Reviewer comments in black and responses in violet.

The manuscript by Tapia et al. describes P/Ca variations in several speleothems from Northern Iberia over two distinct climate transitions. The high increases in P/Ca at the onset of cool events are attributed to enhanced freeze-thaw cycling, facilitating the mobilization of P from the soil. The authors support their interpretation by a model sensitivity study. Overall, I think the data and the final conclusions are sound but I have at major aspects that I suggest to think about or include into the discussion, because these could influence the overall take-home messages a little bit.

In addition, I suggest to reduce the number of Figures in the main text and part of it possibly moved to a supplement. Also, I suggest to use color-blind friendly palettes. Likewise, the main text, in particular the results section, could be streamlined and focused on the most important results. My impression is that a large part of the results section just deals with processes that are then shown to be not dominant, so in my opinion this could be largely shortened or moved to the supplement.

We agree (also Reviewer 2) with the suggestion to reduce the results section focused on the modeling of processes of P incorporation into the calcite; the model results confirm that these processes cannot explain the large transient peaks in P/Ca. We therefore propose to follow the reviewers suggestion to move material to the Supplement, namely moving the former Figure 7 to the supplementary, along with the text in former Section 3.3 of results, which reviewed the outcomes shown in Figure 7. We propose to introduce the main conclusion from this model in the next section (formerly section 3.4) citing the supplementary figure.

Main comments:

1. Soil pH. If I get the authors right, they focus their analyses on the contribution of organic P. However, adsorbed and precipitated P may be released to the soil solution under certain conditions as well, which are highly pH dependent. While I agree that vegetation itself may not have changed too much (L457ff), but soil pH may still have changed. If you have any idea about soil pH values at the sites, and how that could have changed in the past, I suggest to think about this pathway as a potential additional source of P to the drip water. and potentially include it into the discussion as well.

We thank the reviewer for prompting us to broaden the discussion of pH variations from the perspective of P mobility. We propose to clarify in the manuscript that soil pH can affect P mobility.

We propose to add the following addition to the Introduction:

“Soil pH strongly influences the solubility and mobility of inorganic phosphorus by controlling both mineral dissolution and adsorption–desorption processes (Hinsinger, 2001). In acidic soils, phosphate is commonly bound to Fe- and Al-(oxyhydr)oxides, with adsorption maximized at low pH due to positively charged mineral surfaces; as pH increases, surface charge decreases, weakening adsorption and enhancing phosphate release. In alkaline or calcareous soils, phosphorus is often present as low-solubility calcium phosphate minerals such as apatite, whose dissolution is promoted as pH decreases from around 8 toward acidic values (Hinsinger, 2001). Microbial communities can accelerate these processes through localized acidification and the production of organic ligands that solubilize mineral-bound P, even under relatively stable bulk pH conditions (Pastore et al., 2022). Together, these mechanisms suggest that long-term or microbially mediated shifts in soil pH could mobilize mineral-derived P, providing an additional source of P to drip waters beyond organic matter contributions.”

We propose to add to the discussion (in section “4.2 Periods of high P flushing during stadial climate oscillations”):

“In addition to the direct climatic effects on freeze-thaw frequency and hydrology, the climatic transitions into stadial cold events likely also affected soil pH. More positive $\delta^{13}C_{init}$ values during the stadials indicate reduced soil respiration and lower soil CO_2 concentrations, which would raise the pH of soil water. Thus, the elevated P/Ca at the onset and termination of GS events could also reflect pH-driven mobilization of mineral-bound P, if intermediate pH conditions destabilized a fraction of the soil P pool. At the onset of the stadial, the positive pH shift may have enhanced the dissolution of Ca-phosphate minerals in calcareous soils and reduced P adsorption to Fe- and Al-(oxyhydr)oxides, thereby increasing the flux of dissolved P to drip waters. This pathway, acting alongside changes in organic P cycling, may therefore also contribute to the magnitude of the observed P/Ca excursions during these climate transitions.”

2. Sampling bias. I am not quite sure how the authors have taken into account that their stalagmites have varying growth rates, and how that might affect the observed P variability. Its hard to assess from the Figures also because age models have been published elsewhere, but it looks like some of the stalagmites change growth rate quite substantially, and that part of the increased variability in P/Ca could just be because of a sampling bias during periods of high growth rate? Therefore I suggest to down sample the stalagmite P/Ca records to equidistant , comparable resolution, to remove this effect.

The reviewer brings up a very interesting point - we appreciate the suggestion to first of all include clearer documentation of the sampling intervals and resolution and secondly evaluate yhe potential impact of variable growth rates and resolution.

In section 2.2, we now clarify the first order approach to account for varying growth rates.

We report data as the stalagmite P/Ca ratio. Because Ca is the major ion which is accounting for stalagmite growth, increases in P/Ca cannot be only due to higher growth rate because a higher growth rate also means a greater accumulation of Ca per unit time.

Nonetheless, we appreciate the reviewer's suggestion to include clearer documentation of the sampling intervals and resolution and describe how the differences in sample resolution may affect smoothing and amplitude of signals.

To address this, in Table 1 three additional columns will be added to provide the median growth rate, the sample spacing, and the median age difference between successive samples. In addition, the column now notes where sampling was continuous (ie all growth of the stalagmite was sampled by drilling and analyzed) vs where it reflects sampling of the stalagmite interspersed with unsampled growth durations.

In addition to the table, further text in the section 2.2 describes this aspect:

" Stalagmite samples were drilled using a Sherline drill which digitally monitors sampling position. Five of the presented stalagmite datasets are based on drilling resolution that continually sampled all of the deposited stalagmite along the growth axis, using drilling resolution of 0.25 to 1 mm per sample (Table 1). These include most of the presented records covering the PGM-LIG and one of the four records covering the GS22. For these samples, each drilled increment typically integrates 30 to 230 years (Table 1). The other presented records sampled 1 mm of powder but at sample increments ranging from 2 to 5 mm, meaning that only 20 to 50% of the growth conditions were sampled. The typical interval between samples reflects 150 to 400 years for these datasets (Table 1). During GS22, each drilled sample from GAE, NEI and ROW integrates 30 to 70 years based on the median growth rates. "

Age models and growth rates will be added to the Supplementary in Figure S1.

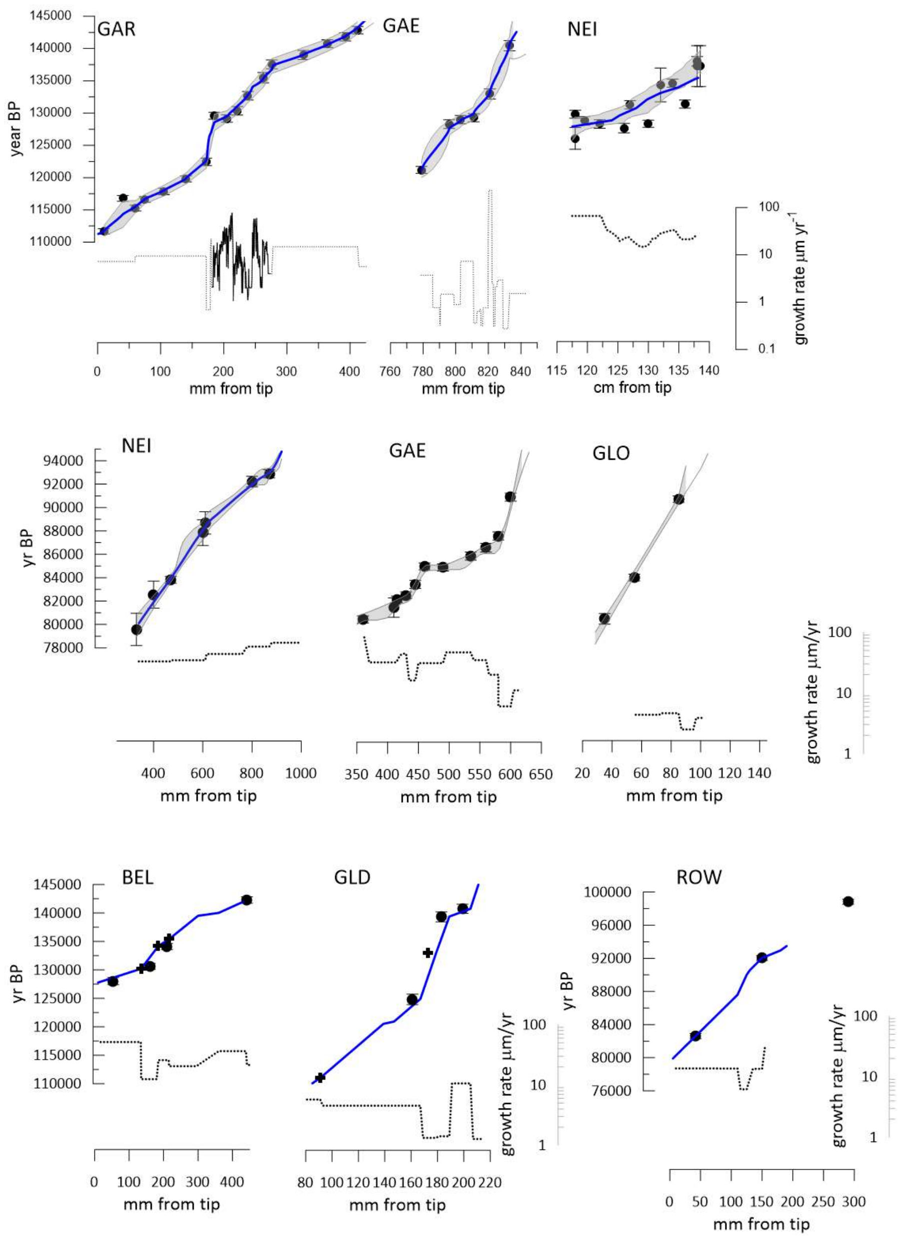

Figure S1. Illustration of age models. Blue lines indicate the employed age model. Gray shading illustrates the 95% confidence intervals for age models generated in BCHRON in top row, from (Stoll et al., 2022)) and in STALAGE in middle row, from (Stoll et al., 2015). Lower row illustrates age models from (Stoll et al., 2023). Growth rate estimated from the age models is shown in solid line for layer counted segments and dashed line for age models based on U/Th ages and tiepoints. Published U/Th dates are shown as circles with error on age assignments. For the lower row, tiepoints to isotope records in GAR (crosses) are also illustrated.

Additionally, the constraints on growth rates and age models will be reviewed at the end of Section 2.1

“ Age models are plotted in Figure S1. Median growth rates of the studied samples range from 4 to 36 µm·yr-1. Age models for GAR include sectors with layer counted growth rates, and in counted sectors layer thickness suggests growth rates varying by 5 to 10-fold on the decadal to centennial scale. In this sample, layer thickness decreases during cold stadial events, and annual layers become too thin for counting between 134.1 and 132.3 ka, as well as during the LIG (Stoll et al., 2022). Age models for other stalagmites are based on the interpolation schemes of BCHRON (Parnell & Parnell, 2018) and STALAGE(Scholz & Hoffmann, 2011) or linear interpolation when the density of dates is low. With the exception of GAR, the available chronological resolution does not permit us to confidently identify variations in growth rates between U/Th dates, and in most samples growth rate variations cannot be identified at timescales shorter than several thousand years (typical interval between dates Fig. S1). “

Because sample GAR is the only sample with highly resolved growth rates, we include an additional supplementary figure (SX) to illustrate the relationship between growth rate and P/Ca with both original and resampled data. This illustrates that the range in P/Ca is not due to the recovery of greater variability during periods of high growth rate.

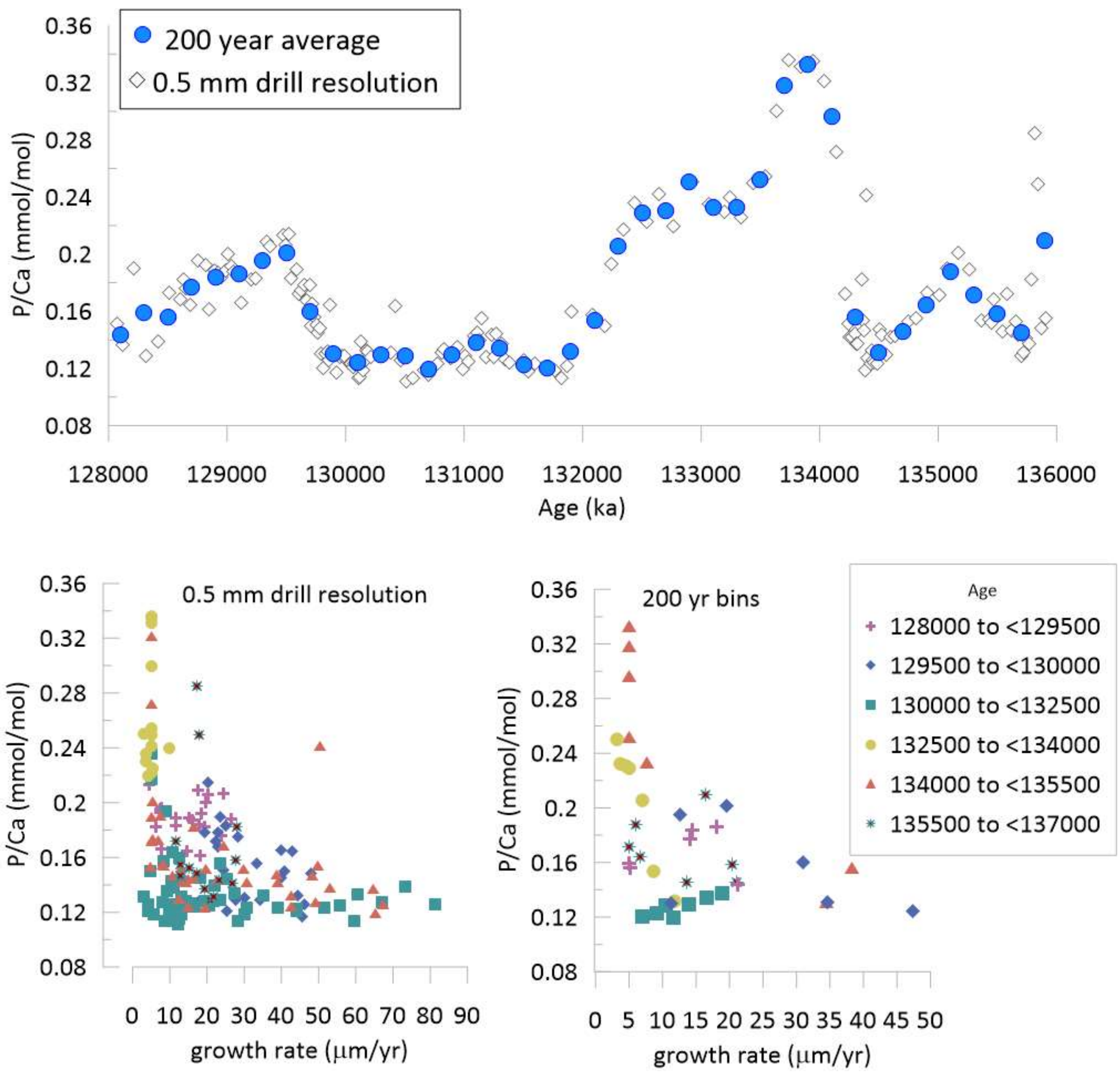

Figure S2. Upper panel shows, for the 136 to 128 ka interval with age model constrained by layer counting in GAR, the P/Ca for Garth for each drilled sample (open diamonds) compared with the average P/Ca in 200 yr age bins (blue circles). The consistency of trends indicates that the transient peaks in P/Ca are not artifacts of different durations integrated by the drill samples due to variable growth rates (Figure S1) and layer thicknesses. Lower panels show the correlations in P/Ca ratios of samples classified by age. Because growth rate slows at the onset of stadial intervals, higher P/Ca are clustered in intervals of slower growth, but slow growth periods also feature low P/Ca.

Because any variations in growth rate on timescales less than several ky are not detected in the other presented stalagmites, it is not possible to devise a robust strategy for downsampling to a fixed common temporal resolution such as 200 yr or 500 yr bins. Thus, for the remaining samples, we will use the newly added information in Table 1 to include relevant caveats on the potential for variation in the degree of smoothing among the different stalagmites recording the same event as a function of the mean growth rate and sample resolution.

Minor comments along the text

L35 and L49: I agree, but if I am not mistaken, none of these aspects is discussed further on of that may be part of the observed variability in the speleothems?

We have included this in the introduction because the organic and soil oxide pools of P are discussed in section 4.2. where we consider potential climatically-related processes which could lead to P loss from these pools at the onset of stadial events.

L116 repeated information – the vegetation description will be made more concise

L153 It would be good to include some numbers of how much soil water pH may change and if that could influence the solubility of other P pools.

Former Figure 7 (now Figure S4) illustrates a range of different initial (soil) water pH which is evaluated in the models. These models explicitly evaluate the evolution of pH in the cave by degassing and its effect on the speciation of P and the resulting effects on its incorporation in calcite.

Now the discussion 4.2 includes a comment on the potential for different initial soil water pH to affect the P mobility. The $\delta^{13}C_{init}$ indicator of changes in soil pCO2 (and pH) is useful to assess changes in soil pCO2 but not absolute values so it is not possible to "translate" this to particular soil pH in the past, only to discuss trends.

L210 In the distributions, is the temporal resolution due to different growth rates / sampling intervals taken into account? some peaks could have been missed/smoothed in slower growing stalagmites

In this section, the text is now revised to acknowledge the greater potential for smoothing in GAE:

The PGM-LIG section of GAE features numerous analyses with higher P/Ca than other samples, despite very slow growth and higher signal smoothing than other coeval stalagmites such as BEL and GAR.

Additionally, the potential for smoothing is also now discussed in section 3.2.

Fig. 2 What to the blue and red colors mean? I wonder how the box plots would change if one would take into account the variations in growth rate.

In the legend we will now clarify that red colors denote samples from Cueva Rosa and blue from La Vallina.

L234 so, synchronously to the shift in d13C?

Now clarify to read:"… increase in the P/Ca ratio synchronous with the shift in $\delta^{13}C_{init}$ marking the onset of the cooling event "

L236 How "low"? could you specify the magnitude of the shift also in d13C?

Propose to clarify to: "The significantly lower amplitude of the $\delta^{13}C_{init}$ anomaly in ROW (2‰ vs 4‰ in the other sampled stalagmites).

Fig. 4 would be nice to have similar lines indicating the "cooling" event(s)/spikes than in Fig 5

We agree, these will be added

L?? section 3.2.2: I again ask myself if some of the observations of high/low variability may have to do with growth rate/sampling bias. I suggest to clarify, and possibly show / discuss records down sampled to equidistant temporal resolution

To section 3.2.2, we will refer to the results from the comparison in GAR:

Comparison of drilled resolution and average P/Ca in 200 yr age bins confirms that these trends are not artifacts of changing sample resolution (Fig. S2).

Fig. 7 the nomenclature is a bit confusing, in the Fig its "models A…D", in the caption "panels A…D", and then it is mixed up with the D from partitioning coefficient… maybe use lower case letters or numbers for the panels?!

Good suggestion to change the cases so they have names which do not overlap with figure panels, and also avoid confusion with model D and the use of "D" as abbreviation for the partitioning coefficient. The figure has been moved to the supplement, following the reviewers suggestions, and we propose to rename models A-D to models 1 through 4.

L457 P export does not necessarily have to be related with a change of vegetation, but also a change in soil pH influencing the stability of the different P pools in the soil

We appreciate this suggestion and have added further discussion in section 4.2, as detailed in response to the general comments.