

## Global patterns of geo-ecological controls on the response of soil respiration to warming

David Haaf, Johan Six, Sebastian Doetterl

### Angaben zur Veröffentlichung / Publication details:

Haaf, David, Johan Six, and Sebastian Doetterl. 2021. "Global patterns of geo-ecological controls on the response of soil respiration to warming." Research Square.  
<https://doi.org/10.21203/rs.3.rs-147557/v1>.

### Nutzungsbedingungen / Terms of use:

CC BY 4.0

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:  
**CC-BY 4.0: Creative Commons: Namensnennung**  
Weitere Informationen finden Sie unter: / For more information see:  
<https://creativecommons.org/licenses/by/4.0/deed.de>



# Global patterns of geo-ecological controls on the response of soil respiration to warming

David Haaf<sup>1</sup>, Johan Six<sup>1</sup>, Sebastian Doetterl<sup>1</sup>

<sup>1</sup> ETH Zurich, Department of Environmental Systems Science, Zurich Switzerland

Corresponding author: Sebastian Doetterl. [sdoetterl@usys.ethz.ch](mailto:sdoetterl@usys.ethz.ch)

## Abstract

While soil respiration is known to be controlled by a large range of biotic and abiotic factors, its temperature sensitivity in global models is largely related to climate parameters. Here, we show that temperature sensitivity of soil respiration is primarily controlled by interacting soil properties and only secondarily by vegetation traits and plant growth conditions. Temperature was not identified as a primary driver for the response of soil respiration to warming. In contrast, the non-linearity and large spatial variability of identified controls stress the importance of the interplay among soil, vegetation and climate parameters in controlling warming responses. Global models might well predict current soil respiration, but not future rates because they neglect the controls exerted by soil development. Thus, to accurately predict the response of soil respiration to warming at the global scale, more observational studies across pedogenetically diverse soils are needed rather than focusing on the isolated effect of warming alone.

## Background

With implemented climate policies struggling to limit global warming to an average of less than 1.5 °C<sup>1</sup>, elucidating the response of an adapting ecosphere to warming is more and more

important. Understanding soil C dynamics is key to this because it directly determines a large portion of future net greenhouse gas (GHG) emissions from terrestrial ecosystems<sup>2</sup>. Soils are considered net sinks for C with current net sequestration estimated at 1 Pg C yr<sup>-1</sup><sup>3</sup>. This is only a minor part of the continuous exchange of C between soil and atmosphere due to C input to soils through plants and release of C through soil respiration, approximately balanced at annual fluxes of 58 - 80 Pg C yr<sup>-1</sup><sup>4-6</sup>. Rising global temperatures are expected to lead to significantly higher decomposition rates of soil C and thus CO<sub>2</sub> release from soils<sup>7,8</sup>, largely because of more energy available for microbial decomposer communities<sup>9</sup>. However, despite its importance, the response of soil C to warming is still one of the great uncertainties in global carbon cycling<sup>10</sup>. Great uncertainties are related to the effect of warming on vegetation, C input<sup>11</sup> across different soil depths<sup>12</sup>, microbial responses<sup>13</sup>, and estimates for losses of soil C the arctic plus high latitudes<sup>14</sup> and tropical plus low latitudes<sup>15</sup>. While the temperature sensitivity of soil carbon has been long studied<sup>10,16</sup>, only now ecosystem models begin to implement mechanistic controls of microbial soil respiration in response to climate and soil changes<sup>17,18</sup>. One issue is that soil properties, often crucially related to subsoils, are hidden from air and space borne sensing techniques that do not “see” soils. Therefore, statistical models are needed to better represent relationships between microscopic and macroscopic processes, especially on broader scales<sup>19,20</sup>. Furthermore, most of our mechanistic understanding of soil processes and warming is derived from studies in temperate zones; their numbers simply dwarf the number of studies in boreal and tropical ecosystems (see Figures S1 and S2). Due to the nature of small-scale studies with often homogenous soil and environmental properties, a holistic, global assessment on factors controlling soil respiration, except for basic variables that integrate various processes at once (i.e. clay content) has not been done yet<sup>16</sup>. Soil is not mechanistically represented in global ecosystem models, but is rather given a mostly budgetary function. Thus, future global soil greenhouse gas emissions might be critically

misrepresented under changing environmental conditions. For example, global climate and ecosystem models<sup>21,22</sup> dealing with warming focus on GHG fluxes from environments where climatic and hydrological barriers are the key controls to limit C decomposition<sup>23</sup>. However, these climate and hydrology driven, geochemically speaking “young”, soil systems do not represent soil conditions found for the largest part of globally relevant soil C stocks<sup>24</sup>. The majority of soil C is stored in geochemically more complex and weathered soil systems, where soils have developed over millennia and the biosphere adapted to warmer conditions over millions of years of evolution<sup>25</sup>. Hence, soils in every (geo-)climatic zone will likely show very different responses in respirations to warming due to their different, soil type dependent, properties and drivers<sup>26</sup>. To the best of our knowledge, previous models of soil  $Q_{10}$  took the average air temperature as main predictor for soil  $Q_{10}$ .<sup>27-29</sup> Thus, the global representation of soils and GHG emissions from them with their drivers and controls are not well represented in earth system models (ESMs) and  $Q_{10}$  is still treated as an average value over all climate zones and state-of-the-art in CMIP5 models to consider temperature sensitivity in soil<sup>29-32</sup>. By using highly averaged values of temperature sensitivity of soil C<sup>28,33-36</sup> that do not represent the underlying processes<sup>16</sup>, or by focusing on selected climatic drivers, current earth system and climate models unintentionally neglect the variability of crucial biogeochemical factors altering the response of soils to climate forcing<sup>37</sup>. Doing so introduces large biases and uncertainties in global estimates of future C emissions from soils.

Here, we brought together large and small-scale controls that have been identified as key variables to explain the soil respiration response to warming - expressed as soil  $Q_{10}$  - at the global scale and used machine learning techniques to identify the most important groups of explaining variables for soil  $Q_{10}$ . More specifically, we combined experimental results with a large database on climate, vegetation and soil related parameters (further called *best data* approach) as proxies of soil respiration influencing factors under warming<sup>38,39</sup> (Table S3).

While  $Q_{10}$  of soil respiration is not a mechanistic depiction of soil C response, it can be interpreted as a phenomenological response of multiple instantaneous processes that differ across geo-climatic and land use settings<sup>38,39</sup> and is widely used in global scale ecosystem models. We compiled 3400 observations from 560 soil warming studies conducted from 1971 to 2018 with incubation lengths of several days to more than three years from all major climate and land use combinations (see methods & Figure S2). For our analyses, we concentrated on climate zones in which rich plant-soil interactions occur and excluded regions with bare soils (polar and non-polar (semi-)deserts and high alpine environments) for which not enough data to train models and/or global maps of independent predictors were available. Then, we (i) built linear and non-linear predictive models for soil  $Q_{10}$ , (ii) derived the relative importance of the derived groups of explaining variables for SRRW and (iii) determined the changing importance of the identified controls in different climate systems and land use zones using partial dependence analyses (Figures 1 & 2). To assess the validity of our interpretation and the robustness of our models, we have repeated i-iii by using only predictors of SRRW derived from global datasets, further referred to as the *generalized data* approach (Table S4 & Table S6).

## Results & Discussion

### *Predicting soil $Q_{10}$ and its controls*

Our model satisfactorily predicted soil  $Q_{10}$  across all included systems for both the *best data* and the *generalized data* approach (Figures 1a & S3a), showing to the best of our knowledge, for the first time how the temperature sensitivity of heterotrophic soil respiration is driven by a combination of soil properties, vegetation and climate interactions at the global scale. Similarly to previous assessments of soil  $Q_{10}$  at the regional scale<sup>40</sup>, non-linear model approaches ( $R^2 = 0.18 - 0.46$ ; RMSE 0.58 - 0.72) greatly outperformed linear models ( $R^2 = 0.07 - 0.08$ ; RMSE 0.76 - 0.77) (Table S6). Both the *best data* and *generalized data* model

approaches performed similarly in explaining the variability in temperature sensitivity of soil respiration ( $R^2 = 0.46$ ) and with reasonable uncertainty (relative RMSE = 24 %). Only a relative small part of soil  $Q_{10}$  was directly controlled by plant growth conditions (11.6%) as well as evapotranspiration and precipitation (12.6%). In contrast, a much larger share of soil  $Q_{10}$  variability was controlled by soil properties (63.1%) (Figure 1b). Interestingly, climate and vegetation variables were more intercorrelated and their effects on soil  $Q_{10}$  were not clearly separable (Table S3).

[Figure 1 about here]

#### *Global patterns of controls on soil $Q_{10}$*

Our analyses also revealed an extremely high variability in the controlling factors for soil respiration (Figure 1c). Vegetation and climate related parameters like growth conditions and evapotranspiration had a strong influence at both extreme ends of their respective range of values, which represent climatic extremes; As a general trend, climate was a strong control at lower temperatures, low precipitation or higher evaporation (Figures 1c, 2). This is likely related to the lack of mineral stabilization of C in these colder climate zones<sup>41</sup> leading to a faster response of microorganisms to warming, and hence a decomposition of labile C once temperature barriers are released<sup>42</sup>. Notably, temperature was not a separate dominant control on SRRW and climatic variables in general exert little influence in environments with more moderate climate; moreover, temperature ceases to influence soil  $Q_{10}$  in warmer climate zones. In contrast, a wide range of biotic and abiotic soil variables controlled the variability of soil  $Q_{10}$  across their full range of values, resulting in the observed high heterogeneity. This dominance of soil variables is most likely because of the variety of parent materials that soils develop from and the various stages of weathering across the globe that affect plant growth and C stabilization. In cold climates, soils show low reactivity due to climatic barriers to chemical

soil weathering<sup>16</sup>. Plant litter, and not microbially processed or mineral associated C, is often the main source of energy for microorganisms under these cold conditions. In temperate climates, soils have generally higher chemical reactivity and high C stabilization potential, thereby diversifying potential C sources for microorganisms. This diversification of energy sources can lead to very variable competitive strategies driving carbon use efficiency<sup>43</sup> and thus soil  $Q_{10}$ <sup>42</sup>. In tropical climates, chemical weathering has depleted many soils of reactive minerals and reduced C stabilization potential, leading to a reduction in the variety of C resources.

Hence, strategies for an efficient recycling of nutrients from litter back into plants are prevailing<sup>26,44</sup>. The implementation of all identified controls in our model resulted into a spatially highly variable map of soil  $Q_{10}$  (Figure 2a,b) and a similarly diverse map of relative uncertainty of prediction (Figure 2c,d). More specifically, in arctic and boreal environments, where temperature is a major barrier for decomposition of labile C, soil  $Q_{10}$  was particularly high across all major land use systems. In contrast, soil  $Q_{10}$  was highly variable in temperate zones where local soil development drives C stabilization and thus responsiveness to warming. Lastly, soil  $Q_{10}$  was generally lowest in tropical environments where soils are deeply weathered and C accessibility is driven by litter quality. Deviations from this general pattern were tied to local variations in climatic, topographic and biogeochemical soil conditions (Table S9). Our uncertainty map (Figure 2c,d) shows high spatial variability especially in data poor regions of the (sub)tropics or in regions with highly diverse soil landscapes (temperate and tropical zones). We explain this with the fact that in data poor regions the model cannot be trained to the same degree as in data-rich regions due to a lack of data and precision for both response and independent variables. In regions of highly developed soils, our results point at the importance of considering local soil development and land use history for predicting SRRW, because these can differ greatly from one geo-climatic region to the next leading to varying

model complexity and strength of predictors (see figure 1) that is not fully captured at the global scale. In summary, our analyses allowed for predicting global patterns of soil Q<sub>10</sub> with reasonable uncertainty at a much higher accuracy and spatial variability than comparable approaches using climatic and vegetation variables alone<sup>27,28,40,45</sup> across major climate zones in which forests, grasslands and agricultural land use appears. Nevertheless, a larger share of variability in soil Q<sub>10</sub> remained unexplained (about 55%). We relate this lack of identifiability to the coarse spatial and temporal resolution of global key datasets, where information on local heterogeneity is lost, paired with a lack of accurate data from data poor regions (i.e. mountains, boreal zones, wetlands, tropics). Furthermore, global studies and predictions are in parts driven by completely different parameters than comparable regional studies, due to the different resolution and data availability<sup>46</sup>. A large number of local to regional scale controls on soil Q<sub>10</sub> and microbial decomposition processes exist (i.e. land management) that cannot be represented currently through proxy variables at the global scale<sup>39</sup>.

[Figure 2 about here]

#### *CO<sub>2</sub> release from soils in the decades to come*

Our study showed much higher and more variable temperature sensitivity of respiration than comparable ecosystem-level assessments<sup>27</sup>. Soil Q<sub>10</sub> predicted by our model was on average  $33 \pm 10$  % higher than compared to soil Q<sub>10</sub> in climate driven models<sup>28</sup>. Our results are consistent with, and can help explain, the predicted reduced uptake of C in soils by the end of the 21<sup>st</sup> century<sup>47,48</sup>. As has been demonstrated before<sup>49</sup>, boreal and temperate climate zones of the northern hemisphere showed increased C release from soils with changing temperature and precipitation while soils of the southern hemisphere showed only limited responses and tropical soils even less. However, based on our results, we would predict that in colder environments, warming will create over time a more reactive soil matrix, similar to those found in temperate climates. Examples for the expected changes in arctic soils are for example, higher rock-



derived nutrient release due to (bio-) chemical weathering, higher potential to stabilize carbon with minerals, thicker soils for higher water retention capacity and larger rooting zones<sup>50-52</sup>. It is thus likely that in many of these changed future soils of arctic, antarctic or alpine environments, plant productivity will increase, C stabilization through various mineral related physico-chemical mechanisms<sup>53</sup> will improve and microbial communities will respond to the changed climatic conditions with, for example, higher carbon use efficiency<sup>43</sup>. Greening and weathering are likely to compensate some of the projected soil C loss from thawing and regressing permafrost<sup>54</sup> losses through additional C sequestration and create new terrestrial C sinks in higher latitudes. However, recent studies show<sup>36</sup> that it is unreasonable to assume that these processes can fully compensate for the additional release of C from soils. Plant growth is limited by more than atmospheric parameters, and weathering leading to nutrient release or C stabilization potential is slow and on decadal timescales<sup>55</sup>. Warming in the next decades could lead to an additional C release from soil that is equal to all other current anthropogenic C emissions.

A warming climate, however, will ultimately lead to lower SRRW in boreal zones in the long term, as plant-soil systems become more adapted to warming<sup>56</sup> with arctic soil systems becoming more similar to boreal or even temperate systems if climate change is progressing as predicted<sup>57</sup>. Predicting these contrasting trends of soil  $Q_{10}$  in changed soil landscapes requires earth-system models to incorporate soil development trajectories as a control for future C fluxes and account correctly for the carbon flux between soil and atmosphere<sup>58</sup>. Indeed, in order to estimate C fluxes further into the future, a more mechanistic approach is needed that includes processes like soil formation (i.e. accelerated soil formation in arctic due to warming and increased weathering) or soil degradation (i.e. in the tropics due to land use change and erosion) to accurately predict the future warming response of these dynamic systems.

**Take home message - A call for action**

Our results illustrate how complex the interplay and strengths of controlling factors for soil Q<sub>10</sub> can be at global scales. First, using a large range of independent variables to predict soil Q<sub>10</sub> in heterogeneous ecosystems, we confirm that controls on soil C responses to climate change are drastically different between climate zones and environmental settings, limiting the transferability of experimental and mechanistic knowledge on soil processes across geoclimatic zones. Second, almost all variables showed spatially varying influence on soil Q<sub>10</sub>, meaning that soil Q<sub>10</sub> is highly non-linear and multifactorial. Lastly, from poles to the equator, temperature has not been identified as the main driving factor for soil Q<sub>10</sub>. While temperature was certainly a limiting and controlling factor for biological activity in high latitudinal environments, soil Q<sub>10</sub> was increasingly stronger related to biogeochemical and physical soil conditions than to warming *per se* in mid and lower latitudes. Thus, large changes to the soil C cycle will occur through a warming induced feedback loop that is more strongly controlled by changing soil parameters and development due to better conditions for chemical weathering than by temperature itself. Our study, focusing on soil development related variables shows which key controls have to be considered in ESMs besides warming to understand and predict a changing terrestrial C sink versus source by the end of the 21<sup>st</sup> century. Lastly, improving our mechanistic understanding of the effects of developing soil characteristics in different climate zones and ecosystems, especially in tropical regions, is required before soil respiration responses to warming can be accurately projected into the future.

## References

- 1 Rogelj, J. *et al.* Paris Agreement climate proposals need a boost to keep warming well below 2 C. *Nature* **534**, 631 (2016).
- 2 Song, J. *et al.* A meta-analysis of 1,119 manipulative experiments on terrestrial carbon-cycling responses to global change. *Nat Ecol Evol*, doi:10.1038/s41559-019-0958-3 (2019).
- 3 Lal, R. Soil carbon sequestration impacts on global climate change and food security. *science* **304**, 1623-1627 (2004).
- 4 Houghton, R. A. The contemporary carbon cycle. *Treatise on geochemistry* **8**, 473-513 (2003).
- 5 Paterson, E., Midwood, A. J. & Millard, P. Through the eye of the needle: a review of isotope approaches to quantify microbial processes mediating soil carbon balance. *New Phytologist* **184**, 19-33 (2009).
- 6 Bader, M. K. F. & Körner, C. No overall stimulation of soil respiration under mature deciduous forest trees after 7 years of CO<sub>2</sub> enrichment. *Global Change Biology* **16**, 2830-2843 (2010).

- 7 Reynolds, L. L., Lajtha, K., Bowden, R. D., Johnson, B. R. & Bridgham, S. D. The carbon quality-temperature hypothesis does not consistently predict temperature sensitivity of soil organic matter mineralization in soils from two manipulative ecosystem experiments. *Biogeochemistry* **136**, 249-260 (2017).
- 8 Knorr, W., Prentice, I. C., House, J. & Holland, E. Long-term sensitivity of soil carbon turnover to warming. *Nature* **433**, 298 (2005).
- 9 Allison, S. D., Wallenstein, M. D. & Bradford, M. A. Soil-carbon response to warming dependent on microbial physiology. *Nature Geoscience* **3**, 336 (2010).
- 10 Kirschbaum, M. U. F. The temperature dependence of organic-matter decomposition—still a topic of debate. *Soil Biology and Biochemistry* **38**, 2510-2518 (2006).
- 11 Feng, X., Simpson, A. J., Wilson, K. P., Williams, D. D. & Simpson, M. J. Increased cuticular carbon sequestration and lignin oxidation in response to soil warming. *Nature Geoscience* **1**, 836-839 (2008).
- 12 Pries, C. E. H., Castanha, C., Porras, R. & Torn, M. The whole-soil carbon flux in response to warming. *Science* **355**, 1420-1423 (2017).
- 13 Li, J. *et al.* Reduced carbon use efficiency and increased microbial turnover with soil warming. *Global change biology* **25**, 900-910 (2019).
- 14 Schaphoff, S. *et al.* Contribution of permafrost soils to the global carbon budget. *Environmental Research Letters* **8**, 014026 (2013).
- 15 Nottingham, A. T., Meir, P., Velasquez, E. & Turner, B. L. Soil carbon loss by experimental warming in a tropical forest. *Nature* **584**, 234-237 (2020).
- 16 Davidson, E. A. & Janssens, I. A. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature* **440**, 165-173, doi:10.1038/nature04514 (2006).
- 17 Koven, C. D. *et al.* The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4. *Biogeosciences* **10**, 7109-7131, doi:10.5194/bg-10-7109-2013 (2013).
- 18 Sulman, B. N., Phillips, R. P., Oishi, A. C., Shevliakova, E. & Pacala, S. W. Microbe-driven turnover offsets mineral-mediated storage of soil carbon under elevated CO<sub>2</sub>. *Nature Climate Change* **4**, 1099-1102, doi:10.1038/nclimate2436 (2014).
- 19 Schmidt, M. W. *et al.* Persistence of soil organic matter as an ecosystem property. *Nature* **478**, 49-56, doi:10.1038/nature10386 (2011).
- 20 Wieder, W. R. *et al.* Explicitly representing soil microbial processes in Earth system models. *Global Biogeochemical Cycles* **29**, 1782-1800, doi:10.1002/2015gb005188 (2015).
- 21 Gonzalez-Dominguez, B. *et al.* Temperature and moisture are minor drivers of regional-scale soil organic carbon dynamics. *Sci Rep* **9**, 6422, doi:10.1038/s41598-019-42629-5 (2019).
- 22 Blankinship, J. C. *et al.* Improving understanding of soil organic matter dynamics by triangulating theories, measurements, and models. *Biogeochemistry* **140**, 1-13, doi:10.1007/s10533-018-0478-2 (2018).
- 23 Koven, C. D. *et al.* Permafrost carbon-climate feedbacks accelerate global warming. *Proceedings of the National Academy of Sciences* **108**, 14769-14774 (2011).
- 24 Angst, G. *et al.* Soil organic carbon stocks in topsoil and subsoil controlled by parent material, carbon input in the rhizosphere, and microbial-derived compounds. *Soil Biology and Biochemistry* **122**, 19-30, doi:10.1016/j.soilbio.2018.03.026 (2018).
- 25 Abramoff, R. *et al.* The Millennial model: in search of measurable pools and transformations for modeling soil carbon in the new century. *Biogeochemistry* **137**, 51-71, doi:10.1007/s10533-017-0409-7 (2017).
- 26 Doetterl, S. *et al.* Links among warming, carbon and microbial dynamics mediated by soil mineral weathering. *Nature Geoscience* **11**, 589-593, doi:10.1038/s41561-018-0168-7 (2018).
- 27 Hamdi, S., Moyano, F., Sall, S., Bernoux, M. & Chevallier, T. Synthesis analysis of the temperature sensitivity of soil respiration from laboratory studies in relation to incubation methods and soil conditions. *Soil Biology and Biochemistry* **58**, 115-126, doi:10.1016/j.soilbio.2012.11.012 (2013).
- 28 Hashimoto, S. *et al.* Global spatiotemporal distribution of soil respiration modeled using a global database. *Biogeosciences* **12**, 4121-4132, doi:10.5194/bg-12-4121-2015 (2015).
- 29 Varney, R. M. *et al.* A spatial emergent constraint on the sensitivity of soil carbon turnover to global warming. *Nat Commun* **11**, 5544, doi:10.1038/s41467-020-19208-8 (2020).
- 30 Wu, D., Piao, S., Liu, Y., Ciais, P. & Yao, Y. Evaluation of CMIP5 Earth System Models for the Spatial Patterns of Biomass and Soil Carbon Turnover Times and Their Linkage with Climate. *Journal of Climate* **31**, 5947-5960, doi:10.1175/jcli-d-17-0380.1 (2018).
- 31 Wieder, W. R. *et al.* Carbon cycle confidence and uncertainty: Exploring variation among soil biogeochemical models. *Glob Chang Biol* **24**, 1563-1579, doi:10.1111/gcb.13979 (2018).
- 32 Koven, C. D., Hugelius, G., Lawrence, D. M. & Wieder, W. R. Higher climatological temperature sensitivity of soil carbon in cold than warm climates. *Nature Climate Change* **7**, 817-822, doi:10.1038/nclimate3421 (2017).
- 33 Mahecha, M. D. *et al.* Global Convergence in the Temperature Sensitivity of Respiration at Ecosystem Level. *Science* **329**, 838-840, doi:10.1126/science.1189587 (2010).
- 34 Foereid, B., Ward, D., Mahowald, N., Paterson, E. & Lehmann, J. The sensitivity of carbon turnover in the Community Land Model to modified assumptions about soil processes. *Earth System Dynamics* **5**, 211-221 (2014).
- 35 Friedlingstein, P. *et al.* Climate-carbon cycle feedback analysis: results from the C4MIP model intercomparison. *Journal of climate* **19**, 3337-3353 (2006).
- 36 Post, H., Vrugt, J. A., Fox, A., Vereecken, H. & Hendricks Franssen, H. J. Estimation of Community Land Model parameters for an improved assessment of net carbon fluxes at European sites. *Journal of Geophysical Research: Biogeosciences* **122**, 661-689 (2017).

- 37 Luo, Y. *et al.* Toward more realistic projections of soil carbon dynamics by Earth system models. *Global Biogeochemical Cycles* **30**, 40-56, doi:10.1002/2015gb005239 (2016).
- 38 Bailey, V. L. *et al.* Soil carbon cycling proxies: Understanding their critical role in predicting climate change feedbacks. *Glob Chang Biol* **24**, 895-905, doi:10.1111/gcb.13926 (2018).
- 39 Conant, R. T. *et al.* Temperature and soil organic matter decomposition rates - synthesis of current knowledge and a way forward. *Global Change Biology* **17**, 3392-3404, doi:10.1111/j.1365-2486.2011.02496.x (2011).
- 40 Meyer, N., Welp, G. & Amelung, W. The Temperature Sensitivity (Q10) of Soil Respiration: Controlling Factors and Spatial Prediction at Regional Scale Based on Environmental Soil Classes. *Global Biogeochemical Cycles* **32**, 306-323, doi:10.1002/2017gb005644 (2018).
- 41 Doetterl, S. *et al.* Soil carbon storage controlled by interactions between geochemistry and climate. *Nature Geoscience* **8**, 780-783, doi:10.1038/ngeo2516 (2015).
- 42 Melillo, J. M. *et al.* Long-term pattern and magnitude of soil carbon feedback to the climate system in a warming world. *Science* **358**, 101-105 (2017).
- 43 Kramer, M. G. & Chadwick, O. A. Climate-driven thresholds in reactive mineral retention of soil carbon at the global scale. *Nature Climate Change* **8**, 1104-1108, doi:10.1038/s41558-018-0341-4 (2018).
- 44 Cusack, D. F. *et al.* Decadal-scale litter manipulation alters the biochemical and physical character of tropical forest soil carbon. *Soil Biology and Biochemistry* **124**, 199-209 (2018).
- 45 Wang, X. *et al.* Are ecological gradients in seasonal Q10 of soil respiration explained by climate or by vegetation seasonality? *Soil Biology and Biochemistry* **42**, 1728-1734, doi:10.1016/j.soilbio.2010.06.008 (2010).
- 46 Warner, D. L., Bond-Lamberty, B., Jian, J., Stell, E. & Vargas, R. Spatial Predictions and Associated Uncertainty of Annual Soil Respiration at the Global Scale. *Global Biogeochemical Cycles* **33**, 1733-1745, doi:10.1029/2019gb006264 (2019).
- 47 Todd-Brown, K., Zheng, B. & Crowther, T. W. Field-warmed soil carbon changes imply high 21st-century modeling uncertainty. *Biogeosciences* **15**, 3659-3671, doi:10.5194/bg-15-3659-2018 (2018).
- 48 He, Y. *et al.* Radiocarbon constraints imply reduced carbon uptake by soils during the 21st century. *Science* **353**, 1419-1424 (2016).
- 49 Haddix, M. L. *et al.* The Role of Soil Characteristics on Temperature Sensitivity of Soil Organic Matter. *Soil Science Society of America Journal* **75**, 56-68, doi:10.2136/sssaj2010.0118 (2011).
- 50 Lara, M. J., Lin, D. H., Andresen, C., Lougheed, V. L. & Tweedie, C. E. Nutrient Release From Permafrost Thaw Enhances CH4 Emissions From Arctic Tundra Wetlands. *Journal of Geophysical Research: Biogeosciences* **124**, 1560-1573 (2019).
- 51 Prater, I. *et al.* From fibrous plant residues to mineral-associated organic carbon—the fate of organic matter in Arctic permafrost soils. *Biogeosciences* **17**, 3367-3383 (2020).
- 52 Åkerman, H. J. & Johansson, M. Thawing permafrost and thicker active layers in sub-arctic Sweden. *Permafrost and periglacial processes* **19**, 279-292 (2008).
- 53 Jilling, A. *et al.* Minerals in the rhizosphere: overlooked mediators of soil nitrogen availability to plants and microbes. *Biogeochemistry* **139**, 103-122, doi:10.1007/s10533-018-0459-5 (2018).
- 54 Jones, M. C. *et al.* Rapid carbon loss and slow recovery following permafrost thaw in boreal peatlands. *Global change biology* **23**, 1109-1127 (2017).
- 55 Korell, L., Auge, H., Chase, J. M., Harpole, W. S. & Knight, T. M. We need more realistic climate change experiments for understanding ecosystems of the future. *Glob Chang Biol*, doi:10.1111/gcb.14797 (2019).
- 56 Raich, J. W. & Schlesinger, W. H. The global carbon dioxide flux in soil respiration and its relationship to vegetation and climate. *Tellus B* **44**, 81-99 (1992).
- 57 Jansson, J. K. & Hofmockel, K. S. Soil microbiomes and climate change. *Nature Reviews Microbiology*, 1-12 (2019).
- 58 Crowther, T. *et al.* The global soil community and its influence on biogeochemistry. *Science* **365**, doi:10.1126/science.aav055 (2019).

#### Online-only methods - General references

- 59 Team, R. C. R: A language and environment for statistical computing. (2017).
- 60 Bond-Lamberty, B. & Thomson, A. Temperature-associated increases in the global soil respiration record. *Nature* **464**, 579-582, doi:10.1038/nature08930 (2010).
- 61 Shapiro, S. S. & Wilk, M. B. An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591-611 (1965).
- 62 Conover, W. J., Johnson, M. E. & Johnson, M. M. A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics* **23**, 351-361 (1981).
- 63 Chen, X., Zhao, P. L. & Zhang, J. A note on ANOVA assumptions and robust analysis for a cross-over study. *Statistics in medicine* **21**, 1377-1386 (2002).
- 64 McGuinness, K. A. Of rowing boats, ocean liners and tests of the ANOVA homogeneity of variance assumption. *Austral Ecology* **27**, 681-688 (2002).
- 65 Zimmerman, D. W. & Zumbo, B. D. Relative power of the Wilcoxon test, the Friedman test, and repeated-measures ANOVA on ranks. *The Journal of Experimental Education* **62**, 75-86 (1993).
- 66 Tomczak, M. & Tomczak, E. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in Sport Science* **1**, 19-25 (2014).
- 67 Thornley, J. & Cannell, M. Soil carbon storage response to temperature: an hypothesis. *Annals of Botany* **87**, 591-598 (2001).
- 68 Lloyd, J. & Taylor, J. On the temperature dependence of soil respiration. *Functional ecology*, 315-323 (1994).

366 69 Libohova, Z. et al. The anatomy of uncertainty for soil pH measurements and predictions: Implications for modellers  
367 and practitioners. *European journal of soil science* 70, 185-199 (2019).  
368 70 Kirkby, C. A. et al. Carbon-nutrient stoichiometry to increase soil carbon sequestration. *Soil Biology and*  
369 *Biochemistry* 60, 77-86 (2013).  
370 71 Bronick, C. J. & Lal, R. Soil structure and management: a review. *Geoderma* 124, 3-22 (2005).  
371 72 Beer, C. et al. Temporal and among-site variability of inherent water use efficiency at the ecosystem level. *Global*  
372 *biogeochemical cycles* 23 (2009).  
373 73 Averill, C., Turner, B. L. & Finzi, A. C. Mycorrhiza-mediated competition between plants and decomposers drives  
374 soil carbon storage. *Nature* 505, 543 (2014).  
375 74 Bradford, M. A. Thermal adaptation of decomposer communities in warming soils. *Frontiers in microbiology* 4, 333  
376 (2013).  
377 75 Friedman, J., Hastie, T. & Tibshirani, R. The elements of statistical learning. Vol. 1 (Springer series in statistics New  
378 York, 2001).  
379 76 Efron, B., Hastie, T., Johnstone, I. & Tibshirani, R. Least angle regression. *The Annals of statistics* 32, 407-499  
380 (2004).  
381 77 Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *Journal of the royal statistical society:*  
382 *series B (statistical methodology)* 67, 301-320 (2005).  
383 78 Kuhn, M. & Johnson, K. Applied predictive modeling. Vol. 26 (Springer, 2013).  
384 79 Breiman, L. Random forests. *Machine learning* 45, 5-32 (2001).  
385 80 Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232 (2001).  
386 81 Breiman, L. Bagging predictors. *Machine learning* 24, 123-140 (1996).  
387 82 Quinlan, J. R. in 5th Australian joint conference on artificial intelligence. 343-348 (World Scientific).  
388 83 Boulesteix, A. L., Janitz, S., Kruppa, J. & König, I. R. Overview of random forest methodology and practical  
389 guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data*  
390 *Mining and Knowledge Discovery* 2, 493-507 (2012).  
391 84 Xu, Q.-S. & Liang, Y.-Z. Monte Carlo cross validation. *Chemometrics and Intelligent Laboratory Systems* 56, 1-11  
392 (2001).  
393 85 Shcherbakov, M. V. et al. A survey of forecast error measures. *World Applied Sciences Journal* 24, 171-176 (2013).  
394 86 James, G., Witten, D., Hastie, T. & Tibshirani, R. An introduction to statistical learning. Vol. 112 (Springer, 2013).  
395 87 Kuhn, M. Building predictive models in R using the caret package. *Journal of statistical software* 28, 1-26 (2008).  
396 88 Grömping, U. Variable importance assessment in regression: linear regression versus random forest. *The American*  
397 *Statistician* 63, 308-319 (2009).  
398 89 Wei, P., Lu, Z. & Song, J. Variable importance analysis: a comprehensive review. *Reliability Engineering & System*  
399 *Safety* 142, 399-432 (2015).  
400 90 Yang, R.-M. et al. Comparison of boosted regression tree and random forest models for mapping topsoil organic  
401 carbon concentration in an alpine ecosystem. *Ecological Indicators* 60, 870-878 (2016).  
402 91 Greenwell, B. M. pdp: an R Package for constructing partial dependence plots. *The R Journal* 9, 421-436 (2017).  
403 92 Elith, J., Leathwick, J. R. & Hastie, T. A working guide to boosted regression trees. *Journal of Animal Ecology* 77,  
404 802-813 (2008).  
405 93 Geomatics, U. (European Space Agency: Louvain la Neuve, Belgium, 2017).  
406 94 Hengl, T. et al. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS one* 12, e0169748  
407 (2017).  
408 95 Moran, P. A. A test for the serial independence of residuals. *Biometrika* 37, 178-181 (1950).  
409 96 Legendre, P. Spatial autocorrelation: trouble or new paradigm? *Ecology* 74, 1659-1673 (1993).  
410

#### 411 Online-only methods - References of statistical packages

412 97 Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas.  
413 *International journal of climatology* 37, 4302-4315 (2017).  
414 98 Zomer, R. J. et al. Trees and water: smallholder agroforestry on irrigated lands in Northern India. Vol. 122 (IWMI,  
415 2007).  
416 99 Zomer, R. J., Trabucco, A., Bossio, D. A. & Verchot, L. V. Climate change mitigation: A spatial analysis of global  
417 land suitability for clean development mechanism afforestation and reforestation. *Agriculture, ecosystems &*  
418 *environment* 126, 67-80 (2008).  
419 100 New, M., Hulme, M. & Jones, P. Representing twentieth-century space-time climate variability. Part I: Development  
420 of a 1961-90 mean monthly terrestrial climatology. *Journal of climate* 12, 829-856 (1999).  
421 101 Didan, K. & Barreto, A. NASA MEaSUREs Vegetation Index and Phenology (VIP) Vegetation Indices Monthly  
422 Global 0.05Deg CMG [Data set]. NASA EOSDIS Land Processes DAAC, doi:10.5067/MEaSUREs/VIP/VIP30.004  
423 (2016).  
424 102 Batjes, N. H. ISRIC-WISE Harmonized Global Soil Profile Dataset. ISRIC-World Soil Information, Wageningen  
425 (2008).  
426 103 Yang, X., Post, W., Thornton, P. & Jain, A. Global gridded soil phosphorus distribution maps at 0.5-degree  
427 resolution. ORNL DAAC (2014).  
428 104 Giradoux, P. pgirmess: spatial analysis and data mining for field ecologists. R Package version 1.6.9 (2018).  
429 105 Graves, S., Piepho, H.-P., Selzer, L. & Dorai-Raj, S. multcompView: visualizations of paired comparisons. R  
430 package version 0.1-7 (2015).

431 106 Hope, R. M. Rmisc: Ryan Miscellaneous. R package version 1.5 (2013).  
 432 107 Kassambara, A. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R package version 0.6.0 (2020).  
 433 108 Fox, J. & Weisberg, S. An R companion to applied regression. Third Edition edn, (Sage Publications, 2019).  
 434 109 Lumley, T. & Lumley, M. T. Package 'leaps' Regression Subset Selection. Thomas Lumley Based on Fortran Code  
 435 by Alan Miller R package version 3.0 (2017).  
 436 110 Hastie, T. & Efron, B. lars: Least Angle Regression, Lasso and Forward Stagewise. R package version 1.2 (2013).  
 437 111 Zou, H. & Hastie, T. elasticnet: Elastic-Net for Sparse Estimation and Sparse PCA. R package version 1.1.1 (2018).  
 438 112 Liaw, A. & Wiener, M. Classification and regression by randomForest. R news 2, 18-22 (2002).  
 439 113 Peters, A. & Hothorn, T. ipred: Improved predictors R package version 0.9-9 (2019).  
 440 114 Wickham, H. The split-apply-combine strategy for data analysis. Journal of Statistical Software 40, 1-29 (2011).  
 441 115 Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A. & Leisch, F. e1071: Misc Functions of the Department of  
 442 Statistics, Probability Theory Group (Formerly: E1071). R package version 1.7-2 (2019).  
 443 116 Ridgeway, G. gbm: Generalized boosted regression models. R package version 1, 55 (2006).  
 444 117 Kuhn, M. & Quinlan, R. Cubist: rule-and instance-based regression modeling. R package version 0.2.2 (2018).  
 445 118 Kuhn, M. caret: Classification and regression training. R package version 6.0-84 (2019).  
 446 119 Schaefer, J. et al. corpcor: Efficient estimation of covariance and (partial) correlation. R package version 1.6.9  
 447 (2017).  
 448 120 Hijmans, R. J. raster: Geographic Data Analysis and Modelling. R package version 2.9-5 (2019).  
 449 121 Bivand, R., Keitt, T. & Rowlingson, B. rgdal: Bindings for the geospatial data abstraction library. R package version  
 450 1.4-4 (2019).  
 451 122 Wickham, H. ggplot2: elegant graphics for data analysis. (Springer, 2016).  
 452 123 Auguie, B. gridExtra: miscellaneous functions for "grid" graphics. R package version 2.3 (2017).  
 453 124 Brownrigg, R., Minka, T. P. & Deckmyn, A. maps: Draw Geographical Maps. R package version 3.3.0 (2018).  
 454 125 Bivand, R. & Lewin-Koh, N. maptools: Tools for reading and handling spatial objects. R package version 0.9-5  
 455 (2019).  
 456 126 Baquero, O. S. ggson: North Symbols and Scale Bars for Maps Created with 'ggplot2' or 'ggmap'. R package version  
 457 0.5.0 (2019).

**Corresponding author**

Reprints and permissions information is available online at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Sebastian Doetterl [[sdoetterl@usys.ethz.ch](mailto:sdoetterl@usys.ethz.ch)].

**Acknowledgements**

Further financial support has been given by: ETH Zurich and the German Research Foundation (DFG, Project No. 387472333). Special thanks goes to Heather Maclean for language proofreading.

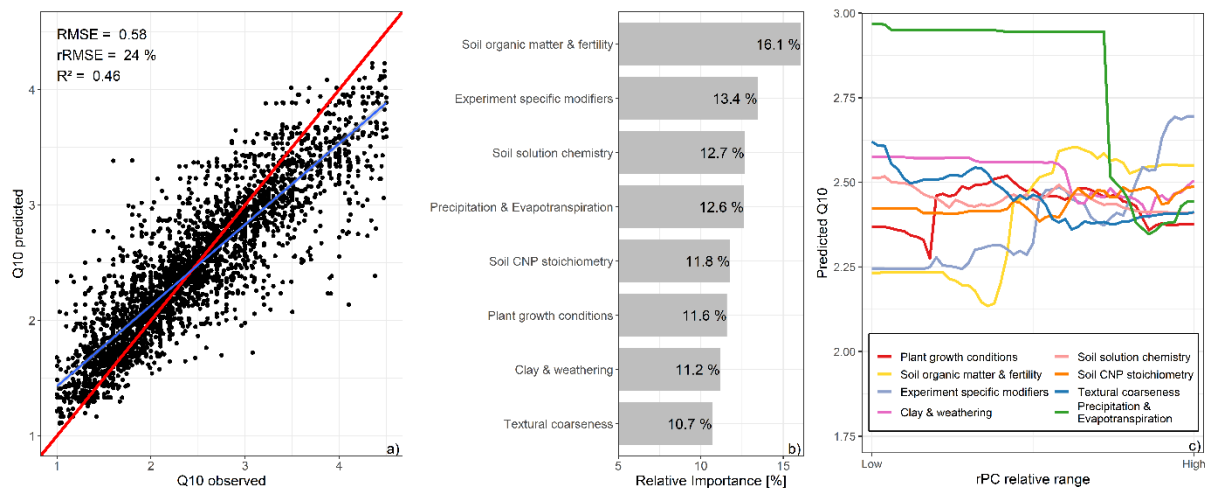
**Author contributions**

SD designed the research. DH conducted the data assembly and statistical analyses. DH and SD processed the data. All authors interpreted the data and contributed to the writing of the paper.

**Supplementary Information**

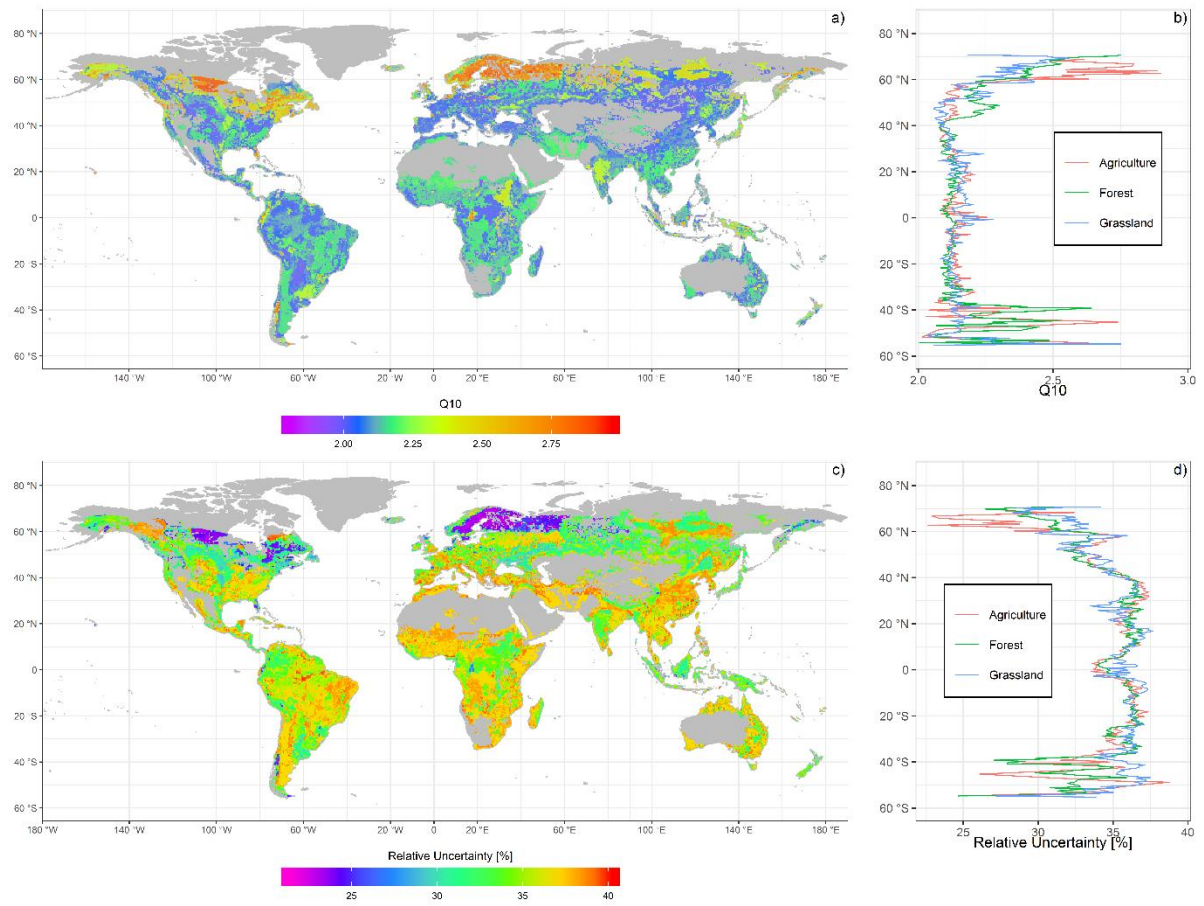
Supplementary Information, including source data information, a documented and annotated R code of all applied statistics as well as a database containing all input data files is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

## Figures



**Figure 1.** Predictions for soil  $Q_{10}$ , expressed as  $Q_{10}$  of soil respiration, (panel a, best data approach, random forest model) show a good fit across the complete data range. An assessment of the relative importance of rPCA derived variables (panel b) shows the dominance of the sum of soil parameters on the prediction over climate and vegetation related and experiment specific modifiers. Partial dependence plots (panel c) illustrate the variable effect of identified controls on averaged soil  $Q_{10}$ .





**Figure 2.** Map of the predicted average soil  $Q_{10}$  for terrestrial, non-desert environments (a), and the averaged latitudinal  $Q_{10}$ -pathway for different major land-use types (b), aggregated at  $0.25^\circ$  latitudinal resolution, and the corresponding distribution of relative uncertainty (c & d) (see methods section “Global soil  $Q_{10}$  mapping for details and uncertainty).

## **Online only methods section**

**Statistical analysis. Basics.** Statistical analysis was performed in R<sup>59</sup> Version 3.4.1 with additional packages (Table S8). For all statistical tests, a significant level of  $p < 0.05$  was used. A documented and annotated R code of all applied statistics as well as a database containing all input data.

**Database assembly and pre-processing.** Global  $Q_{10}$  data of soil basal respiration was collected from existing scientific databases<sup>60</sup> and published laboratory and field studies (for a full overview of all included studies, see supplementary data file “Q10\_Database“ and “Q10\_Database References”). The biggest part of soil  $Q_{10}$  values were taken from the Global Database of Soil Respiration Data (SRDB)<sup>60</sup>, together with information about the experiment temperatures. All compiled information from the database were counterchecked with the reported values of the original reference and duplicate  $Q_{10}$  values for the same temperature ranges were removed to keep only the temperature range related to the original study. Furthermore, web-based search engines like Google Scholar, ResearchGate or Web of Science were examined with catchphrases like “(laboratory/field measured) temperature sensitivity of SOC decomposition” or “soil  $Q_{10}$  (for laboratory/field experiments)” and added to our database. In total, data from 67 laboratory studies and from 493 field studies were compiled located between 68 °N to 43 °S and 176 °E to 156 °W (Fig. S1).

In our compiled database, soil  $Q_{10}$  data in these studies were taken from temperature ranges -5 - +50 °C, conducted from 1971 to 2018 with incubation lengths of several days to more than three years. We constrained our study to observations of topsoil samples (weighted averages for 0-30cm soil depth) and excluded studies that targeted autotrophic soil respiration. Reported  $Q_{10}$  in these studies represent the average soil  $Q_{10}$  during the length of the experiment and are considered as soil basal respiration.

The included soil  $Q_{10}$  data was tested for fulfilling normal distribution using the Shapiro-Wilk Normality Test<sup>61</sup> and for fulfilling homogeneity of variances with the Fligner-Killeen Test<sup>62</sup>. Comparability of soil  $Q_{10}$  and to avoid introducing potential biases was tested in several ways. To identify experiment specific influencing factors (measures taken by the experimentalist, see figure A2) we used one-way Analysis of Variances (ANOVA)<sup>63,64</sup> and in case of significant rejection of the statistical requirements for ANOVA, using the Kruskal-Wallis Test<sup>65</sup>, to test for differences in soil  $Q_{10}$  between (i) lab and field studies, (ii) studies reporting explicitly heterotrophic respiration vs. mixed respiration where remnants of autotrophic respiration cannot be excluded, (iii) sequential vs. parallel warming of soils and (iv) explicit pre-treatments of the samples vs. non-treated samples. Results of this test indicated only minor differences between the above compared studies (Figure S2). Furthermore, we evaluate the effect size of the applied Kruskal-Wallis test pairs to show the strength of the analysed relationship of statistical significant differences between sub-groups of the database. We computed the effect size as follows<sup>66</sup>:

$$\eta^2(H) = (H - k + 1)/(n - k) \quad (1)$$

where  $H$  is the test statistic,  $n$  is the number of observations and  $k$  is the number of groups in the model.

The analyses revealed that among all pairs, only grouping by climate zone has a strong effect on  $Q_{10}$  differences between subgroups. Other pairings, including the division of lab vs. field derived  $Q_{10}$  did not show a significant effect size (Table S2). Additionally, we tested our model performance on a data-rich and environmentally diverse region (Continental Europe, Scandinavia and the British Isles) using the same independent predictor variables and model structures as for the global approach, but predicting SRRW with only subsets of the data: one prediction where we use both field and lab data ( $n = 786$ ) combined, and one prediction each where we used only lab ( $n = 237$ ) or field data ( $n=549$ ). Our results (Figure S3) show that no

difference in model performance or potential bias can be observed based on the origin of parts of our data. Hence, we continued with a unified dataset for all other analyses, but included these experiment specific criteria in our later modelling approach as a confining factor (see methods section “statistical analyses”, results in tables S3-S5 and figure 1).

From the compiled  $Q_{10}$  data, values  $<1$  and  $>4.5$  were excluded from further analysis, as (i) we want to represent natural conditions that follow current paradigm, namely that soil basal respiration increases with incubation temperatures<sup>27</sup> and (ii) that  $Q_{10} >4.5$  are the result of the decomposition of large amounts of poorly decomposed, isolated organic matter (litter, roots<sup>67</sup>) in litter layers or de-frosting former permafrost soils. Furthermore, including these values would lead to inaccuracy in calculation with exponential equations<sup>68</sup>. These criteria led to the exclusion of 8% of the compiled observations (262 observations), resulting in a total of 3413 observations remaining across all major land use (grassland, cropland, forest and wetland) for the boreal, temperate, subtropical and tropical climate zones of the northern and southern hemisphere used in this study (Figure S2).

**Included independent variables.** To analyse the influence of soil properties, vegetation and climate parameters on  $Q_{10}$ , five climatic and vegetation as well as eight soil parameters were selected as independent variables. These parameters were used for all further statistical analyses. Where available, we used high-resolution local data taken from the included studies directly, resulting in our “*best data*” dataset. Where local studies did not include all the desired independent variables, global datamaps and satellite remote sensing data were used to fill gaps in climate and soil properties (Table S3). Note that values of  $\text{pH} < 3$  were replaced with a  $\text{pH} = 3$ , due to the fact that soils with a  $\text{pH} < 3$  do not occur in the ecosystems investigated in this study<sup>69</sup> and are an artefact created during the assembly of the original dataset (best data approach: 9 datapoints replaced; generalized data approach: 0 datapoints replaced). Note that these global datamaps of independent controls show variable spatial resolutions ranging from

250m - 0.5° and represent averages over 1-30 years (see Table S1 for details). To assess the potential impact of spatially highly variable data in our analyses, we used the data in the highest available resolution and did not transformed the data to match resolutions. In addition, to represent potential controls that result from the interaction of soil parameters with climate and vegetation, a series of interaction terms were included. Organic carbon/organic nitrogen/total phosphorus ratios were included to represent effects of nutrient stoichiometry in soils<sup>70</sup>. Clay content/mean annual temperature ratios were included to represent soil weathering and changes in mineral surface area<sup>71</sup>. Base saturation/clay content and potential cation exchange capacity (CEC)/clay content as well as base saturation/CEC ratios were used to assess mineral surface charge effects. Base saturation / pH ratios were used to assess soil acidity effects. Mean annual precipitation (MAP) / Potential evapotranspiration (PET) and PET/normalized vegetation index (NDVI) ratios were used to assess plant productivity as well as precipitation and evapotranspiration related effects<sup>72</sup>.

The resulting dataset of independent variables is not inclusive for all experimentally identified controls (i.e. variability of microbial decomposers and their strategies are not included)<sup>73,74</sup>. However, key criteria for their selection in our modelling exercise was availability as global datasets to fill data gaps of the metadata of the included warming studies. Furthermore, all included variables stand in a causal relationship for controlling biological processes and C cycling between soils and atmosphere and vary across a large range of possible values (Table S1) that represent the majority of conditions in which biological processes take place in soils (i.e. very acidic, to very basic, very low and very high temperatures, etc.). This compilation of empirical data was selected to bridge a crucial gap from experimental finding to implementation of soil processes into earth system models.

**Statistical analysis. Rotated principal component analysis.** To increase the identifiability of larger groups of controls and to reduce the number of independent variables that are

autocorrelated, we used rotated principal component analysis (rPCA), performed for both our *best data* model building (Table S3) as for our *generalized data* approach (Table S4) and interpreted the loading of each principal component according to their underlying relevance as a controlling factor for SRRW. To minimize multicollinearity effects, the Variance Inflation Factor (VIF) was estimated for all independent predictor variables and maximal VIF was eliminated until all independent variables possessed a  $VIF < 5$ . As rotation method and to minimize multicollinearity, variance maximizing (VARIMAX) was used. The selection of an optimal number of principal components was done on the basis of the Kaiser-Guttman rule and limited to principal components with an Eigenvalue  $> 1$ . This resulted in 8 rotated principal components (rPC), identifying the eight most important groups of explaining variables for SRRW (Table S3).

**Statistical analysis. Predictive modelling.** To build and identify the best model for predicting soil  $Q_{10}$  and using the results of the rPCA analyses, regression modelling was conducted including four different linear and four different non-linear regression types. Linear regression included models without (LM) and with (LEAPS) stepwise selection<sup>75</sup> as well as models such as least angle regression (LARS)<sup>76</sup> and Elastic Net (ENET)<sup>77</sup> that use a penalizing term to the regression coefficients of those variables with minor influence on the prediction<sup>78</sup>. Non-linear regressions included the tree- and rule-based (=representing the path of partitioned regression(s) by using distinct if-then rules to create prediction models)<sup>77</sup> models random forest (RF)<sup>79</sup> and boosted tree model (BOOSTED)<sup>80</sup>, as well as model bagged tree (BAGGED)<sup>81</sup> and cubist (CUBIST)<sup>82</sup>. All models, except for the LM linear regression and the BAGGED model, have built-in feature selection procedures and were tuned individually, to increase the accuracy and control the complexity of the models<sup>78</sup>. As part of the tuning process, the following steps have been taken: LEAPS models were trained for the maximal number of variables. For penalizing models, penalty terms for feature reduction (i.e. lowering the effect of less important

variables on the final linear equation) varied between 0 and 0.1 in 0.01 steps. The RF models were constrained by setting the maximum number of allowed trees to 1000. The number of included predictors were set to the maximum number of possible predictors divided by three<sup>83</sup>. BOOSTED were trained with a minimum of 10 to a maximum of 100 trees with 1-7 nodes, a shrinkage factor of 0.01 or 0.1 and a maximum size of 5. To train the CUBIST models, 1-9 by 2 neighbours and 1, 5, 10, 50, 75 and 100 communities were used. For all models, Monte-Carlo-cross-validation<sup>84</sup>, with 100 repeated data resamples and a ratio of 80 % training to 20 % validation data were used to assess the uncertainty of model structures and prevent overfitting. Root mean square error of cross-validation (RMSE) and  $R^2$  were estimated for all tuned models and used to analyze the residual variance and accuracy of the models<sup>85</sup> and as a criterion for ranking model performance (Table S5). For an easier interpretation of the uncertainty of estimated SRRW, relative root means square error (rRMSE) was estimated by dividing the absolute error by the global mean of  $Q_{10}$ . Random Forest regressions resulted in the best model performance within one-standard error of minimal RMSE<sup>86</sup> and were used for all further analyses of variable importance. Furthermore, residual plots for the global best model (Figure 4S) and the three data-rich examples of continental Europe (Figure S3) were created. All residual plots show random patterns, indicating a good fit of the used random forest models for the global and the European models.

**Statistical analysis. Assessing variable importance.** To estimate the influence of the identified rPC variables for predicting  $Q_{10}$ , we assessed variable importance using permutation variable importance measurements (PVIM) through the variable importance tool implemented in R caret package<sup>87</sup> for the model with the highest accuracy and prediction quality (Random Forest RF). Briefly, to assess the error of prediction in the model, the PVIM method calculates the mean square error for every given regression tree with out-of-bag estimates<sup>79,88</sup>. The resulting measure of variable importance of RF models represents the influence of the predictor

variables on the model results<sup>89</sup>. For better comparability all independent controls in our models<sup>90</sup>, the included independent rPC control variables were normalized on a scale of 0-100% to represent relative importance for the model outcome.

**Statistical analysis. Partial dependency of controls.** Partial dependence analyses using the R package pdp<sup>91</sup> were used to test effects between predicted Q<sub>10</sub> and independent controls across the whole range of possible values that were included in the RF modeling. Briefly, the method results in a statement about the global relationship of an independent variable to the predicted across the whole range of all potential values by removing and averaging out the effect of other independent controls and isolating the effect of the targeted independent variable(s)<sup>80</sup>. In contrast to the assessment of the relative importance of an independent variable overall, partial dependence analyses and their visual representations (partial dependence plots, PDP) can illustrate the average marginal effect of one or more independent variables on the predicted outcome of a machine learning model<sup>80</sup> across a specific range of values. For example, a PDPs can show whether the relationship between the predicted variable and an independent control is linear, monotonic or complex<sup>92</sup>. The shape and knickpoints of the PDP curve can then be used to interpret and identify areas where an independent has a particular strong and direct effect on the predicted, and where its control is rather indirect, for example through influencing other independent variables. For simpler interpretation of the PDPs x-axis from low to high, the curves of rPCs with dominant negative loading (best data approach: rPC1, rPC7; Table S3) were reversed.

As an example in our study, PDPs illustrate that precipitation and evapotranspiration has a weak effect and control on Q<sub>10</sub> at lower ranges, but a stronger effect at higher ones (Figures 1c). As the loading of our rPC variable “precipitation and evapotranspiration” is not mixed with other controls (Table S3, S4), the PDP allows a direct interpretation of the variable’s



value. In contrast, temperature has a complex relationship to the predicted soil  $Q_{10}$ , mostly through affecting plant growth conditions, experimental setup and weathering.

**Statistical analysis. Global soil  $Q_{10}$  mapping.** A map of the global distribution of soil  $Q_{10}$ , expressed as  $Q_{10}$  of soil basal respiration and a corresponding map of the relative uncertainty of prediction (Figure 2) was derived using our *best data* rPCA structure and scores (Table S4) and a RF model with the included global climate, vegetation and soil datasets (Table S1) that we used to build our *generalized data* model of soil  $Q_{10}$ . Using the datasets of the *generalized data approach*, we calculated factor maps based on the primary input variables for our 8 rPC scores for each according raster cell before using them to calculate a spatial explicit map of global soil  $Q_{10}$ . In consequence, the resulting map corresponds in quality to the results of our RF model results without experiment specific modifiers as explanatory variables (Table S7;  $R^2 = 0.42$ , RMSE 0.61). For this mapping exercise at a global scale, input variables were run at a  $0.5^\circ$  resolution and later aggregated at  $0.25^\circ$  latitudinal resolution to derive a mean  $Q_{10}$  value separately for major land use systems at the respective latitude. Land use was derived using the 2015 ESA CCI-LC<sup>93</sup> land cover maps (300m original resolution) and summarized to agriculture, forest and grassland systems. We excluded those areas from our prediction where (i) data in any of the required predictors was missing, (ii) land use was different to the aggregated land use systems listed above or (iii) areas where located in climate zones which were not targeted by our model (polar and non-polar (semi-) deserts). Predictors that were available at a higher resolution were resampled using geostatistics to match a  $0.5^\circ$  resolution. The resulting map's averages shows significant differences for distinct USDA and WRB<sup>94</sup> soil orders across climate zones and land use systems (Table S9). To assess the uncertainty related to the creation of the map due to resampling of data and unexplained variability not captured by the rPC scores we run the model also at a finer resolved  $1 \text{ km}^2$  grid or those areas where input variables were available at this higher resolution. This analysis revealed an overall

uncertainty of our global soil  $Q_{10}$  map averaging at  $27.4 \pm 10\%$ . The corresponding map of relative uncertainty of prediction was built by displaying standard deviation/divided by the mean of prediction based on the results of our final random-forest model with standard deviation related to the range of possible predictions based on the build-up of the used decision tree after 500 model runs.

## **Caveats**

### *The “real” controls and the influence of experimental modifications*

The identification of variables for regression models, including their importance and dependency assessments, are highly dependent on the range in which the included variables can vary. In our global model design, we addressed this by including independent variables that vary across a large range of possible values in which biological processes take place and which represent the majority of conditions that can occur in soils (Table S1). To assess the validity of our interpretation and the robustness of our models, we repeated all statistical analyses that involve independent predictors by using data only derived from global datamaps, further referred to as the *generalized data* approach (Table S4 & Figure S3). An approach that excluded experiment specific modifiers (Table S7) generally yielded in less performance than fully parameterized models, but differences were marginal ( $R^2 = 0.03 - 0.42$ ; RMSE 0.61 - 0.79). Together with our analyses of potential biases in the database that yielded negative results (Figure S2) this suggests that experimental and climatic conditions, if made comparable across larger gradients, do not exceed the control of soil variables on soil  $Q_{10}$ .

### *Spatial autocorrelation*

Building our predictive models of soil  $Q_{10}$  (Figures 1a 3Sa), we tested for and quantified spatial autocorrelation of modelled residuals using Moran I test<sup>95</sup>. Results indicated only a minor influence of spatial autocorrelation for all linear models (Moran I = ca. 0.3 for all models).

746 Further corrections taking into account spatial variability and the accuracy of geographic  
747 coordinates<sup>96</sup> in the modelling structure of the linear models showed no improvement. In  
748 combination with the good results of the machine learning models (Table S6 & S7), we  
749 interpret these results as supportive to our finding that the relationship of soil Q<sub>10</sub> and the  
750 included independent controls are primarily non-linear.