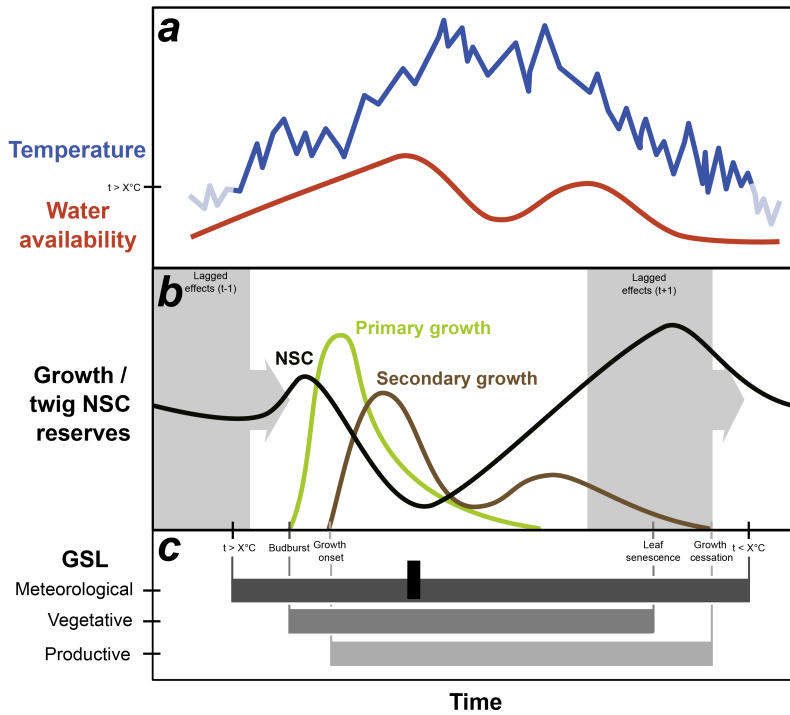**Box. Growth, growing season length and the challenge of standardized metrics**

A major challenge in determining how growth responds to longer growing seasons is the complexity of each, which means that neither can have one simple definition. Here we show the simplified climate of one year (a), which drives variation (b) in primary growth (root, shoot and leaf elongation) and secondary growth (radial wood growth), both of which often depend on growth from previous seasons. Each of these types of growth could define the growing season length (GSL, c) but similarly it can be defined meteorologically (e.g., days above 5°C with some level of soil moisture) or by large-scale measures of plant productivity (Körner *et al.*, 2023).

Of studies in our literature review, the largest proportion used metrics related to secondary growth, quantifying growth by measuring radial growth (e.g., through increment cores or dendrometers, $n = 28$), but a number also looked at metrics related to primary growth, including C assimilation (e.g. net ecosystem productivity or gross primary productivity, $n = 20$). A smaller number used metrics that reflect combined primary and secondary growth, including biomass, height, or number of stems ($n = 9$), and root:shoot ratio ($n = 1$). Some studies used modeled estimates of photosynthesis (e.g., Smith *et al.* (2014) relied on daily photosynthesis estimates derived from the LPJ-GUESS photosynthesis model, while Chen *et al.* (2000) estimated photosynthese using the Integrated Terrestrial Ecosystem C-budget model, InTEC). Others measured photosynthesis at the leaf level, through flux towers, or used greenness metrics (NDVI).

For growing season length, the largest number of studies used vegetative (e.g., budburst to leaf senescence in our figure above, 26 studies) or wood phenology 11 studies) as their definition, while a smaller number used a meteorological definitions or fixed dates (7 studies). We found 14 that did not directly measure GSL (e.g., Zhu *et al.*, 2021; Dow *et al.*, 2022; Zohner *et al.*, 2023). Further, these definitions of GSL are simplified and could not be easily aligned, as we found 14 different metrics of start of season, 16 metrics of end of season (25 metrics of growing season length. See also 'The challenge of metrics: Measuring growth and growing season length' in the Supplement.

Below from Supplement!

The challenge of metrics: Measuring growth and growing season length

Understanding the diverse drivers and testing underlying hypotheses (Fig. **??**) for growth × season length relationships requires a common language. We found 14 different metrics of start of season, 16 metrics of end of season (25 metrics of growing season length), and 21 different metrics of growth across 59 studies—highlighting just part of the problem (see also *Box: Growth, season length and the challenge of standardized metrics* in the main text). Definitions and metrics for external and internal drivers were myriad, with papers reporting dozens of tests of different aspects of climate over different temporal windows. Although this is understandable given the differing goals of these papers, it also slows progress.

A common framework for explanatory and response variables would accelerate research by easing communication between fields and providing a path to comparable quantitative estimates. This should also include expected statistical tests, as we found a number of papers failed to directly test for growth × growing season length relationships (Fig. **??**), often instead testing only certain hypothesized indirect relationships (e.g. spring temperature × growth in Dow *et al.*, 2022).

Measuring growth

Tree growth can be defined and measured in a variety of ways. It is often divided into primary (growth from the root and shoot tips that results in increased height and length) versus secondary (growth that increases the thickness or girth of stems and branches above ground, as well as below ground parts), or leaf production versus wood production. Growth measurements vary across disciplines and study types, posing further challenges to an interdisciplinary approach to understanding how growing season length relates to growth. Greenhouse or growth chamber studies and provenance trials were more likely to measure height or biomass, whereas larger scale syntheses and remote-sensed studies are more likely to use metrics of carbon assimilation. The inconsistency of current metrics across fields is not surprising, but may be confusing– including photosynthesis as a metric of growth may seem unconventional for ecophysiologists, for example,

who may interpret the process of photosynthesis as distinct from growth– and also stalls progress toward a unified model across disciplines.

Aligning across the range and scale of growth metrics will be critical for an integrated understanding of growth-growing season length relationships and implications under continued climate change. There is decoupling among some metrics of growth, as different types of growth (leaves, shoots, wood) have different phenologies within a season, and may vary across years, as well (e.g., with different lags, Fig. **??**). For example, vegetation photosynthesis may be poorly correlated with tree radial growth, and this relationship can vary seasonally (Cabon *et al.*, 2022). Further, tree radial growth is not a perfect indicator of whole tree growth, since plants allocate carbon to their roots, leaves, reproductive structures, and stores in addition to aboveground biomass. Relationships among different metrics of growth are not simple, so selecting relevant ones and aligning across the most widely used ones will be necessary, though not easy: the relationship between photosynthesis, radial growth, and carbon uptake has large implications for future carbon sequestration and it remains widely debated (Green & Keenan, 2022). Further, there is a need to understand how to scale up across these varying metrics- from leaf and individual level to populations, communities, and ecosystems- while incorporating the variation that exists within and across levels.

Measuring growing season length

Growing season length is defined in myriad ways due to varying definitions and metrics for the start of season and the end of season. For example, the end of season could be defined as budset, representing the end of vegetative phenology, or as xylogensis, representing the end of wood phenology. Additionally, several studies leveraged climate data as a proxy for start or end of season metrics, such as snowmelt for start of season or growing season GDDs. We therefore grouped all reported growing season length metrics into four separate categories: 1) 'Climate and date' simulated by climate data, 2) 'Vegetative phenology' through field observations or satellite derived measurements, 3) 'Wood phenology' measured by dendrometers and xylogenesis, and 4) not measured, which includes all studies that only measured start or end of seasons but not both.

Numerous studies in the review reported multiple metrics for end of season (e.g., budset versus leaf coloring) and found differing relationships between growth and growing season length depending on the end of season definition. We reported all combinations of assessments per study in our review.

Extending disciplinary focus to help bridge the internal-external drivers divide

Each field studying growth × growing season length today has its own historical aims, and thus, its own biases towards certain species, methods and metrics. For example, dendrochronology's original focus on using tree growth to estimate climate has led to sampling biases (e.g. to 'climate-sensitive' individual trees, Klesse *et al.*, 2018; Nehrbass-Ahles *et al.*, 2014) and statistical detrending (Rollinson *et al.*, 2021), which may obscure patterns where the signal of longer growing seasons and biotic drivers may be most apparent (such as rapid growth phases). Dendrochronology also generally focuses on conifers (gymnosperms, Zhao *et al.*, 2019), creating a major split from most studies of leaf phenology, which focus almost entirely on deciduous angiosperm species (see Fig. **??**). By contrast, phenology research has been strongly focused on spring events (e.g. budburst, leafout), with limited data on fall events and thus limited data to calculate growing season length. This focus on spring events may have been justified decades ago, when most shifts from anthropogenic warming occurred in the spring, but less justified as increasing research suggests important complexity in fall shifts (Gill *et al.*, 2015; Zohner *et al.*, 2023) and a need to scale up phenological research to understand tree growth.

1 References

Cabon, A., Kannenberg, S.A., Arain, A., Babst, F., Baldocchi, D., Belmecheri, S., Delpierre, N., Guerrieri, R., Maxwell, J.T., McKenzie, S. *et al.* (2022) Cross-biome synthesis of source versus sink limits to tree growth. *Science* **376**, 758–761.

Chen, W., Chen, J., Liu, J. & Cihlar, J. (2000) Approaches for reducing uncertainties in regional forest carbon balance. *Global Biogeochemical Cycles* **14**, 827–838.

Dow, C., Kim, A.Y., D'Orangeville, L., Gonzalez-Akre, E.B., Helcoski, R., Herrmann, V., Harley, G.L., Maxwell, J.T., McGregor, I.R., McShea, W.J. *et al.* (2022) Warm springs alter timing but not total growth of temperate deciduous trees. *Nature* **608**, 552–557.

Gill, A.L., Gallinat, A.S., Sanders-DeMott, R., Rigden, A.J., Gianotti, D.J.S., Mantooth, J.A. & Templer, P.H. (2015) Changes in autumn senescence in northern hemisphere deciduous trees: a meta-analysis of autumn phenology studies. *Annals of Botany* **116**, 875–888.

Green, J.K. & Keenan, T.F. (2022) The limits of forest carbon sequestration. *Science* **376**, 692–693.

Klesse, S., DeRose, R.J., Guiterman, C.H., Lynch, A.M., O'Connor, C.D., Shaw, J.D. & Evans, M.E. (2018) Sampling bias overestimates climate change impacts on forest growth in the southwestern united states. *Nature communications* **9**, 5336.

Körner, C., Möhl, P. & Hiltbrunner, E. (2023) Four ways to define the growing season. *Ecology Letters* .

Nehrbass-Ahles, C., Babst, F., Klesse, S., Nötzli, M., Bouriaud, O., Neukom, R., Dobbertin, M. & Frank, D. (2014) The influence of sampling design on tree-ring-based quantification of forest growth. *Global change biology* **20**, 2867–2885.

Rollinson, C.R., Alexander, M.R., Dye, A.W., Moore, D.J., Pederson, N. & Trouet, V. (2021) Climate sensitivity of understory trees differs from overstory trees in temperate mesic forests. *Ecology* **102**, e03264.

Smith, B., Wårlind, D., Arneth, A., Hickler, T., Leadley, P., Siltberg, J. & Zaehle, S. (2014) Implications of incorporating n cycling and n limitations on primary production in an individual-based dynamic vegetation model. *Biogeosciences* **11**, 2027–2054.

Zhao, S., Pederson, N., D'Orangeville, L., HilleRisLambers, J., Boose, E., Penone, C., Bauer, B., Jiang, Y. & Manzanedo, R.D. (2019) The international tree-ring data bank (itrdb) revisited: data availability and global ecological representativity. *Journal of Biogeography* **46**, 355–368.

Zhu, L., Liu, S., Arzac, A., Cooper, D.J., Jin, Y., Yuan, D., Zhu, Y., Zhang, X., Li, Z., Zhang, Y. *et al.* (2021) Different response of earlywood vessel features of fraxinus mandshurica to rapid warming in warm-dry and cold-wet areas. *Agricultural and Forest Meteorology* **307**, 108523.

Zohner, C.M., Mirzagholi, L., Renner, S.S., Mo, L., Rebindaine, D., Bucher, R., Palouš, D., Vitasse, Y., Fu, Y.H., Stocker, B.D. *et al.* (2023) Effect of climate warming on the timing of autumn leaf senescence reverses after the summer solstice. *Science* **381**, eadf5098.