JGR Biogeosciences

RESEARCH ARTICLE
10.1029/2024JG008420

Measuring Gas Transfer Velocity in a Steep Tropical Stream: Method Evaluation and Implications for Upscaling

Adam T. Rexroade[1], Marcus B. Wallin[2], and Clément Duvert[1,3]

[1]Research Institute for the Environment and Livelihoods, Charles Darwin University, Darwin, NT, Australia, [2]Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala, Sweden, [3]College of Science and Engineering, James Cook University, Cairns, QLD, Australia

Key Points:

- We compared four field-based and four model-based methods to estimate gas transfer velocity in a steep, tropical stream
- When compared to in situ measurements, empirical models tended to underestimate gas transfer velocity at high values
- The use of CO_2 as a tracer gas requires accurate quantification of other CO_2 fluxes before estimating gas transfer velocity

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. T. Rexroade,
rexroade@cdu.edu.au

Citation:

Rexroade, A. T., Wallin, M. B., & Duvert, C. (2025). Measuring gas transfer velocity in a steep tropical stream: Method evaluation and implications for upscaling. *Journal of Geophysical Research: Biogeosciences*, *130*, e2024JG008420. https://doi.org/10.1029/2024JG008420

Received 4 SEP 2024
Accepted 20 JAN 2025
Corrected 12 AUG 2025

This article was corrected on 12 AUG 2025. See the end of the full text for details.

Abstract Greenhouse gas emission estimates from streams rely, in part, on accurate measurements or estimates of the gas transfer velocity, which describes the physical efficiency for gas exchange across the water-air interface. Numerous methods for measuring or modeling gas transfer velocity exist, yet few studies compare these different methods. Additionally, current models of gas transfer velocity in streams are predominantly derived from measurements in low-gradient, temperate, or boreal streams. Here, we measured gas transfer velocity using four different methods in a high-energy, tropical headwater stream under a range of flow conditions, and compared these measurements to indirect estimates from four empirical models. Our results show that, when compared to the use of a biologically inert gas tracer (propane), floating chambers produced lower gas transfer velocity values. Using carbon dioxide (CO_2) as a tracer gas was unreliable without considering other natural sources and sinks of CO_2 and yielded gas transfer velocities lower than when using propane. Existing empirical models tended to underestimate gas transfer velocity, compared to the inert tracer gas. When using empirical models to upscale the emission flux along an entire stream reach, implementing them at finer spatial resolutions yielded flux estimates closer to measured fluxes. We also highlight the extreme spatial variability of gas transfer velocity across small spatial scales, which contrasts with its relative stability across changing hydrological conditions. The discrepancies between methods highlight the need for further research in measuring and upscaling gas transfer velocity, particularly in very turbulent steep streams.

Plain Language Summary Streams and rivers emit greenhouse gases and the rate at which this process occurs is determined, in part, by the gas transfer velocity. This parameter can be measured or estimated in different ways, yet few studies compare these different methods. Most of the studies that examine gas transfer velocity have occurred in small, flat streams in temperate or boreal ecosystems. Here, we measured the gas transfer velocity in a steep, tropical stream using eight different methods. We aimed to understand how existing methods compare to each other and how well these methods work in systems other than flat streams. Our results show that methods that use a biologically active tracer gas to measure gas transfer velocity require more careful consideration of other natural processes than methods that use biologically inert tracer gases. Models that estimate gas transfer velocity tend to overestimate this parameter when it is low and underestimate this parameter when it is high. We also found that although gas transfer velocity varies substantially in space, it remains relatively stable over time. The differences we observed between methods highlight the need for further research on how to best measure or model gas transfer velocity, particularly in steep streams.

1. Introduction

Streams contain biologically active and chemically important dissolved gasses such as oxygen (O_2), carbon dioxide (CO_2), and methane (CH_4) (Hall & Ulseth, 2020). Both O_2 and CO_2 are of interest to stream ecologists for their role in photosynthesis, respiration, or the combined process of metabolism, which is influential in determining whether a given stream is a sink or source of carbon (Battin et al., 2023; Bernhardt et al., 2018). In addition to internal cycling of CO_2 and O_2, streams emit significant amounts of CO_2 and CH_4 to the atmosphere and have received increasing attention, especially over the past two decades, for their role in global carbon cycling (Liu et al., 2022; Raymond et al., 2013; Rocher-Ros et al., 2023). CO_2 produced from soil respiration or internal stream metabolism, as well as CH_4 produced within stream sediments or riparian soils, can diffuse across the water surface and into the atmosphere, representing a significant source for atmospheric greenhouse gases (GHGs) globally (Liu et al., 2022; Rocher-Ros et al., 2023). However, both the concentration and flux to the atmosphere of these gases can be highly variable across time and space in streams, which makes them difficult to generalize on a

reach-scale (Natchimuthu et al., 2017; Wallin et al., 2011). Steep streams in particular have been shown to emit gases to the atmosphere at higher rates than their lower-gradient counterparts (Maurice et al., 2017; Ulseth et al., 2019), with global syntheses highlighting the disproportionate emission contribution from streams draining steep landscapes (Horgby et al., 2019; Liu et al., 2022). Yet large uncertainties remain in estimating gas transfer velocities (k) for turbulent streams, where bubble-mediated emission is often prevalent (Hall Jr. & Madinger, 2018; Ulseth et al., 2019). Additionally, tropical regions have been found to have elevated concentrations of both CO_2 and CH_4 when compared to temperate regions (Liu et al., 2022; Regnier et al., 2013; Rocher-Ros et al., 2023). Despite their potential contributions to global carbon cycling, tropical streams, especially headwaters, tend to be underrepresented in global upscaling studies (Lauerwald et al., 2023). This underrepresentation may be the result of fewer gas concentration observations from tropical streams, as well as coarse resolution of available digital elevation models which limit the ability to integrate gas measurements in headwater streams with catchment characteristics typically used for stream emission modeling (Lauerwald et al., 2015; Liu et al., 2022).

Movement of a gas across a water surface is a product of the concentration gradient (the difference between the concentration of the gas in the surface water and the concentration in the water at saturation with the atmosphere) and k (Raymond & Cole, 2001). The concentrations of CO_2, CH_4, and O_2 in both the surface water and the overlying atmosphere are typically easily measured, but k has proven more difficult to quantify. k is largely dependent on surface turbulence which promotes mixing of water with the overlying air (Hall & Ulseth, 2020; Jähne & Haußecker, 1997). Characteristics that increase surface turbulence—such as channel slope, channel depth, energy dissipation rate (ε_D), flow rate, and flow velocity—have been found to scale with k (Raymond et al., 2012; Wallin et al., 2011). Spanning multiple orders of magnitude from less than 5 m d^{-1} in slow moving, relatively flat streams to over 4,000 m d^{-1} in steep, alpine streams, k has a significant influence in determining GHG emissions from rivers (Raymond et al., 2012; Ulseth et al., 2019). In addition to k, the gas concentration gradient also influences flux and lower gas solubility in warm waters, often found in the tropics (van Vliet et al., 2013), results in a lower gas concentration at saturation, which would change the concentration gradient and subsequent gas flux.

A particular challenge in quantifying k is its spatial and temporal variability on short/small and long/large scales. Within a single stream, small scale changes in hydro-morphological conditions can have significant impacts on k and thus gas emissions (Botter et al., 2021, 2022; Kokic et al., 2018; Vautier et al., 2020). Because k is influenced by hydrological regimes, it also experiences temporal variability. Largely due to changes in discharge, k can span an order of magnitude at a single stream site over an annual timescale (McDowell & Johnson, 2018).

Methods used to compute k in streams range from heavily instrumented in situ measurements (direct turbulence measurements using acoustic Doppler velocimetry (Kokic et al., 2018)), to indirect but still in situ measurements (tracer gas additions, floating chambers or eddy covariance in parallel with in-stream gas concentration to resolve k), to empirical modeling based on hydrology and/or geomorphological characteristics (Hall & Ulseth, 2020). Each of these methods have distinct advantages and disadvantages dependent on the type of stream system they are applied to, and the scale in space and time that the measurements or modeling represent.

Floating chambers provide an in situ measurement of gas flux, which can be used to calculate k if the concentration of the specific gas in the atmosphere and water is known. Although this method provides a direct measurement of gas flux that can easily be used to resolve a k value, these are spot measurements representing local areas in space and time. This small footprint may pose a weakness when deriving k in streams that have highly variable morphology across space and hydrology across time. Additionally, chambers floating on the stream surface (whether tethered to a location or allowed to drift with the current) can influence the surface conditions and hence alter the gas flux and k value (Lorke et al., 2015; Vachon et al., 2010).

Tracer gas additions also provide an in situ measurement of k (Tsivoglou, 1967). Historically, SF_6 and propane have often been used as tracer gases due to being easy to measure on a gas chromatograph, and for being biologically inert (and in the case of propane, inexpensive) (Mulholland et al., 2001; Wanninkhof et al., 1990); however, recent years have seen the use of alternative gases such as argon (Hall & Madinger, 2018; Ulseth et al., 2019). The k measured using a gas tracer is specific to that gas and must be converted to the gas of interest (e.g., O_2 or CO_2) using a Schmidt scaling number adjusted for temperature, which can introduce uncertainties (Hall Jr. & Madinger, 2018). Recently, CO_2 has been tested as a tracer gas to avoid the need to scale between gases (McDowell & Johnson, 2018). CO_2 is an attractive tracer gas option since it is relatively easy to measure using an infrared gas analyzer, or in situ sensor. Unlike SF_6, propane, and argon, CO_2 is biologically active in

streams and may require additional consideration of natural sources and sinks. Tracer gas additions cover a larger spatial footprint than chambers; however, they still have a limited temporal footprint (ca. 1 hour). They work best where the study reach is well mixed and are thus well adapted to high-gradient, turbulent streams where chamber measurements may not be applicable.

Empirical models provide estimates of k based on parameters often obtained from geospatial products or monitoring stations (slope, flow rate, or flow velocity) (e.g., Raymond et al. (2012)). For this reason, they are typically used when upscaling atmospheric flux at regional or global scales. Depending on how they are applied and how the required data inputs are obtained, an empirical model can provide an estimate of k for a stream without ever having to go to the field. However, this comes at a cost of higher uncertainties, particularly in morphologically complex headwater streams (Vautier et al., 2020). Not only do these models have high uncertainties but different models can also yield vastly different results for the same location and time (see Figure S3 in Duvert et al. (2018)). Recent studies using tracer gas injections in steep streams have led to the development of empirical models adapted to these high-gradient systems (Hall Jr. & Madinger, 2018; Maurice et al., 2017; Ulseth et al., 2019), and also suggest that previous models developed in lower gradient streams underestimate k in these steep systems (Ulseth et al., 2019).

Despite a wide body of work which relies on accurate estimates of k across large spatial scales and significant effort developing methods suited to different types of streams, there are limited studies which compare these methods (e.g., Lorke et al. (2015), McDowell and Johnson (2018), and Vautier et al. (2020)). There is a limited number of studies that measure k (rather than model it) in steep, turbulent streams (Horgby et al., 2019; Ulseth et al., 2019), and direct measurements in steep, tropical streams are also missing from the literature. These data gaps leave global upscaling efforts exposed to significant uncertainties in GHG fluxes from such areas and methodology choices regarding k may lead to large discrepancies between studies. By simultaneously measuring k using four field-based methods (a floating chamber, and three variations of tracer gases) and four empirical models (Equations 2 and 5 from Raymond et al. (2012) and the models developed in Natchimuthu et al. (2017) and Ulseth et al. (2019)), we aim to.

1. Apply and compare different field-based methods of measuring k in a high-energy tropical stream.
2. Compare existing commonly used empirical models against in situ measurements to estimate k.
3. Demonstrate the impact of model selection and spatial resolution of in situ measurements on upscaled gas flux calculations.

2. Materials and Methods

2.1. Study System and Design

Shady Creek is a steep, first-order and low pH (5.2–5.5) stream in Litchfield National Park (Northern Territory, Australia) that drains an approximately 350 ha sandstone dominated catchment. Located in the wet-dry tropics of Australia (Köppen-Geiger climate classification Aw; (Beck et al., 2018)), it receives over 90% of its average 1,685 mm (median 1,770 mm) of annual precipitation between the months of November and April (Bureau of Meteorology station number 14021 http://www.bom.gov.au/climate/data). The streambed is composed of a mix of exposed bedrock, gravel, and sand. Diffuse groundwater inflows through the streambed are probable in small pockets of sand/gravel and discrete groundwater inflows can be observed throughout the stream, especially on steep hillslopes. The stream is turbulent with few areas where water is stagnant.

Three reaches located within a 197-m stretch of Shady Creek comprised our study sites (Figure 1a). These reaches differed in length, slope, depth, and flow velocity (Table 1). The three sites were loosely classified into a high, medium, and low energy reach. These energy classifications were not only based on measured parameters such as flow velocity and slope but also on subjective visual judgments of surface turbulence. Although we refer to these three different sites as low/medium/high energy, these labels are relative to each other and all these sites would generally be described as high-energy if comparing to most global stream networks.

Briefly, the high-energy reach is 12.7 m long with an average slope of -0.193 m m^{-1}, although nearly all of the change in elevation along the reach occurs at a 3.3 m tall waterfall 4.5 m downstream from the upstream end of the reach. The medium energy reach is 57.8 m long with an average slope of -0.032 m m^{-1} with numerous distinct

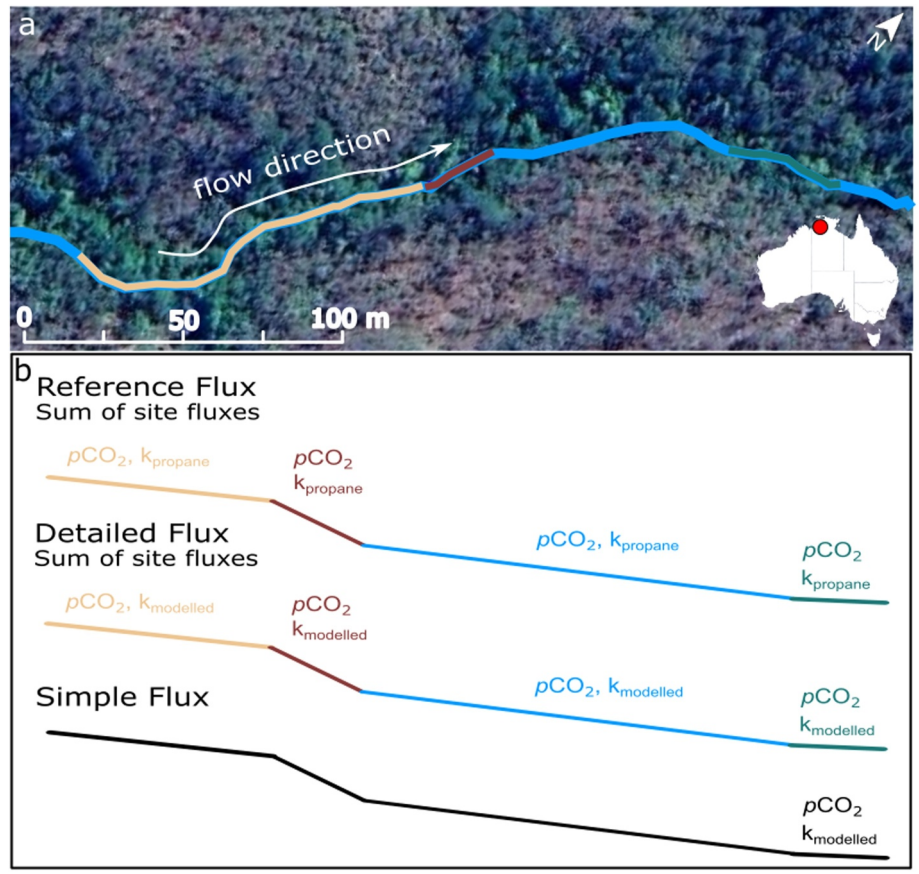

Figure 1. (a) Site map denoting the 197-m long section of Shady Creek that was used in the current study. The three study reaches are shown in tan (mid-energy), red (high-energy), and green (low-energy). (b) Conceptual diagram depicting how the three different types of fluxes were calculated. In the reference and detailed fluxes, the pCO_2 values denote the pCO_2 measured at the up and downstream end of each individual reach. In the detailed and simple flux, the modeled k uses velocity/discharge measurements were taken at the downstream end of each reach, or in the case of the simple flux, the downstream end of the entire 197-m reach. Figure 1b is not drawn to scale; see SI1 for a detailed elevation profile.

cascades of approximately 0.5–1.0 m tall. The low energy reach is 21.9 m long with an average slope of −0.030 m m^{-1} and has a single small (<0.5 m) cascade (Table 1; Figure SI1).

Between October 2022 and November 2023, we sampled the three stream reaches over 16 days resulting in 37 unique sampling "events." At each event, we measured basic hydrological (discharge and velocity) and morphological (width and depth) parameters, as well as computed k for our study reach using eight methods of direct measurements and modeling. These methods included (a) propane as a tracer gas using a standard first-order model to calculate k, (b) CO_2 as a tracer gas using a standard first-order model to calculate k, (c) CO_2 as a tracer gas and mass balance model to calculate k, (d) floating chambers coupled with in-stream CO_2 concentration, (e) and (f) models 2 and 5 from Raymond et al. (2012), (g) the empirical model developed by Natchimuthu et al. (2017), and finally (8) the piecewise linear regression model by Ulseth et al. (2019). Methods 1 through 4 compute a k specific to either propane or CO_2 at the measured temperature; these values were converted to k_{600} (see Sections 2.3 and 2.4), the k value for CO_2 at 20°C, to allow comparison between these four methods as well as with the empirical models which calculated k_{600}. Due to the wide usage of method 1, here we consider the k estimates obtained from this method as "reference" values, to which we compare k derived from the other seven methods. It

Table 1
Average Conditions for Each of the Three Sites

	Low	Medium	High
Length (m)	21.9	57.8	12.7
Width (m)	2.1 ± 0.4	2.2 ± 0.2	3.4 ± 1.2
Depth (m)	0.4 ± 0.07	0.4 ± 0.06	0.2 ± 0.1
Avg. Slope (m m^{-1})	−0.030	−0.032	−0.193
Avg Flow (L s^{-1})	220 ± 196	192 ± 167	191 ± 161
Avg flow velocity (m s^{-1})	0.25 ± 0.2	0.27 ± 0.2	0.39 ± 0.2
Avg CO_2 concentration (ppm)	882 ± 253	1,073 ± 260	1,017 ± 311

Note. Standard deviation is shown when appropriate.

should be noted that these reference k values should not be considered as "true values," although we believe that these values are likely similar to what the true value of k would be. Throughout the manuscript, "overestimate" and "underestimate" are used in reference to this propane derived k value, not necessarily the true value of k.

2.2. Hydro-Morphological Measurements

A topography survey using an optical level (NA2; Leica Geosystems; St. Gallen, Switzerland) was conducted to determine the length and slope (S; m m^{-1}) of each reach as well as the entire 197-m stream section (Figure S1 in Supporting Information S1). Lengths of each reach were checked against manual measurements using a tape measurer. Depth ($\bar{z}$; m) and width were measured at each reach on each sampling day at 3–4 cross-sections distributed along the length of the reach. At each cross-section, depth was measured every 20–50 cm (depending on total width). Discharge (Q; L s^{-1}) and flow velocity (v; m s^{-1}) were measured at each site and each day with salt dilution gaging (Day, 1977) using conductivity sensors at the up and downstream ends of each study reach (ProSolo handheld and EXO2 multiparameter sonde, both equipped with conductivity sensor, YSI Inc.; Yellow Springs, Ohio, USA). Depending on the flow rate, a slug of between 100 and 1,000 g of salt were added ca. 10 m upstream of the upstream measurement point to allow adequate mixing. Using these measures of slope and velocity, as well as the gravitational acceleration (9.8 m s^{-2}), ε_D (m^2 s^{-3}) was calculated according to Equation 1 (Moog & Jirka, 1999).

$$\varepsilon_D = 9.8 \times S \times \nu \tag{1}$$

2.3. Propane Diffusion

Propane was diffused into the stream at a constant rate of 8–10 L min^{-1} for 30–40 min, approximately 10 m upstream of each study reach. The time of 30–40 min was chosen based on the time it took for the measurements from a C-sense CO_2 sensor (Turner Designs, United States) placed at the downstream end of the reach to stabilize after the addition of CO_2 (see Section 2.4), indicating that the gas had ample time to diffuse into the water and travel the length of the study reach. The diffusion points were located immediately upstream of highly turbulent zones (e.g., a small cascade) to ensure thorough mixing across the channel. After this time, eight 20 mL glass crimp top vials were filled with water from the upstream and downstream ends of the study reach (four at each location distributed across the channel) and sealed with an aluminum screw cap with septa. Samples contained no headspace and were refrigerated until analysis. Shortly before analysis, 10 mL of water was removed from each vial. Samples were analyzed for propane at Charles Darwin University on a Clarus 680 gas chromatograph (Perkin Elmer) with a flame ionization detector (FID).

The gas transfer velocity of propane, $k_{C_3H_8}$ (m d^{-1}), was calculated by solving the following equation (Wanninkhof et al., 1990):

$$C_x = C_0 e^{-\frac{(k_{C_3H_8})x}{v\bar{z}}} \tag{2}$$

where C_0 and C_x are the concentration of propane at the up and downstream measurement points and x is the length of the study reach (m). C_x and C_0 were not adjusted to reflect an increase/decrease of flow at the two sampling points due to minimal changes in flow rate at the locations of C_x and C_0. (see Text S2 in Supporting Information S1). To compare to other methods, $k_{C_3H_8}$ was converted to k_{600} using the following equation:

$$k_{600} = \left(\frac{600}{Sc_{C_3H_8}}\right)^{-0.5} \times k_{C_3H_8} \tag{3}$$

where k_{600} is the gas transfer velocity standardized to CO_2 at 20°C (m d^{-1}) and $Sc_{C_3H_8}$ is the Schmidt number for propane, calculated from (Witherspoon & Saraf, 1965):

$$Sc_{(C_3H_8)} = 2864 - 154.14T + 3.791T^2 - 0.0379T^3 \tag{4}$$

where T is water temperature (°C). In total, 23 of the 37 sampling events generated quality checked data used for further analysis.

2.4. CO_2 Diffusion

CO_2 was also diffused into the stream at a constant rate of 8–10 L min^{-1} at the same locations and times as the propane injection. From October 2022 until March 2023, CO_2 was measured continuously using two Turner Designs C-sense pCO_2 sensors placed at each end of the stream reach during gas injection. CO_2 concentration was monitored in real time at both points and CO_2 values used to compute k were recorded once the CO_2 concentration at the downstream end of the reach had stabilized. This stable value was typically reached after 30–40 min and suggests adequate travel time from the point of diffusion and sensor response time. From April through November 2023, CO_2 was measured using a headspace equilibration method. Briefly, 45 mL of water was equilibrated with 15 mL of atmospheric air by gently shaking a 60 mL syringe containing the two components for 2 minutes. After equilibration, the gas was transferred to a 20 mL syringe via a 3-way stopcock valve. Syringes were kept cool until analysis on a trace gas analyzer (LI-7810; Li-Cor Environmental; Lincoln, Nebraska, USA) within 12 hr of sample collection. Using these CO_2 measurements, k for CO_2 (k_{CO_2}), was calculated in two different ways described below.

2.4.1. First Order Power Function

k_{CO2} was calculated in a way similar to Equation 2, except that C_0 and C_x were the concentration of CO_2 at the up and downstream sampling points after ca. 40 minutes of diffusion minus the concentration of CO_2 present before CO_2 was diffused into the stream. k_{CO2} was then converted to k_{600} using Equation 3, except the Schmidt value of CO_2 ($Sc_{(CO_2)}$) was used (Wanninkhof, 1992):

$$Sc_{(CO_2)} = 1911 - 118.11T + 3.453T^2 - 0.0413T^3 \quad (5)$$

This method treats CO_2 as an unreactive tracer gas, the same way C_3H_8 is treated in the previous section. However, this is an assumption and does not consider that CO_2 is naturally occurring in streams and can be altered by groundwater additions, produced or consumed by aquatic metabolism, or be lost due to carbonate equilibria shifts. However, in this study system, the latter two of these are likely minimal due to the low pH water that would limit a strong carbonate shift, and a (likely) low rate of metabolism. A near-by and similar system exhibited a net primary production of 0.84–4.06 g C m^{-2} d^{-1} (Solano et al., 2023), and we expect Shady Creek to be lower due to its high turbulence and short residence time. Despite not accounting for these pathways, we were able to produce estimates of k_{600} during 14 sampling events.

2.4.2. Mass Balance

To incorporate the CO_2 added from groundwater, we also used a mass balance approach to calculate k. This method builds on the methods described in McDowell and Johnson (2018), which uses a mass balance model to calculate k. We have added a component which considered the CO_2 added from groundwater inputs. Briefly, excess CO_2 in the stream (ΔCO_2) was calculated as follows:

$$\Delta CO_2 = \left(\frac{(C_{US} + C_{DS})}{2} - C_{air}\right) \quad (6)$$

where C_{US} and C_{DS} are the mass of dissolved CO_2 at the up and downstream locations (g CO_2-C L^{-1}) and C_{air} is the mass of carbon at saturation (g CO_2-C L^{-1}). Flux of carbon to the atmosphere (g CO_2-C m^{-2} s^{-1}) was calculated as follows:

$$F = \left(\frac{(C_{US.corr} - (C_{DS.corr} - C_{gw.gain}))}{A}\right) \times Q \quad (7)$$

where $C_{US.corr}$ and $C_{DS.corr}$ are the up and downstream concentrations of CO_2 (g CO_2-C L^{-1}) corrected for the background concentration by subtracting the concentrations of CO_2 present before CO_2 was diffused in the

Table 2
Empirical Models Used to Estimate k_{600}

Empirical model	Reference
$k_{600} = 5937 \times \left(1 - 2.54 \times \left(\frac{v}{(9.8\bar{z})^{0.5}}\right)^2\right) \times (vS)^{0.89} \times \bar{z}^{0.58}$	Raymond et al. (2012); Equation 2
$k_{600} = vS \times 2841 + 2.02$	Raymond et al. (2012); Equation 5
$\log_{10}k_{600} = 0.319 + (2.110 \times v) + (1.026 \times \log_{10} S)$	Natchimuthu et al. (2017); model no. 2
$ln[k_{600}] = 3.10 + 0.35 \times ln[\varepsilon_D]$ for $\varepsilon_D < 0.02$	Ulseth et al. (2019)
$ln[k_{600}] = 6.43 + 1.18 \times ln[\varepsilon_D]$ for $\varepsilon_D > 0.02$	

Note. For simplicity, the standard deviations of coefficients and constants were removed.

stream; Q is flow rate (L s^{-1}); and A is surface area of the stream reach (m^2) calculated as the site length times the average width of the site. $C_{gw.gain}$ is the mass of carbon (g CO_2-C L^{-1}) added to the stream reach from groundwater. This is calculated as:

$$C_{gw.gain} = \frac{Q_{gw} \times l \times C_{gw}}{Q} \quad (8)$$

where Q_{gw} is the length-standardized discharge gain over the stream (L s^{-1} m^{-1}; see Text S1 in Supporting Information S1), l is the length of the reach (m), and C_{gw} is the concentration of CO_2 in the groundwater (g C-CO_2 L^{-1}; see Text S1 in Supporting Information S1). k_{CO2} (m d^{-1}) was then calculated by dividing the emission flux by excess CO_2:

$$k_{CO_2} = \frac{F}{\Delta CO_2} \quad (9)$$

k_{CO2} was converted to k_{600} as described in the previous section using eq3 and eq5. The mass balance method considers CO_2 added from groundwater, but assumes no consumption or inputs of CO_2 related to aquatic metabolism or shifts in the carbonate equilibrium.

2.5. Floating Chambers

On the last two sampling days (September 29 and 10 November 2023), k was also measured using a floating chamber. Within each of the three study reaches, a custom-made, tethered, floating chamber connected to the LI-7810 was used to measure gas flux at three locations. The three chamber locations were selected to capture different features of the reach that influence the rate of emission (e.g., slow riffles, pools, etc.). The chamber was deployed for three to 5 minutes for each measurement, and three replicate measurements were taken for each location. The rate of increase of gas in the chamber was used to calculate the flux (μmol m^{-2} s^{-1}) of CO_2 as follows:

$$F = \frac{P \times m \times h}{T \times R} \quad (10)$$

where P is the atmospheric pressure (Pa), R is the gas constant (8.314; Pa m^3 mol^{-1} K^{-1}), T is the air temperature (K), m is the rate of increase (μmol mol^{-1} s^{-1}), and h is the height of the chamber (m). k_{CO2} was then calculated using Equation 9 and converted to k_{600} as described in the previous section.

The three replicates at each location within each sampling reach were averaged to get a single value for each location within each reach. A weighted average was then obtained for each reach from the k_{600} values measured at the three locations. Weights were based on the proportion of the reach that each chamber measurement represented. The weights were the product of both objective measurements of slope and subjective field observations.

2.6. Empirical Models

In addition to the above methods of measuring k, four empirical models were used to estimate k based on a combination of slope, depth, flow velocity, and discharge (Table 2). Although not developed specifically for streams such as Shady Creek, two models from Raymond et al. (2012) (Equations 2 and 5) were used because they are commonly employed in regional or global upscaling calculations (e.g., Lauerwald et al. (2015), and Rocher-Ros et al. (2023)). The two models presented by Natchimuthu et al. (2017) and Ulseth et al. (2019) were also used to model k because these models were developed for small headwater streams and high-energy streams, respectively. Unique to the Ulseth model is a piecewise linear regression function based on energy dissipation. The piecewise function uses two different linear regressions to calculate k_{600} for higher energy streams ($\varepsilon_D > 0.2\ m^2\ s^{-3}$) and lower energy streams ($\varepsilon_D < 0.02\ m^2\ s^{-3}$). This breakpoint of $0.02\ m^2\ s^{-3}$ is the theoretical distinction between when the flux of systems is dominated by diffusivity in low energy streams versus bubble-mediated gas exchange in high energy streams (Ulseth et al., 2019). All models were calculated at a spatial resolution which matches the length of the site; for example, a model based on slope and velocity used the velocity and slope of the high energy reach to calculate k_{600} for the high energy site. Together, these four models provide estimates of k that would be similar to k values calculated as part of a large upscaling effort.

2.7. Impact of Method Choice on Upscaled CO_2 Flux Estimates

To assess how the method used to obtain k may impact a flux estimate for a given stream reach, CO_2 flux for the entire 197-m long study reach (Figure 1a) was calculated in three different ways that combine different models or measurements for k and different spatial resolutions of CO_2 concentration, channel slope, velocity, and discharge. Detailed methods for calculating the three different flux measurements are provided in Text S2 in Supporting Information S1. Briefly, (a) a "reference" measured CO_2 flux demonstrates a flux value obtained from intensive field measurements of k using propane and which is likely closest to the "true" flux (Figure 1b). This method also leverages multiple measurements of CO_2 concentration throughout the stream reach. (b) A "detailed" modeled CO_2 flux employs the same approach but uses empirically derived estimates of k rather than measurements made with propane. Comparing the detailed flux estimate to the reference estimate provides an idea of how empirical model selection may impact flux estimates. (c) A "simple" modeled CO_2 flux reflects perhaps the most common yet imperfect sampling procedure in the literature, using one CO_2 measurement and a modeled k value based on measurements of slope, discharge, and/or velocity from the outlet to upscale GHG fluxes to an entire catchment. The comparison of the simple and detailed modeled fluxes provides insight on how the spatial resolution of input parameters (slope, flow velocity, k, and CO_2 concentration) can impact the accuracy of flux estimates. In short, both the reference and detailed method use measurements collected from multiple locations throughout the 197-m reach, whereas the simple method uses measurements collected at only one location. Additionally, the simple and detailed methods rely on modeled values of k, whereas the reference flux uses in situ measurements of k.

2.8. Impact of Hydrology on k_{600}

The impact of hydrology on k_{600} was evaluated using energy dissipation, flow rate, and flow velocity. Simple linear regressions were used to evaluate each hydrological variable against the k_{600} values calculated using the propane tracer gas method. Regressions were calculated at both a site level and at an inter site level.

2.9. Statistical Methods

All calculations, statistical tests, and figures were done using *R* version 4.2.2 (R Core Team, 2022). Because chamber measurements were only made on two occasions, any comparisons with chamber derived k values were only between measurements that occurred on the same 2 days. Due to non-normally distributed data and small sample size (n=3), the differences in upscaled CO_2 fluxes were not formally tested for significant differences. Instead, we simply used boxplots to visually assess potential differences.

3. Results

3.1. k_{600} Estimates Across Methods

k_{600} values measure in Shady Creek were generally higher, or on the high end, than reported values globally. This pattern was true between the different methods we evaluated and across our different sites–even our

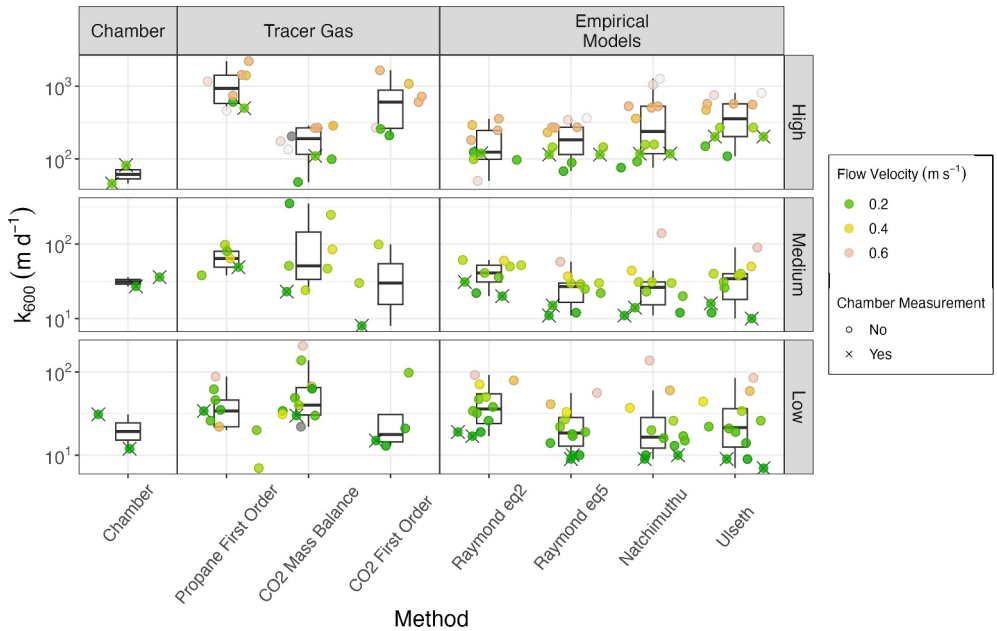

Figure 2. k_{600} values computed using in situ methods and empirical models for the three sites (rows) in Shady Creek. Individual data points are shown, and colors represent the flow velocity; gray denotes k_{600} values when velocity was not measured. Because chambers were only used on two occasions, chamber boxplots are not directly comparable to boxplots for other methods. Crossed circles denote the two data points that were collected on the same day as chamber measurements and are directly comparable.

"low" energy site had high k values, compared to globally reported sites (Ulseth et al., 2019). Across our 16 sampling campaigns, the mean k_{600} values consistently increased from the low-to high-energy reaches, with mean values across all measurement methods of 23, 43, and 399 m d^{-1} at the low, medium, and high energy reaches, respectively (Figure 2). We found substantial variations between methods at individual reaches, with the chamber-based estimates being generally lower than other estimates and the propane-based values being in the higher range. The two CO_2-based methods yielded results generally higher than the chamber but lower than the propane. The empirical models fell within the range of these other methods.

Over the 14-month sampling period, 23 k_{600} values were obtained by using propane as a tracer gas. Propane-based k_{600} estimates spanned three orders of magnitude, with the mean values increasing from the low to high energy reaches (37, 65, and 1,067 m d^{-1}, respectively).

CO_2 was successfully used to calculate k_{600} on 14 occasions using the first-order method (Equation 2 adapted to CO_2). k_{600} values ranged from 8 m d^{-1} to 1,656 m d^{-1} and averaged 363 m d^{-1}. As with propane, the mean increased from the low to high energy site (36, 46, and 681 m d^{-1}, respectively). Using the mass balance method, CO_2 yielded 29 values of k_{600} and had an average of 122 m d^{-1}, less than half of that from the first-order method. These k_{600} values followed the same increasing patterns from the low to high energy reach, but with comparatively much lower estimates for the high-energy reach (mean 64, 118, and 188 m d^{-1} in the low, medium, and high reach, respectively).

The chamber estimates of k_{600} had an average of 39 m d^{-1} across all three reaches (based on two sampling events only). Like the other in situ methods, the mean k_{600} values calculated from the chamber increased from low to high and had means of 21, 31, and 64 m d^{-1}, respectively. These values were generally lower than those obtained from the tracer gas methods.

All four empirical models generally followed an increasing pattern from the low to high energy reaches, with a few minor deviations from this pattern between the low and medium energy reaches. The Raymond Equation 2 model had average k_{600} values of 43, 41, and 144 m d^{-1} for the low, medium and high energy reaches, respectively. The Raymond Equation 5 mean values were 23, 27, and 203 m d^{-1}. The Natchimuthu model

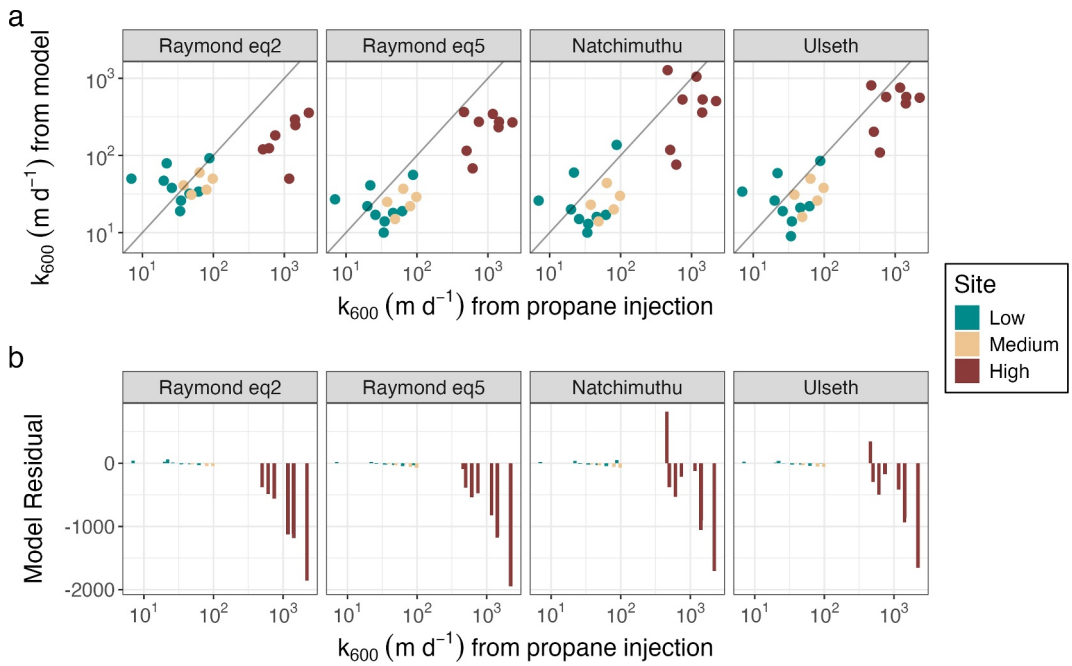

Figure 3. (a) k_{600} values from empirical models are compared to k_{600} values computed from propane injections. The solid black diagonal line represents the 1:1 line. (b) Residuals between the k_{600} values measured by propane diffusion and empirical models. A positive value represents an overestimation by the empirical model and a negative value represents and underestimation by the empirical model. Individual models are shown in separate facets and sites are denoted by color.

followed the same pattern at the low to high energy sites, but with significantly higher estimates for the high-energy reach, with mean k_{600} values of 31, 35, and 414 m d^{-1}. Finally, the Ulseth model followed the same increasing pattern with mean values of 30, 35, and 412 m d^{-1} at the low, medium, and high energy reaches, respectively.

3.2. Chamber Versus Propane and Empirical Models

Chambers were consistently higher than empirical models by a factor of 1.9 and 2.0 in the low and medium energy reaches, respectively. In the high energy reach, chambers produced lower k_{600} values by a factor of 2.2 when compared to the average of the four models. There was only one day when both chambers and propane were used to measure k_{600} and on this day, the propane measurements were higher than the chamber measurements by a factor or 1.1, 1.4, and 10.8 in the low, medium, and high energy reaches, respectively.

3.3. Propane Versus Empirical Models

When compared to measured k_{600} values computed from propane, all of the tested empirical models tended to yield lower values of k_{600} in the medium and high energy reaches, while estimates of k_{600} were relatively similar across the two methods in the low energy reach (Figure 3). On average, k_{600} values from the Raymond Equations 2 and 5, the Natchimuthu and the Ulseth models were higher than the propane derived k_{600} by factors of 2.1, 1.1, 1.2, and 1.4 respectively, in the low energy reach. However, although these average factors suggest overestimation, all four models had lower k_{600} values more than or nearly half of the time. In the medium energy reach, all four models underestimated k_{600} by factors ranging from 0.4 (Raymond, Equation 5) to 0.7 (Raymond, Equation 2). In the high energy reach, the Raymond models yielded much lower values k_{600} by factors of 0.1 and 0.3 whereas estimates of k_{600} from the Natchimuthu and Ulseth models were much closer to the measured k_{600} with underestimation factors of 0.8 and 0.6, respectively.

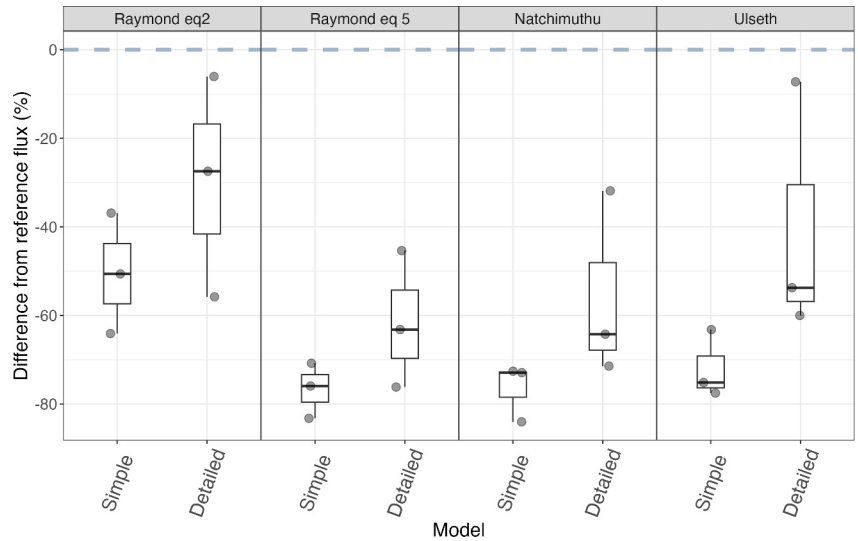

Figure 4. Difference between reference (spatially explicit, k_{600} measured using propane diffusion) and modeled fluxes (k_{600} modeled using an empirical model). A perfectly modeled flux would appear at zero (dotted gray line), whereas underestimated fluxes are negative and overestimated fluxes are positive.

3.4. Comparing Scaling Approaches

Shady Creek was consistently supersaturated in CO_2. Average concentrations were 882, 1,073, and 1,017 ppm in the low, medium, and high energy reaches, respectively (Table 1). However, out of the 16 sampling days, on only three occasions were all three sites sampled during the same day, providing the data needed to compute flux estimates for the entire stream reach. Two of these days were in the peak wet season of 2023 and had reference CO_2 fluxes of 1,280 mol d^{-1} and 639 mol d^{-1} (Figure 4). The third day was in the late dry season of 2023 and had a reference CO_2 flux of 329 mol d^{-1}. The mean reference flux across all three days was 749 mol d^{-1}.

Both implementation approaches tended to underestimate total flux compared to the reference flux when using the Raymond and Natchimuthu models (average 214 and 362 mol d^{-1} across all four models for the simple and detailed approaches, respectively). While we did not test for significant differences between flux estimates due to small sample size, choice of model proved influential in determining the upscaled fluxes (Figure 4). Across all four models, the detailed implementations had smaller differences from the true flux than the simple implementations.

3.5. Changes in k Across Flow Conditions

We found that at a given site, k_{600} was only marginally affected by changes in flow conditions (Figure 5). Energy dissipation rate (ε_D) was significantly positively correlated with propane-based k_{600} when considering data across all three sites ($p < 0.001$, $R^2 = 0.63$; Figure 5a). However, when looking at sites individually, where the stream slope was constant but flow velocity fluctuated, changes in ε_D had a limited effect on k_{600}, with nonsignificant regressions ($p > 0.20$) and R^2 values < 0.2.

Discharge had no significant impact on k_{600}. Across all sites, the linear regression between discharge and k_{600} was insignificant ($p = 0.7$, $R^2 = 0.005$; Figure 5b). Site specific linear regressions between Q and k_{600} were also insignificant (p-values > 0.2, R^2-values < 0.2). Flow velocity, across all three sites, was significantly related with increasing k_{600} ($p = 0.009$, $R^2 = 0.28$; Figure 5c), but large changes in flow velocity resulted in limited changes in k_{600}. Individual site regressions between velocity and k_{600} and ε_D and k_{600} were identical due to ε_D being a function of velocity and slope, the latter of which does not change within each site.

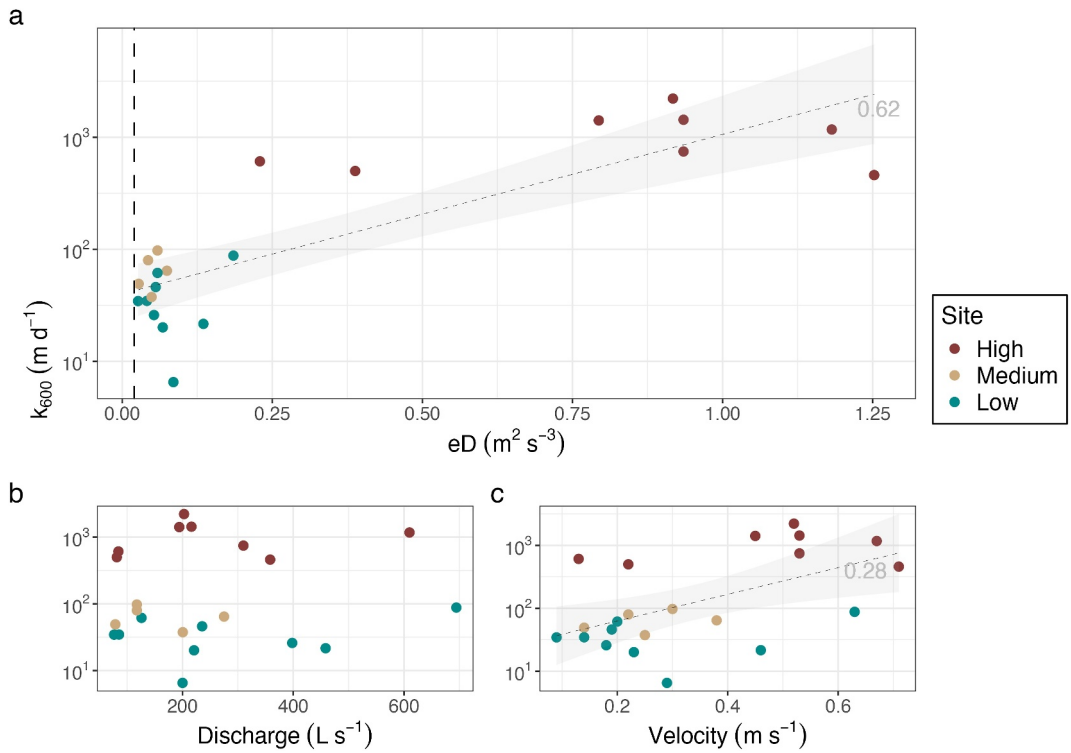

Figure 5. (a) Relationship between energy dissipation and k_{600} obtained from propane injections. The dashed vertical black line represents $\varepsilon_D = 0.02$ m^2 s^{-3}, the breakpoint identified in the Ulseth et al. (2019) model for k_{600} where gas exchange shifts from predominantly diffusion driven to predominantly bubble-mediated. (b) Relationship between stream discharge and k_{600} obtained from propane injections. (c) Relationship between flow velocity and k_{600} obtained from propane injections. In each panel, a linear regression, 95% confidence interval, and r^2-value is shown in gray, when a significant linear regression exists.

4. Discussion

Regardless of method used, our estimated k values in Shady Creek are much higher than the reported global mean of 5.7 m d^{-1} (Raymond et al., 2013). This discrepancy is unsurprising given the steep and turbulent nature of our study system. Despite significant loss of CO_2 to the atmosphere facilitated by the high k values, Shady Creek supported supersaturated levels of CO_2 along its entire length. These data demonstrate the ability of turbulent headwater streams to support large gas fluxes in the absence of source limitations. Combined, these elevated values of k and CO_2 concentrations position steep, tropical headwater streams as global hotspots for GHG emissions (Lauerwald et al., 2015; Rocher-Ros et al., 2023).

Additionally, our data showcase differences in methods commonly used to measure or estimate k and the potential implication the choice of method may have on upscaled estimations of gas flux to the atmosphere. In general, there is an inability of empirical models to accurately estimate k in the steepest sections of Shady Creek. These results make clear that, in order to improve estimates of GHG fluxes to the atmosphere from steep mountain streams, we must first improve our understanding of what measurable stream features drive k, as well as our ability to measure and predict these with greater certainty. Our study also highlights the heterogenous nature of k across small spatial scales as a result of morphological heterogeneities, and the relative stability of k across changing hydrological conditions, particularly in more turbulent reaches.

4.1. Propane as a More Practical Option Than CO_2 to Measure k

Although a variety of tracer gases (e.g., SF_6, argon, and propane) are frequently used to measure k (e.g., Hall Jr. & Madinger (2018), Ulseth et al. (2018), and Vautier et al. (2020)), the use of propane and CO_2 as tracer gases in Shady Creek yielded mixed results. Propane-derived estimates of k were consistent with our understanding of k—high energy reaches had higher values of k than low energy reaches, and k generally increased with flow velocity and discharge rate, although the influence of both parameters was only marginal. Because of the ease of use, tracer

gases such as propane remain a popular method for measuring gas exchange in streams (e.g., Knapp et al., 2019; Schelker et al., 2016; Ulseth et al., 2018).

Using CO_2 to measure k proved to be more difficult than propane, despite similar underlying theories. Because CO_2 is not inert, changes in CO_2 concentration from pathways other than flux to the atmosphere made this method unreliable without significant additional data. The first-order method, which does not account for any sources or sinks of CO_2 other than flux to the atmosphere, sometimes yielded negative concentration gradient (suggesting an influx of CO_2 from the atmosphere into the stream), something practically impossible given the supersaturated state of the stream and surface turbulence. When considering added CO_2 due to groundwater inputs (our "mass balance model"), the model yielded much more realistic, positive values, and in the case of the low and medium energy streams, values similar to the propane-derived values. In the high-energy reach, this method failed to capture the higher (>300 m d^{-1}) k values. These data do not provide evidence of why this method does not capture these higher values, but it is likely the result of poorly quantified groundwater inputs. In general, poor performance by CO_2 as a tracer is likely the result of poorly quantified natural pathways of CO_2 rather than a theoretical shortcoming of CO_2 as a tracer. These results alone should not be used as justification that mass balance methods may not work in high energy systems, but rather as a motivation to more rigorously consider natural pathways of CO_2 (groundwater inputs, in-stream metabolism, carbonate buffering) if using CO_2 as a tracer gas for measuring k. Without these considerations, and another method to validate results, this method may be unreliable. However, the benefits of using CO_2 as a tracer may be realized if used repeatedly in a single stream over a longer period time using a more permanent setup (e.g., in the case of McDowell and Johnson (2018)), where these additional pathways are better accounted for.

4.2. Chambers Failed to Capture Extreme k Values

There is no clear consensus on the ability of chambers to accurately measure gas exchange in streams and our data offer limited clarification. Some studies have reported overestimates caused by artificial turbulence (Lorke et al., 2015), others found underestimation compared to eddy covariance methods (Huotari et al., 2013), whereas some suggest that chambers can be a suitable method in headwater streams (Crawford et al., 2013). In addition to an unclear tendency to over/under estimate, chambers have been found to have a high level of uncertainty in high-energy streams (Vingiani et al., 2021). Our data show a general, though not consistent in magnitude, underestimation of chamber-derived k across sites and, to a lesser extent, across comparisons to other methods.

Different conclusions on the accuracy of chambers may be due to their small spatial and temporal footprints when compared to other methods. Because chambers measure the flux over a small surface of the stream, entire reach-scale estimates of k also have significant uncertainty when upscaling the flux from the chamber to the reach. For example, the low energy reach of Shady Creek has an average surface area of 47 m^2, yet the combined surface area of the three chamber measurements used to calculate the flux and subsequent k is only 0.06 m^2, just 0.1% of the total. The locations of these three chamber measurements, and their inability to capture any turbulent hotspots (e.g., small cascades, rapids) where k may be orders of magnitude higher than in surrounding areas with more laminar flow (Botter et al., 2022; Peruzzo et al., 2023; Vautier et al., 2020), can have drastic implications for a k value computed for the entire reach. Generally, it is likely that uncertainties from upscaling chamber measurements to entire stream reaches will increase in streams that are highly heterogeneous in morphology and surface turbulence. Within a single reach on the same sampling day, our measurements of k using a chamber could span two orders of magnitude and measurements from the same sampling location could vary by as much as a factor of six between the two sampling days, despite the flow rate and velocity only varying by a factor of ca. two. In addition to their small spatial footprint, chambers are limited in how long they can be deployed by the time needed for the chamber headspace to equilibrate with the stream. In streams with low CO_2 concentration and a high k, this could be as short as a few minutes.

Ultimately, chambers, whether tethered or drifting, may not offer a consistently appropriate method for measuring k in high energy systems. There is not a clear black and white set of conditions when chambers will and will not work, but perhaps a gray area where this transitions happens, as evidenced by the fact that chamber-based measurements did capture k well in the low energy reach, but not in the medium or high. It is also worth pointing out that all of these chamber measurements were made when flow was low (relative to flow conditions observed in the year-long study) and our conclusions might be different if chambers were used at higher flows. Given their spatial and temporal limitations, as well as the lack of consensus in the existing literature, chambers

may not offer a solution in measuring k in high energy and heterogenous streams, at least without more work that clarifies how they may influence flux measurements (Kremer et al., 2003; Lorke et al., 2015; Rawitch et al., 2021; Vachon et al., 2010). A better understanding of what types of systems chambers can accurately perform in is imperative to their continued use in studying gas exchange in streams.

4.3. Empirical Models: High Uncertainties and Need for Greater Discretion in Their Use

Computationally the easiest of the different methods evaluated in this study, empirical models offer valuable estimates of k, at least at larger spatial scales. Depending on the required input data, these models can be used with high frequency data to compute temporally explicit values of k and they can also be easily implemented over a large spatial range. However, our data suggests the benefits of these models are lost when trying to apply them at the reach scale and at very high values of k. Each model is capable of estimating an average k or a range of probable k values, but head to head comparisons with a measured reference k almost never yielded a perfect match due to their uncertainties.

The existing models we evaluated, particularly the Raymond models and Natchimuthu model, were developed with data derived from streams with predominantly lower slopes and flow velocity, and thus these models are not designed for use in short, high energy reaches like those in our study system. This is especially evident in the k estimates by Raymond Equation 2 in the high energy reach, which are much lower than the propane derived k values (Figure 3). The Natchimuthu model had similar average residuals but both over and underestimated at times; for this reason, it may perform better in scenarios when it is applied to many sites over a broad range of conditions and these over and underestimations may average out to a near-true value. The Ulseth model, which was developed using systems more similar to ours, had similar residuals in the low and medium energy reaches, but performed better in our high energy reach. Unique to the Ulseth model, and potentially the cause of its robust performance across sites, is that it uses a piece wise function to estimate k_{600} differently under conditions when bubble mediated flux may be significant (when $\varepsilon_D > 0.02\ m^2\ s^{-3}$). In every instance we used this model, flux was calculated under this scenario. The high energy site, which has low residuals, is dominated by a single cascade over rough rock substrate that generates high turbulence and bubble entrainment. In contrast, the low and medium energy sites contain small cascades similar to the one in the high energy reach, but also long pools with lower surface turbulence and minimal bubble entrainment. The Ulseth model may have performed better than the other models in this stream because of its ability to accurately account for the role bubble mediated flux plays in each site under a range of flow conditions.

These comments are not criticisms of any of these models, but rather observations on potential limitations when applying them in environments other than those they were trained in. Empirical models are built on the foundation that k increases with flow rate or velocity, yet k values in Shady Creek were only weakly positively correlated with flow velocity (Figure 5). This limited influence of hydrology on k can be thought of as a lack of transfer limitation in these systems, where high slope and streambed heterogeneity together generate high turbulence even under relatively low flow conditions, and where an increase in flow rate may only marginally increase gas exchange. This situation contrasts with low-gradient streams that are mostly transfer-limited, particularly at low flow (e.g., Solano et al. (2023) and Taillardat et al. (2022)). It is possible that this minor role of flow conditions for driving changes in k is common in steep, turbulent streams, where gas emission predominantly occurs via bubble entrapment (Ulseth et al., 2019), but additional data are needed to confirm this hypothesis.

Gas exchange can be caused by a complex interaction of stream characteristics that are difficult to encapsulate in a simple model based on hydrology and morphology alone. Turbulent streams with multiple flow paths and confluence points may experience different patterns of surface turbulence (Yuan et al., 2024). Cascades of any size lead to heterogeneous and irregular bubble entrainment and subsequent gas exchange (Peruzzo et al., 2023). Additionally, and in the case of Shady Creek, shallow streams can have additional induced surface turbulence caused by rough streambed surfaces that are not accounted using only velocity, slope, or discharge, as is the case with these models (Moog & Jirka, 1999). Ulseth et al. (2019) did find k_{600} increased with higher streambed roughness, but this metric had weak predictive power on k_{600}. Small, irregular uniformities in the stream bed may cause micro hotspots of gas emission (Botter et al., 2021) that are too irregular and random for empirical models to account for.

Models must improve their accuracy at high k values while maintaining their proven effectiveness at lower k values. A key to this is better understanding and being able to capture inter-stream variability as well as intra-

stream variability. For example, energy dissipation was able to capture differences between sites, but its predictive power significantly decreased for predicting k at a single site during different flow conditions (Figure 5). In addition to the need to improve models in a general sense, more discretion on when a specific model is appropriate is needed. This latter point is critical in constraining global gas flux estimates from streams around the world due to the large variation in stream characteristics. Global emission estimates that take the advantage of multiple empirical models for different size streams may reduce biases introduced from a single model that tends to underestimate high k_{600} values and overestimate low k_{600} values (e.g., Liu et al. (2022)). Underestimations by commonly used models for k in high-energy headwater streams that are rich in CO_2 and CH_4 can translate into high underestimations of GHG emission globally.

4.4. Implications for Upscaling Flux

Our flux estimate results demonstrate how the choice and implementation of an empirical model can influence conclusions on stream GHG flux to the atmosphere. The most striking discrepancies tended to be between the simple and detailed implementation of a specific model, rather than between empirical models in the same implementation.

Implementing these models at a higher spatial resolution (i.e., our detailed flux estimates) did improve the flux estimates (Figure 4). This is largely due to variable slope, velocity and discharge throughout the study reaches (Table 1, Figure 1 and Figure S1 in Supporting Information S1).

Shady Creek has a heterogeneous morphology at small spatial scale, and CO_2, slope, and flow rate varied consistently across a small range of values. A high energy stream that is more irregularly heterogeneous may see even larger differences in upscaled fluxes calculated using coarse and fine scale data (e.g., Botter et al. (2021)). Depending on the hydrological and morphological characteristics of a system and the scale at which a study is focused (e.g., a single reach, regional, or global), the added benefit of implementing an empirical model at a higher spatial resolution may outweigh the computational costs of doing so. While recent global upscaling studies (e.g., Liu et al. (2022); Rocher-Ros et al. (2023)) have provided critical insights into large-scale patterns of riverine GHG fluxes, they typically rely on coarse spatial estimates of water velocity, discharge or slope. Our findings highlight the potential value of incorporating higher-resolution data in such studies, particularly in headwater streams with high spatial variability, to further improve the accuracy of flux estimates.

5. Conclusions

These data showcase inconsistencies between different in situ measurements of k and a general inability of existing empirical models to capture representative k values with high confidence in high energy stream environments. These inconsistencies can lead to significant differences in upscaled gas flux estimates, even over relatively short distances (<1 km). Steep, high energy systems are either outside the scope of existing methods to measure or estimate k or have complex drivers that results in large uncertainties in modeled estimates of k. Given that high values of k can result in high rates of GHG emissions to the atmosphere, future work should aim to reduce uncertainties in models for k in turbulent streams.

High measured values of k and CO_2 in Shady Creek also demonstrate the ability of tropical streams to have large exports of GHG to the atmosphere. Steep or high energy streams are underrepresented in literature regarding k, and tropical streams are underrepresented in global GHG data collection efforts. Future work should aim to constrain k estimates in high energy, headwater streams to better capture high GHG fluxes in tropical headwaters. A critical component of constraining these estimates is exploring how k can be predicted with high certainty for the same stream over different flow conditions to better capture temporal variability of GHG fluxes.

Data Availability Statement

Data presented in this study are available on Figshare (Rexroade et al., 2024). Code used to create figures in the manuscript can be found in Supporting Information S2. All data analysis and visualization was carried out using RStudio Version 2022.12.0 + 353 (2022.12.0 + 353) using version *R* version 4.2.2 (2022-10-31) (R Core Team, 2022).

Acknowledgments
We would like to thank Adam J. Bourke, Francesco Ulloa-Cedamanos, Yihan Li, and Vanessa Solano for their assistance with fieldwork. Additionally, Matthew Northwood and Hao Wang provided technical and laboratory support. We thank two anonymous reviewers and the editor for comments and suggestions which greatly improved this manuscript. This work was funded by the Australian Research Council Grant DE220100852. M.B.W was supported by funding from the Swedish Research Council (2021-04058). Open access publishing facilitated by Charles Darwin University, as part of the Wiley - Charles Darwin University agreement via the Council of Australian University Librarians.

References

Battin, T. J., Lauerwald, R., Bernhardt, E. S., Bertuzzo, E., Gener, L. G., Hall, R. O., et al. (2023). River ecosystem metabolism and carbon biogeochemistry in a changing world. *Nature*, *613*(7944), 449–459. https://doi.org/10.1038/s41586-022-05500-8

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*(1), 180214. https://doi.org/10.1038/sdata.2018.214

Bernhardt, E. S., Heffernan, J. B., Grimm, N. B., Stanley, E. H., Harvey, J. W., Arroita, M., et al. (2018). The metabolic regimes of flowing waters. *Limnology & Oceanography*, *63*(S1), S99–S118. https://doi.org/10.1002/lno.10726

Botter, G., Carozzani, A., Peruzzo, P., & Durighetto, N. (2022). Steps dominate gas evasion from a mountain headwater stream. *Nature Communications*, *13*(1), 7803. https://doi.org/10.1038/s41467-022-35552-3

Botter, G., Peruzzo, P., & Durighetto, N. (2021). Heterogeneity matters: Aggregation bias of gas transfer velocity versus energy dissipation rate relations in streams. *Geophysical Research Letters*, *48*(17), e2021GL094272. https://doi.org/10.1029/2021GL094272

Crawford, J. T., Striegl, R. G., Wickland, K. P., Dornblaser, M. M., & Stanley, E. H. (2013). Emissions of carbon dioxide and methane from a headwater stream network of interior Alaska. *Journal of Geophysical Research: Biogeosciences*, *118*(2), 482–494. https://doi.org/10.1002/jgrg.20034

Day, T. J. (1977). Field procedures and evaluation of a slug dilution gauging method in mountain streams. *Journal of Hydrology*, *16*(2), 113–133.

Duvert, C., Butman, D. E., Marx, A., Ribolzi, O., & Hutley, L. B. (2018). CO2 evasion along streams driven by groundwater inputs and geomorphic controls. *Nature Geoscience*, *11*(11), 813–818. https://doi.org/10.1038/s41561-018-0245-y

Hall, R. O., Jr., & Madinger, H. L. (2018). Use of argon to measure gas exchange in turbulent mountain streams. *Biogeosciences*, *15*(10), 3085–3092. https://doi.org/10.5194/bg-15-3085-2018

Hall, R. O., & Ulseth, A. J. (2020). Gas exchange in streams and rivers. *WIREs Water*, *7*(1), e1391. https://doi.org/10.1002/wat2.1391

Horgby, Å., Segatto, P. L., Bertuzzo, E., Lauerwald, R., Lehner, B., Ulseth, A. J., et al. (2019). Unexpected large evasion fluxes of carbon dioxide from turbulent streams draining the world's mountains. *Nature Communications*, *10*(1), 4888. https://doi.org/10.1038/s41467-019-12905-z

Huotari, J., Haapanala, S., Pumpanen, J., Vesala, T., & Ojala, A. (2013). Efficient gas exchange between a boreal river and the atmosphere. *Geophysical Research Letters*, *40*(21), 5683–5686. https://doi.org/10.1002/2013GL057705

Jähne, B., & Haußecker, H. (1997). Air-water gas exchange. *Annual Review of Fluid Mechanics*, *30*, 443–468. https://doi.org/10.1146/annurev.fluid.30.1.443

Knapp, J. L. A., Osenbrück, K., Brennwald, M. S., & Cirpka, O. A. (2019). In-situ mass spectrometry improves the estimation of stream reaeration from gas-tracer tests. *Science of The Total Environment*, *655*, 1062–1070. https://doi.org/10.1016/j.scitotenv.2018.11.300

Kokic, J., Sahlée, E., Sobek, S., Vachon, D., & Wallin, M. (2018). High spatial variability of gas transfer velocity in streams revealed by turbulence measurements. *Inland Waters*, *8*(4), 461–473. https://doi.org/10.1080/20442041.2018.1500228

Kremer, J. N., Nixon, S. W., Buckley, B., & Roques, P. (2003). Technical note: Conditions for using the floating chamber method to estimate air-water gas exchange. *Estuaries*, *26*(4), 985–990. https://doi.org/10.1007/BF02803357

Lauerwald, R., Allen, G. H., Deemer, B. R., Liu, S., Maavara, T., Raymond, P., et al. (2023). Inland water greenhouse gas budgets for RECCAP2: 2. Regionalization and homogenization of estimates. *Global Biogeochemical Cycles*, *37*(5), e2022GB007658. https://doi.org/10.1029/2022GB007658

Lauerwald, R., Laruelle, G. G., Hartmann, J., Ciais, P., & Regnier, P. A. G. (2015). Spatial patterns in CO2 evasion from the global river network. *Global Biogeochemical Cycles*, *29*(5), 534–554. https://doi.org/10.1002/2014GB004941

Liu, S., Kuhn, C., Amatulli, G., Aho, K., Butman, D. E., Allen, G. H., et al. (2022). The importance of hydrology in routing terrestrial carbon to the atmosphere via global streams and rivers. *PNAS*, *119*(11), e2106322119. https://doi.org/10.1073/pnas.2106322119

Lorke, A., Bodmer, P., Noss, C., Alshboul, Z., Koschorreck, M., Somlai-Haase, C., et al. (2015). Technical note: Drifting versus anchored flux chambers for measuring greenhouse gas emissions from running waters. *Biogeosciences*, *12*(23), 7013–7024. https://doi.org/10.5194/bg-12-7013-2015

Maurice, L., Rawlins, B. G., Farr, G., Bell, R., & Gooddy, D. C. (2017). The influence of flow and bed slope on gas transfer in steep streams and their implications for evasion of CO2. *Journal of Geophysical Research: Biogeosciences*, *122*(11), 2862–2875. https://doi.org/10.1002/2017JG004045

McDowell, M. J., & Johnson, M. S. (2018). Gas transfer velocities evaluated using carbon dioxide as a tracer show high streamflow to be a major driver of total CO2 evasion flux for a headwater stream. *Journal of Geophysical Research: Biogeosciences*, *123*(7), 2183–2197. https://doi.org/10.1029/2018JG004388

Moog, D. B., & Jirka, G. H. (1999). Stream reaeration in Nonuniform flow: Macroroughness Enhancement. *Journal of Hydraulic Engineering*, *125*(1), 11–16. https://doi.org/10.1061/(ASCE)0733-9429(1999)125:1(11)

Mulholland, P. J., Fellows, C. S., Tank, J. L., Grimm, N. B., Webster, J. R., Hamilton, S. K., et al. (2001). Inter-biome comparison of factors controlling stream metabolism. *Freshwater Biology*, *46*(11), 1503–1517. https://doi.org/10.1046/j.1365-2427.2001.00773.x

Natchimuthu, S., Wallin, M. B., Klemedtsson, L., & Bastviken, D. (2017). Spatio-temporal patterns of stream methane and carbon dioxide emissions in a hemiboreal catchment in Southwest Sweden. *Scientific Reports*, *7*(1), 39729. https://doi.org/10.1038/srep39729

Peruzzo, P., Cappozzo, M., Durighetto, N., & Botter, G. (2023). Local processes with a global impact: Unraveling the dynamics of gas evasion in a step-and-pool configuration. *Biogeosciences*, *20*(15), 3261–3271. https://doi.org/10.5194/bg-20-3261-2023

Rawitch, M. J., Macpherson, G. L., & Brookfield, A. E. (2021). The validity of floating chambers in quantifying CO2 flux from headwater streams. *Journal of Water and Climate Change*, *12*(2), 453–468. https://doi.org/10.2166/wcc.2020.199

Raymond, P. A., & Cole, J. J. (2001). Gas exchange in rivers and estuaries: Choosing a gas transfer velocity. *Estuaries*, *24*(2), 312–317. https://doi.org/10.2307/1352954

Raymond, P. A., Hartmann, J., Lauerwald, R., Sobek, S., McDonald, C., Hoover, M., et al. (2013). Global carbon dioxide emissions from inland waters. *Nature*, *503*(7476), 355–359. https://doi.org/10.1038/nature12760

Raymond, P. A., Zappa, C. J., Butman, D., Bott, T. L., Potter, J., Mulholland, P., et al. (2012). Scaling the gas transfer velocity and hydraulic geometry in streams and small rivers. *Limnology and Oceanography: Fluids and Environments*, *2*(1), 41–53. https://doi.org/10.1215/21573689-1597669

R Core Team. (2022). *R: A language and environment for statistical computing (version 4.2.2)*. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Regnier, P., Friedlingstein, P., Ciais, P., Mackenzie, F. T., Gruber, N., Janssens, I. A., et al. (2013). Anthropogenic perturbation of the carbon fluxes from land to ocean. *Nature Geoscience*, *6*(8), 597–607. https://doi.org/10.1038/ngeo1830

Rexroade, A., Wallin, M., & Duvert, C. (2024). Supporting Data for "Measuring gas transfer velocity in a steep tropical stream: method evaluation and implications or upscaling" (Version 3) [Data set]. *Figshare*. https://doi.org/10.6084/m9.figshare.26893507.v3

Rocher-Ros, G., Stanley, E. H., Loken, L. C., Casson, N. J., Raymond, P. A., Liu, S., et al. (2023). Global methane emissions from rivers and streams. *Nature*, *621*(7979), 530–535. https://doi.org/10.1038/s41586-023-06344-6

Schelker, J., Singer, G. A., Ulseth, A. J., Hengsberger, S., & Battin, T. J. (2016). CO2 evasion from a steep, high gradient stream network: Importance of seasonal and diurnal variation in aquatic pCO2 and gas transfer. *Limnology & Oceanography*, *61*(5), 1826–1838. https://doi.org/10.1002/lno.10339

Solano, V., Duvert, C., Birkel, C., Maher, D. T., García, E. A., & Hutley, L. B. (2023). Stream respiration exceeds CO evasion in a low-energy, oligotrophic tropical stream. *Limnology & Oceanography*, *68*(5), 1132–1146. https://doi.org/10.1002/lno.12334

Taillardat, P., Bodmer, P., Deblois, C. P., Ponçot, A., Prijac, A., Riahi, K., et al. (2022). Carbon dioxide and methane dynamics in a peatland headwater stream: Origins, processes and implications. *Journal of Geophysical Research: Biogeosciences*, *127*(7), e2022JG006855. https://doi.org/10.1029/2022JG006855

Tsivoglou, E. C. (1967). Tracer measurement of stream reaeration. Washington DC: Federal water pollution control administration.

Ulseth, A. J., Bertuzzo, E., Singer, G. A., Schelker, J., & Battin, T. J. (2018). Climate-induced changes in spring snowmelt impact ecosystem metabolism and carbon fluxes in an alpine stream network. *Ecosystems*, *21*(2), 373–390. https://doi.org/10.1007/s10021-017-0155-7

Ulseth, A. J., Hall, R. O., Boix Canadell, M., Madinger, H. L., Niayifar, A., & Battin, T. J. (2019). Distinct air–water gas exchange regimes in low- and high-energy streams. *Nature Geoscience*, *12*(4), 259–263. https://doi.org/10.1038/s41561-019-0324-8

Vachon, D., Prairie, Y. T., & Cole, J. J. (2010). The relationship between near-surface turbulence and gas transfer velocity in freshwater systems and its implications for floating chamber measurements of gas exchange. *Limnology & Oceanography*, *55*(4), 1723–1732. https://doi.org/10.4319/lo.2010.55.4.1723

van Vliet, M. T. H., Franssen, W. H. P., Yearsley, J. R., Ludwig, F., Haddeland, I., Lettenmaier, D. P., & Kabat, P. (2013). Global river discharge and water temperature under climate change. *Global Environmental Change*, *23*(2), 450–464. https://doi.org/10.1016/j.gloenvcha.2012.11.002

Vautier, C., Abhervé, R., Labasque, T., Laverman, A. M., Guillou, A., Chatton, E., et al. (2020). Mapping gas exchanges in headwater streams with membrane inlet mass spectrometry. *Journal of Hydrology*, *581*, 124398. https://doi.org/10.1016/j.jhydrol.2019.124398

Vingiani, F., Durighetto, N., Klaus, M., Schelker, J., Labasque, T., & Botter, G. (2021). Evaluating stream CO2 outgassing via drifting and anchored flux chambers in a controlled flume experiment. *Biogeosciences*, *18*(3), 1223–1240. https://doi.org/10.5194/bg-18-1223-2021

Wallin, M. B., Öquist, M. G., Buffam, I., Billett, M. F., Nisell, J., & Bishop, K. H. (2011). Spatiotemporal variability of the gas transfer coefficient KCO2 in boreal streams: Implications for large scale estimates of CO 2 evasion. *Global Biogeochemical Cycles*, *25*(3), GB3025. https://doi.org/10.1029/2010GB003975

Wanninkhof, R. (1992). Relationship between wind speed and gas exchange over the ocean. *Journal of Geophysical Research*, *97*(C5), 7373–7382. https://doi.org/10.1029/92JC00188

Wanninkhof, R., Mulholland, P. J., & Elwood, J. W. (1990). Gas exchange rates for a first-order stream determined with deliberate and natural tracers. *Water Resources Research*, *26*(7), 1621–1630. https://doi.org/10.1029/WR026i007p01621

Witherspoon, P. A., & Saraf, D. N. (1965). Diffusion of methane, ethane, propane, and n-butane in water from 25 to 43°. *The Journal of Physical Chemistry*, *69*(11), 3752–3755. https://doi.org/10.1021/j100895a017

Yuan, S., Lin, J., Tang, H., Zhu, Y., Ran, Q., Constantinescu, G., & Gualtieri, C. (2024). Near-surface turbulent dissipation at a laboratory-scale confluence: Implications on gas transfer. *Environmental Fluid Mechanics*, *24*(6), 1099–1122. https://doi.org/10.1007/s10652-023-09964-8

Erratum

The originally published version of this article contained some errors. Figures 2, 3, and 4 have been replaced and the supporting information has been added. In the eighth sentence of the Abstract, "choice of model was more influential than the spatial resolution of model implementation" has been corrected to the following: "implementing them at finer spatial resolutions yielded flux estimates closer to measured fluxes." In the first sentence of the second paragraph of Section 2.3, "x is the length of the study reach (m), C_x" has been corrected to the following: "x is the length of the study reach (m). C_x." The third and fourth sentences of Section 2.9 have been corrected to the following: "Due to nonnormally distributed data and small sample size (n=3), the differences in upscaled CO_2 fluxes were not formally tested for significant differences. Instead, we simply used boxplots to visually assess potential differences." In the third sentence of the first paragraph of Section 3.1, "42, 66, and 413" has been corrected to the following: "23, 43, and 399." The last sentence of the same paragraph has been corrected to the following: "The empirical models fell within the range of these other methods." The second sentence of the fifth paragraph of Section 3.1 has been corrected to the following: "Finally, the Ulseth model followed the same increasing pattern with mean values of 30, 35, and 412 m d^{-1} at the low, medium, and high energy reaches, respectively." The first sentence of Section 3.2 has been corrected to the following: "Chambers were consistently higher than empirical models by a factor of 1.9 and 2.0 in the low and medium energy reaches, respectively." The second sentence of the same paragraph has been corrected to the following: "In the high energy reach, chambers produced lower k_{600} values by a factor of 2.2 when compared to the average of the four models." In the same paragraph, the third sentence has been removed. Section 3.3 has been corrected to the following: "When compared to measured k_{600} values computed from propane, all of the tested empirical models tended to yield lower values of k_{600} in the medium and high energy reaches, while estimates of k_{600} were relatively similar across the two methods in the low energy reach (Figure 3). On average, k_{600} values from Raymond Equations 2 and 5, the Natchimuthu and the Ulseth models were higher than the propane derived k_{600} by factors of 2.1, 1.1, 1.2, and 1.4 respectively, in the low energy reach. However, although these average factors suggest overestimation, all four models had lower k_{600} values more than or nearly half of the time. In the medium energy reach, all four models underestimated k_{600} by factors ranging from 0.4 (Raymond, Equation 5) to 0.7 (Raymond, Equation 2). In the high

energy reach, the Raymond models yielded much lower values k_{600} by factors of 0.1 and 0.3, whereas estimates of k_{600} from the Natchimuthu and Ulseth models were much closer to the measured k_{600} with underestimation factors of 0.8 and 0.6, respectively." The second paragraph of Section 3.4 has been corrected to the following: "Both implementation approaches tended to underestimate total flux compared to the reference flux when using the Raymond and Natchimuthu models (average 214 and 362 mol d^{-1} across all four models for the simple and detailed approaches, respectively). While we did not test for significant differences between flux estimates due to small sample size, choice of model proved influential in determining the upscaled fluxes (Figure 4). Across all four models, the detailed implementations had smaller differences from the true flux than the simple implementations." The last sentence of the same paragraph has been removed. The fourth sentence of the second paragraph of Section 4.3 has been corrected to the following: "The Ulseth model, which was developed using systems more similar to ours, had similar residuals in the low and medium energy reaches, but performed better in our high energy reach." The fifth sentence of the same paragraph has been corrected to the following: "Unique to the Ulseth model, and potentially the cause of its robust performance across sites, is that it uses a piece wise function to estimate k_{600} differently under conditions when bubble mediated flux may be significant (when $\varepsilon_D >$ 0.02 m^2 s^{-3})." The last sentence of the same paragraph has been corrected as follows: "The Ulseth model may have performed better than the other models in this stream because of its ability to accurately account for the role bubble mediated flux plays in each site under a range of flow conditions." The first and second sentences of the first paragraph of Section 4.4 has been corrected to the following: "Our flux estimate results demonstrate how the choice and implementation of an empirical model can influence conclusions on stream GHG flux to the atmosphere. The most striking discrepancies tended to be not between the simple and detailed implementation of a specific model, rather than between empirical models in the same implementation." The last sentence of the same paragraph has been removed. The first sentence of the second paragraph of Section 4.4 has been corrected as follows: "Implementing these models at a higher spatial resolution (i.e., our detailed flux estimates) did improve the flux estimates (Figure 4)." The second sentence of the same paragraph has been corrected as follows: "This is largely due to variable slope, velocity and discharge throughout the study reaches (Table 1, Figure 1 and Figure S1 in Supporting Information S1)." In the same paragraph the last sentence has been removed. The first sentence of the third paragraph of Section 4.4 has been corrected to the following: "Shady Creek has a heterogeneous morphology at small spatial scale, and CO_2, slope, and flow rate varied consistently across a small range of values." The second sentence of the same paragraph has been corrected to the following: "A high energy stream that is more irregularly heterogeneous may see even larger differences in upscaled fluxes calculated using coarse and fine scale data "(e.g., Botter et al. (2021))." The third sentence of the same paragraph has been removed. The following has been added to the end of the paragraph: "While recent global upscaling studies (e.g., Liu et al. (2022); Rocher-Ros et al. (2023)) have provided critical insights into large-scale patterns of riverine GHG fluxes, they typically rely on coarse spatial estimates of water velocity, discharge or slope. Our findings highlight the potential value of incorporating higher-resolution data in such studies, particularly in headwater streams with high spatial variability, to further improve the accuracy of flux estimates." The reference for Rexroade et al., 2024 has been corrected to the following: Rexroade, A., Wallin, M., & Duvert, C. (2024). Supporting Data for "Measuring gas transfer velocity in a steep tropical stream: method evaluation and implications or upscaling" (Version 3) [Data set]. *Figshare*. https://doi.org/10.6084/m9.figshare.26893507.v3. This may be considered the authoritative version of record.