aim of this exercise.

Lines 32-33 of the Supplement:

Table S1. Trends in the model residuals and change in forested area in the catchment area of each gauging station. Significant trends are marked in bold (p-value < 0.05). LR: Linear Regression. The units for LR slopes are $m^3/s/day$.

- The Y-axes in figures S9 and S10 are not appropriate because variables of diverse ranks are shown together. The ratio of slope to mean value (%) could be better as Y-axe units, as this could increase the visibility of low values and allow direct comparison between gauging stations because the axes could be equal.

Response: Thank you for your comment, we have changed the Y-axis units in Figures S9 and S10 to increase visibility of smaller values (and therefore we have excluded Figure S10b). The Y-axis is still different for each gauging stations, but the objective of these figures is only to show if trends have the same direction (negative or positive) using observed and simulated discharge, and not to compare between gauging stations.

Lines 89-96 of the Supplement:

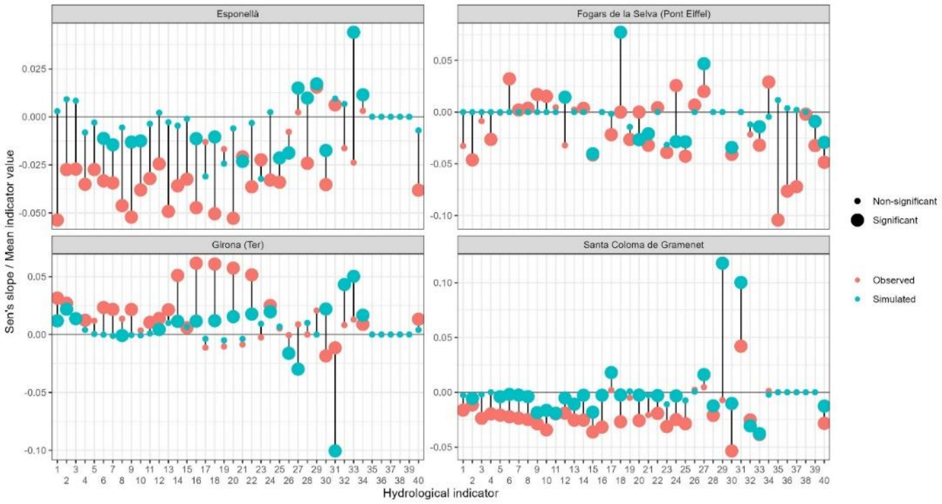

Figure S9: Ratio of Sen's slope to mean indicator value for observed and simulated trends. See Table S3 to match each indicator to the number used in the figure.

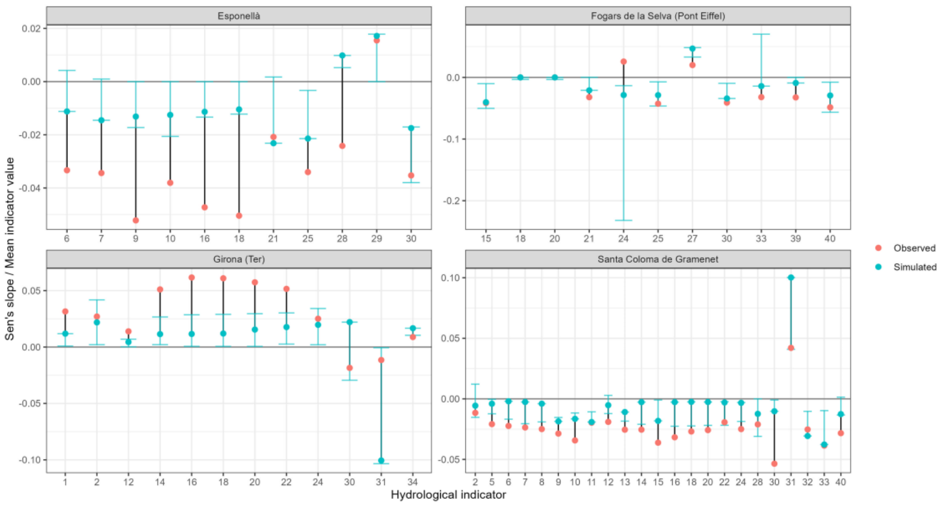

Figure S10: Ratio of Sen's slope to mean indicator value only for significant pairs of observed and simulated trends. Uncertainty associated to simulated Sen's slope is also shown. For indicators 18 and 20 in the Fogars de la Selva plot, the Sen's slope is not divided by the indicator value. This is because due to the Sen's slope being very close to 0, the uncertainty value became too large when standardizing, and so it masked the other indicators. See Table S3 to match each indicator to the number used in the figure.

Reviewer 1

The authors have taken on the comments and suggestions made in the first review concerning the reformulation of objectives and the definition of end-user input. The supplements are interesting and useful.

In my opinion, the manuscript is acceptable in this 2nd version.

Just one last comment:

The residue study in supplement 1 is a good idea. Trends are presented basin by basin (which is interesting). But these trends are only compared with the percentage increase in forest area over the entire study area. It would have been more convincing to show the variation in forest area (and/or urbanized area) in each basin beside the residual trends.

Response: Thank you very much for your review and your suggestion to improve the analysis of trends in model residuals. We have calculated for each of the 50 gauging stations the actual forest increase within their catchment area while also accounting for the period for which we have observed discharge, as some gauging stations are not representative of the whole 2001-2022 simulated period.

Lines 8-15 of the Supplement:

However, we must consider that the increase in forested area varies locally, and not all gauging stations have observed data for the whole period (2001-2022). Therefore, we have used the Corine Land Cover maps from 2000, 2006, 2012 and 2018 (EEA, 2000, 2006, 2012, 2018) to determine for each gauging station the increase in forested area within their drainage area during the closest period to the actual observations (Table S1). Of the 50 gauging stations, most (74%) present an increase in forested area, consistent with the general increase for the whole study area, while 14% present a decrease, and the remaining 12% either do not present any change in forested area or it could not be calculated because the observations start in 2018. Therefore, if the hypothesis that these land use changes are a main driver for streamflow reduction is true, we would expect to observe decreasing trends in model residuals for the gauging stations with an increase in forested area, and vice versa.

Lines 24-34 of the Supplement:

However, of the 37 gauging stations with an increase in forested area, only 54% show a decreasing trend in model residuals (55.6% when only considering statistically significant trends). Similarly, of the 7 gauging stations with a decrease in forested area, 4 show an increasing trend in model residuals, but only 2 are significant. In summary, of the 44 gauging stations where a change in forested area is observed, only 17 (38.6%) present a trend in model residuals consistent with the change. Moreover, we also must consider that R^2 is very small for all trends, with a maximum of 0.098, although 82% of trends present values of $R^2 < 0.01$. Therefore, as we do not observe clear trends in model residuals, we can reasonably assume that land use changes in our study are not a main driver influencing streamflow and their omission for the purpose of our analysis is not incorrect.

Table S1. Trends in the model residuals and change in forested area in the catchment area of each gauging station. Significant trends are marked in bold (p-value < 0.05). LR: Linear Regression. The units for LR slopes are m3/s/day.

Gauging station	Change in forested area (%)	LR slope	LR p-value	R^2	Likelihood
Abrera	+16.13	-7.43E-04	**5.55E-04**	3.21E-03	virtually certain
Balsareny	+11.60	2.74E-04	**3.49E-18**	9.55E-03	virtually certain
Berga	+0.71	2.25E-04	**4.21E-07**	4.91E-03	virtually certain
Cardona	-1.27	2.00E-04	**1.14E-17**	1.08E-02	virtually certain
Castellar de n'Hug	+10.26	1.60E-05	**2.71E-02**	7.93E-04	very likely
Castellbell i el Vilar	+16.05	5.04E-04	**3.69E-16**	8.31E-03	virtually certain
Castellbisbal	-1.59	2.25E-04	1.93E-01	5.58E-04	likely
Castellet i la Gornal	+16.89	-1.26E-05	**3.39E-03**	2.08E-03	virtually certain
El Papiol	+7.26	3.68E-05	**4.20E-03**	2.10E-03	virtually certain
Esponellà	+8.09	-3.20E-04	**5.70E-12**	6.00E-03	virtually certain
Fogars de la Selva (Can Simó)	+4.99	-2.07E-04	**1.64E-15**	8.84E-03	virtually certain
Fogars de la Selva (Pont Eiffel)	+2.05	-1.05E-04	**2.22E-05**	2.42E-03	virtually certain
Girona (Onyar)	+19.66	-1.65E-06	9.35E-01	8.18E-07	very unlikely
Girona (Ter)	+8.64	-1.10E-04	3.96E-01	9.36E-05	about as likely as not
Guardiola de Berguedà	+14.72	1.42E-04	**1.96E-07**	3.47E-03	virtually certain
Guixers (Aigua de Valls)	0	8.39E-04	**7.85E-23**	2.19E-02	virtually certain
Guixers (Cardener - Monegal)	-	-7.24E-04	**1.38E-34**	9.80E-02	virtually certain
Jorba	+20.84	-1.08E-05	6.43E-02	9.39E-04	very likely
La Cellera de Ter	+8.71	2.24E-04	1.35E-01	3.39E-04	likely
La Coma i la Pedra	+32.51	-2.33E-04	**2.05E-11**	1.73E-02	virtually certain
La Garriga	+9.62	3.55E-05	**6.73E-09**	4.19E-03	virtually certain
La Pobla de Claramunt	-0.37	-2.85E-05	**4.83E-09**	8.48E-03	virtually certain
Les Masies de Roda (Ter i Gurri)	+8.05	3.38E-04	**1.42E-05**	2.95E-03	virtually certain
Les Masies de Roda (Ter)	-0.13	-4.37E-04	**4.90E-02**	1.26E-03	very likely
Martorell	-0.26	-9.67E-06	8.40E-01	9.89E-06	unlikely
Montornès del Vallès	+6.65	-1.25E-05	**9.17E-03**	9.57E-04	virtually certain
Montseny	+20.40	-3.41E-06	6.52E-01	3.57E-05	about as likely as not
Navès	+10.24	1.98E-05	7.38E-02	6.49E-04	very likely
Olot	+6.48	-7.09E-05	**1.27E-14**	7.67E-03	virtually certain
Puig-reig	-3.58	2.04E-05	1.76E-01	4.04E-04	likely
Ripoll	+7.43	9.11E-05	6.15E-02	4.35E-04	very likely
Riudellots de la Selva	+11.10	-1.91E-04	**2.43E-34**	2.40E-02	virtually certain
Sallent	+14.88	3.81E-05	**4.14E-11**	6.75E-03	virtually certain
Sant Celoni	+12.71	-2.90E-06	6.22E-01	3.08E-05	about as likely as not
Sant Feliu de Buixalleu	+9.10	4.65E-05	**7.89E-03**	1.91E-03	virtually certain
Sant Gregori	+5.72	-4.30E-05	**3.03E-05**	3.13E-03	virtually certain
Sant Joan de les Abadesses	+9.36	5.34E-05	**2.58E-02**	6.28E-04	very likely
Sant Joan Despí	+14.78	5.75E-04	**9.37E-08**	3.73E-03	virtually certain
Sant Sadurní d'Anoia	+14.53	2.43E-05	9.73E-02	3.54E-04	very likely
Sant Vicenç dels Horts	-1.48	1.56E-03	**7.14E-20**	1.38E-02	virtually certain
Santa Coloma de Gramenet	+5.48	-1.17E-04	**5.37E-06**	2.84E-03	virtually certain
Santa Cristina d'Aro	+36.08	3.14E-06	8.01E-02	5.22E-04	very likely
Santa Perpètua de Mogoda	+2.90	-1.04E-04	**1.10E-24**	2.79E-02	virtually certain
Serra de Daró	+11.39	-8.87E-05	**7.51E-03**	1.09E-03	virtually certain
Torelló	0	1.21E-04	**2.92E-02**	1.62E-03	very likely
Torroella de Montgrí	+9.54	-4.62E-04	**9.68E-03**	8.90E-04	virtually certain
Tortellà	+0.69	-3.08E-04	**1.94E-06**	1.05E-02	virtually certain

Vilada (Merdançol)	0	-3.75E-06	4.59E-01	2.02E-04	about as likely as not
Vilada (Riera Vilada)	0	2.36E-05	4.87E-01	1.41E-04	about as likely as not
Vilanova de Sau	-	1.00E-03	**9.33E-13**	5.01E-02	virtually certain