Thermal-infrared remote sensing of stream temperatures at multiple spatial scales


1 Department of Earth and Space Sciences, University of Washington, Seattle, WA, 98195, USA
2 Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, 98195, USA

Abstract

Stream temperature is an important indicator of water quality, particularly where endangered fish populations are sensitive to elevated water temperature. Regional assessment of stream temperatures from the ground is limited by sparse sampling in both space and time. Remotely sensed thermal-infrared (TIR) images can make spatially distributed measurements of the radiant skin temperature of streams. We quantify and discuss the accuracy and uncertainty limits to recovering stream temperatures in the Pacific Northwest for a range of stream widths (10 – 500 m), and TIR pixel sizes (5 – 1000 m) from remotely sensed airborne and satellite TIR images. Among locations with more than three pixels across the stream, the difference between the image and in-stream temperature measurements averaged 1.15 °C with a standard error (SE) of 0.16 °C, while the corresponding uncertainty (spatially weighted standard deviation in temperature from the image subsets) for these locations averaged ±0.27 °C (SE = 0.03 °C). For streams with one to three pixels across, mixing with bank elements increased the average difference to 2.25 °C (SE = 0.28 °C) and the uncertainty to ±0.39 °C (SE = 0.07 °C). For a fraction of a pixel across the stream the average difference was 7.65 °C (SE = 1.17 °C) and the uncertainty averaged ±0.54 °C (SE = 0.13 °C). These results show that reliable satellite TIR measurement of stream temperatures is limited to large rivers (currently, ~180-m across for Landsat ETM+), unless novel unmixing algorithms are used effectively.
INTRODUCTION

Stream temperature is an important regional indicator of water quality that is used for land-use monitoring (Brown and Krygier, 1970; Chen et al., 1998). In the Pacific Northwest, endangered fish populations are sensitive to elevated water temperatures (Washington Department of Ecology, 1998), and stream temperatures must meet regulatory guidelines and quality standards, which aim to create a healthy environment for salmon populations. For example, the United States Environmental Protection Agency (EPA) water-quality standards specify that the 7-day moving average of maximum daily temperature should not exceed 13 °C during salmon spawning and incubation (Environmental Protection Agency, 2003). The National Water Quality Assessment Program focuses on summer peak-temperatures (United States Geological Service, 1999). Migrating salmon can survive in cool-water refugia, even when temperatures at measurement locations are at or above the recommended maximums (Torgersen et al., 1999). Spatially extensive measurements of stream temperature are necessary to locate these refugia, identify the location of ground- and surface-water inputs to the stream channel, and to identify thermal pollution sources. Ultimately we are interested in the bulk, or kinetic, temperature of the water ($T_k$), since this is both biologically important and is the definition of temperature used for management purposes.

Regional assessment of stream temperatures has been limited by sparse sampling of temperatures in both space and time. Streams temperatures have typically been measured using a network of in-stream gauges, which record the temporal change at specific locations. For example, the State of Washington recorded water quality conditions at 76 stations within the Puget Lowlands ecoregion, which contains 12,721 km of streams and rivers (Washington Department of Ecology, 1998). Such gages are typically located in larger streams and rivers, and give limited information about the spatial distribution of stream temperatures. Remote sensing of stream temperatures is an attractive alternative if temperatures can be determined with suitable and known accuracies and uncertainties.
The remote sensing of surface water temperatures using measurements of emitted thermal-infrared (TIR) radiation (8 to 14 µm) can provide spatially distributed values of radiant temperature (T_r) in the ‘skin’ layer of the water (top 100 µm). This is a well-established practice, particularly in oceanography, where daily observations of regional and global sea-surface temperatures (SST) are now made (Anding and Kauth, 1970; Emery & Yu, 1997; Kilpatrick et al., 2001). In the terrestrial environment the TIR remote sensing of surface water temperatures has focused on lakes (LeDrew and Franklin, 1985; Bolgrien & Brooks, 1992), water-cooling ponds (Chen et al., 2003), and streams (Torgerson et al., 2001; Li et al., 1997). However, there is still a lack of research addressing accuracy and uncertainty issues associated with these measurements.

In terms of remote sensing, a stream is a more complex environment than a lake because it is usually much smaller and is often not resolved at the resolution (pixel size) of the TIR data. Streams often have a complex morphology of braided channels, islands and in-stream rocks, and vary greatly in hydrological characteristics such as ground-water inputs, water velocity and turbulence. They also vary in the amount of bank vegetation, which reduces the full stream width that the sensor can resolve. Emitted radiation from the near-bank environment of vegetation and rocks can reach the sensor directly or be scattered from the stream surface.

Individual local studies show the applicability of remote sensing to stream temperature estimation. Li et al. (1997) for example, used forward-looking infrared (FLIR) videography to monitor thermal gradients in the John Day River, Oregon. Torgersen et al. (2001) used fine pixel-size (0.2-0.4 m) airborne TIR images to evaluate the accuracy of T_r measurements, and found that the remotely sensed T_r were within 0.5°C of in-situ measurements of T_k. Torgersen et al. (2001) identified several issues with the thermal remote sensing of streams, including reflected TIR radiation, vertical thermal stratification in the stream, and thermal boundary-layer effects at the water surface. Using images where the stream was fully resolved, they concluded that fine pixel-size measurements of stream temperature are useful for studying
fine-scale spatial variation and patterns in stream temperature related to hydrological features such as ground-water inputs.

The present research augments these fine-scale studies of Torgersen et al. (2001) and Li et al. (1997) by quantifying how the image pixel size limits the accuracy and uncertainty associated with stream temperature measurements of streams that are not fully resolved. This is necessary to determine the applicability of airborne- and satellite-derived TIR measurements for regional assessment of stream temperatures, particularly since approximately 75% of the total length of streams in the conterminous USA is made up of small channels (first and second order) (Leopold et al., 1964). Remote sensing of stream temperatures involves a trade-off between having fine enough pixel sizes to identify spatial patterns and coarse enough data so that the cost of flying large areas is not prohibitive. Satellite observations cover a large area, but they are commonly too coarse to resolve the stream channel, whereas airborne- and ground-based platforms produce data at very fine pixel-sizes over narrow swath widths but limited area.

In recent years there has been increasing interest in the issue of scale in remote sensing (Dungan et al., 2002; Andrefouet et al., 2002). Spatial scale is important to the TIR remote sensing of stream temperatures since the pixel size determines how big an object must be on the ground before we can resolve it, as well as the fraction of pixels that mix bank and water. We address the impact of spatial scale on the accuracy and uncertainty of the stream temperature measurements by using of the number of pixels across the stream to combine stream width and pixel size.

In this paper, we determine the limits to determining stream temperatures using remotely sensed ground-, airborne- and satellite-based TIR measurements in the Pacific Northwest, and various combinations of steam widths (10 – 500 m) and pixel sizes (5 – 1000 m). We quantify the associated accuracies and uncertainties, and discuss sources of this uncertainty, including the image processing methodology, the complexity of the stream environment, and the problem of mixing in large pixels.
2 APPROACH

2.1 Study Areas

Image and field data were collected in the summers of 2001 and 2002 for a range of stream widths and environmental conditions in three basins in Washington State, USA (Figure 1). The Green River basin is located in a marine ecoregion (Bailey, 1995). The flow of the Green River is frequently controlled, and it originates in a forested area before passing through increasingly rural and urbanized environments further downstream. Stream widths at the study sites range from 8 to 50 m. The Yakima River basin is located in a temperate desert ecoregion (Bailey, 1995). The flow of streams and rivers in this basin are frequently controlled, and irrigated cropland areas alternate with a sparsely vegetated arid environment. Stream widths at the study sites range from 15 to 120 m. The third study area is situated along the Columbia River, and climatic and hydraulic conditions are similar to the Yakima study area. River widths at the study sites range from 420 to 450 m.

Figure 1 about here

2.2 Image Data

The NASA EOS MODIS sensor, on the same Terra platform as ASTER, has ten TIR bands (6.54 to 14.39 µm) with a pixel size of 1000 m (http://modis.gsfc.nasa.gov/about/specs.html), and an estimated TIR radiometric precision, or noise-equivalent temperature difference (NE\(\Delta T\)), of 0.05 °C at 27 °C. We used one Terra-MODIS scene acquired over the Columbia River study sites, and concurrent with a Landsat scene. Data were obtained as radiance at sensor (http://edcimswww.cr.usgs.gov/) (Table 1).

The NASA EOS Advanced Spaceborne Thermal Emission and Reflection (ASTER) radiometer (Kahle et al., 1991; Yamaguchi et al., 1998), located on the Terra spacecraft, has five TIR bands (8.12 to 11.65 µm), with 90-m pixel size, and an estimated NE\(\Delta T\) of \(\leq 0.3^{\circ}\)C at 27 °C (Gillespie et al., 1998). We used four ASTER scenes, obtained over the Green River and Yakima River study sites (Table 1). Data were obtained as radiance at sensor (http://edcimswww.cr.usgs.gov/).
The Landsat ETM+ sensor on Landsat-7 is in the same orbit as Terra, which allows images to be acquired ~20 minutes apart. The single TIR band on Landsat (10.40 to 12.50 µm) has a pixel size of 60 m (http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook_toc.html) and an NEΔT of 0.22 °C at 7 °C (Barsi et al., 2003). We used four Landsat ETM+ scenes, obtained over the Green River, Yakima River, and Columbia River study sites (Table 1). Data were obtained as radiance at sensor (http://landsat7.usgs.gov/).

Airborne TIR images were collected over the Green and Yakima study areas using the MODIS/ASTER (MASTER) airborne simulator (Hook et al., 2001). MASTER has ten TIR bands (10.15 to 11.45 µm) with an NEΔT during our data imaging that ranged from 0.46 to 1.67 °C for all TIR bands, and the sensor scans ±43° from nadir. The MASTER sensor was flown on a King Air B200 fixed-wing aircraft at two altitudes of ~2000 and ~6000 m, which gives approximate pixel sizes of 5- and 15- m, respectively. The MASTER team processed the data using onboard calibration targets and an onboard GPS to record the aircraft’s location, and they made the data available as radiance at sensor (http://masterweb.jpl.nasa.gov/). We used fourteen MASTER 5-m scenes and ten MASTER 15-m scenes, obtained over the Green River and Yakima River study sites (Table 1).

The FLIR, from FLIR Systems Inc. (Portland, Oregon), is a solid state IQ Series Model 812, and it records TIR radiance in a band with a radiometric precision of 0.1 °C at 273 °C. We used two FLIR images taken current with the MASTER data (August 27th 2001) from a bridge overlooking the Green River, one looking down-stream and one with a near-nadir viewing geometry. Pixel sizes from these images range from 5 to 50cm. Data from one TIR band (10.15-11.45 µm) were calibrated by a linear regression between raw image values from stirred water targets at different temperatures, and concurrent with T_r and T_k measurements.

**Table 1 about here**
2.3 Ground Validation Data

Ground validation data were collected concurrent with the remotely sensed TIR images, mostly during summer peak-temperature conditions. We measured in-stream temperatures at sites across the three study areas, using Optic StowAway loggers (http://www.onset.com), which have a precision of 0.16 °C, and an accuracy of ±0.2 °C. Loggers located 30cm above the stream bed recorded measurements at 15-30 minute intervals. Temperatures were extracted from the time-series of logger data ($T_{Logger}$) for the overpass times of the airborne and satellite sensors. If $T_{Logger}$ was within five minutes of the overpass time, then that measurement was used, otherwise $T_{Logger}$ averaged the two measurements closest in time. These $T_{Logger}$ measurements form the main control dataset for the images, under the assumption that there is a consistent relationship between $T_k$ and $T_r$.

Additional $T_k$ measurements in the near-surface layer of the stream were collected at selected validation sites within the study area using Checktemp digital thermometers (Hanna Instruments, Inc., http://www.hannainst.com/), which have a precision of 0.1 °C and an accuracy of ± 0.3 °C. These measurements were used to investigate any vertical temperature difference between the near-surface layer and the stream bottom where the $T_{Logger}$ measurements were made.

We did not compare ground-based measurements of $T_r$ to the remotely sensed $T_r$ because these measurements were sparse and variable in space and time. Rapid fluctuations of up to 3 °C are apparent in measurements made using a hand-held broadband (8 –18 um) TempTestr-IR radiometer, (http://www.4oakton.com/) at the Columbia River study site on 17th September 2002, concurrent with a Landsat overpass. These fluctuations are attributed to instrument instability, but are complicated by wind-driven surface evaporative cooling. During this period $T_k$ measurements made in the near-surface layer of the river varied only within a range of 0.2 °C.

Auxiliary datasets for each study site included digital elevation models, photographs of the study sites, archived aerial orthophotos, and field records of vegetation characteristics. The stream width at each study site was calculated from the orthophotos after averaging five transects across the stream. For
Thermal-infrared remote sensing of stream temperatures at multiple spatial scales

comparison with modeled results, atmospheric profiles of temperature and humidity and pressure were collected during the overpasses of the MASTER sensor from a radiosonde launched at the Yakima River (Roza Campground, August 28th 2001) study areas.

2.4 Image processing

Raw TIR images were calibrated to ground-leaving TIR radiance prior to the extraction of image subsets at the study sites (Figure 2). Small errors in the radiometric correction can have a large effect on the resulting stream temperatures. For example, for the average stream temperature of 17.2 °C in our study area with a blackbody emission at 10 µm of 8.43 Wm⁻²µm⁻¹sr⁻¹, increases in the estimated temperature of 0.1 and 1.0 °C will result from increases in the estimated radiances of 0.02 and 0.15 Wm⁻²µm⁻²sr⁻¹ respectively.

The Penn State/NCAR mesoscale model (MM5) (Dudhia, 1993) was used to estimate the atmospheric conditions for airborne and satellite data across all study areas and dates. This information was then used to calculate the spectrally varying atmospheric transmissivity (τ) and path radiance (Lp) with the MODTRAN 4.0 (Ontar Corporation, 2001; Berk et al., 1989) radiative transfer model. The contribution of upwelling reflected sky irradiance is small. These correction factors were adjusted to the spectral response function of each band and sensor and used to compensate the data for atmospheric effects (Kay, 2002).

The emissivity for each TIR band was calculated from published values of the reflectance of distilled water (http://speclib.jpl.nasa.gov/), with adjustments for the spectral response function of each band and sensor. While the emissivity of the water can vary with the amount of suspended sediment (Salisbury and Aria, 1992), the streams used in the present study did not carry a large near-surface suspended sediment load during the summer study period, and suspended sediment was therefore not considered in the emissivity calculation.
Of the images collected, the ones used in this study were only those with concurrent ground validation data and modeled atmospheric parameters. If multiple images were available at a particular pixel size of the airborne data, then to maintain radiometric consistency between scenes, the flight lines in which the study sites were viewed nearest to nadir were used. Although Kay (2002) found that fully resolved lake pixels from airborne data in the Green River study areas were not affected by an off-nadir viewing geometry, we examine the effect of viewing geometry on the TIR remote sensing of streams across all study areas.

When comparing the remotely sensed $T_r$ to $T_{Logger}$ it is necessary to use more than one pixel to capture the natural variation in the measurements. Since we wished to compare measurements between images with a range of pixel sizes (5-1000 m), we extracted image subsets of 1-km$^2$ around the locations of all in-stream loggers with concurrent validation data (Table 1). Visual examination of the 1-km$^2$ image subsets did not reveal any obvious gradients around the logger locations. Sixty-seven 1-km$^2$ image subsets were used, with all necessary validation data, in addition to one smaller FLIR image (Figure 8).

Due to the difficulty of discriminating the stream channel from the background of vegetation and shadow, both manual identification and spectral classification of VNIR and TIR bands and auxiliary data was used to locate stream centerline pixels. Manual quality control was increasingly necessary as the pixel size of the images became larger, and misclassifications of the stream using the automatic techniques increased. With the larger pixel sizes of the satellite data it was difficult to identify the stream channel, and sometimes as few as three image pixels were identified as stream channel within the 1-km$^2$ image subset. Class statistics of the mean, standard deviation, minimum, and maximum temperatures were calculated for these stream centerline pixels. $T_r$ for each pixel and band was calculated using Planck's Law, for both raw data and radiometrically corrected radiance values. Corresponding $T_{Logger}$ at these sites and dates averaged 17.2 °C ± 5.2 °C.

**Figure 2 and Figure 3 about here**
2.5 Best estimate of temperature

When only one TIR band was available, the best estimate of stream temperature for the 1-km² image subset ($T_{\text{Best}}$) was the average of the radiometrically corrected $T_r$ from all stream centerline pixels classified as centerline stream. When multiple TIR bands were available, the ‘best’ bands were selected to avoid contamination from atmospheric effects or bands with sensor noise. The standard deviation across multiple bands, as distinct from the within-band spatial variability, was used as an independent assessment of the quality of the processed data. For MASTER the best bands were bands 43, 44, 46, 47 and 48 (Kay, 2002), and for ASTER the best bands were bands 11 and 13.

Temperature measurements from these best bands will vary due to instrument sensitivity, measurement error, and errors in image processing methods such as atmospheric compensation, so it is not possible to choose a single best band. An additional measure of band quality is the within-band standard deviation of $T_r$ in that band, which assumes that the actual stream temperature does not change within the 1-km² image subset. We therefore weight each band $k$ with $w_k$, the inverse of the variance of the radiometrically corrected $T_r$ for all centerline pixels measured at that band, normalized proportional to the sum of all non-normalized weights so that all $w_k$ add to one. We use these normalized weights to calculate $T_{\text{Best}}$ for each 1-km² image subset as the weighted average across the best bands.

Quantitative assessment of the uncertainty in stream temperature measurements is necessary to compare measurements to regulatory standards of precision and accuracy. We first assess the accuracy of our remotely sensed TIR measurements by calculating $(T_{\text{Best}} - T_{\text{Logger}})$. The uncertainty in the remotely sensed $T_r$ results from a combination of the instrument sensitivity, errors introduced by the complexity of the stream environment and uncertainty introduced during the radiometric correction. This uncertainty is integrated in the measurements of $T_r$ made at each pixel and band within each 1-km² image subset.

We quantify this uncertainty within each 1-km² image subset by calculating the spatially weighted variance for $T_{\text{Best}}$, which is the sum of the within-subset variance of $T_r$ in each band $k$ weighted by $w_k^2$. 


The weights are the same as described previously. The variance of $(T_{\text{Best}} - T_{\text{Logger}})$ therefore follows as the sum of this within-subset spatially weighted variance and the variance in $T_{\text{Logger}}$, and its square root is our uncertainty estimate ($U_{\text{Best}}$). In practice the variance in $T_{\text{Logger}}$ is small, so the within-image spatially weighted variance will dominate $U_{\text{Best}}$.

### 3 RESULTS

#### 3.1 Comparing $T_{\text{Best}}$ to $T_{\text{Logger}}$

$T_{\text{Best}}$ over-predicts $T_{\text{Logger}}$ (Figure 4) for nearly all data used in this study. Prior to radiometric correction (Figure 4a), the data fall into three groups, divided by study site and sensor. The first group contains cooler temperatures from the Green River MASTER data obtained on August 25th 2001, and the second group is from the warmer Yakima River MASTER temperatures measured on August 28th 2001. These groups are consistent with the different atmospheric conditions on each day, particularly since the radiometric corrections have corrected for differences between the image dates, locations, and sensors (Figure 4b).

Our data are for many sites with known mixed pixels and Figures 4 and 5 display the loose relationship that exists between $T_{\text{Logger}}$ and $T_{\text{Best}}$ when all sites are considered together. If we group the data based on the number of pixels across the stream, A) more than 3, B) 1 to 3, and C) a fraction of one, there is a definite systematic relationship between $T_{\text{Best}}$ and $T_{\text{Logger}}$. Where there are more than three pixels across the stream (Figure 5b), a least-squares linear regression between $T_{\text{Best}}$ and $T_{\text{Logger}}$ shows that this relationship is near-linear, with over 93% of the variance being explained by this model. Where the stream is resolved by one to three pixels there is a much weaker relationship between $T_{\text{Logger}}$ and $T_{\text{Best}}$ (Figure 5c), and there is no relationship when the stream is resolved by a fraction of a pixel (Figure 5d), since this predominately contains data from satellite images, and the number of pixels, identified as water is subsequently small.

**Figure 4 and Figure 5 about here**
3.2 Quantifying accuracy and uncertainty

Our ability to resolve the stream is determined by the number of pixels across its width, which integrates the stream width and the size of the image pixels across spatial scales. For all combinations of stream width and pixel size within our study areas, there is a systematic pattern in how \((T_{\text{Best}} - T_{\text{Logger}})\) varies with the number of pixels across the stream (Figure 6a). As seen previously, \(T_{\text{Best}}\) overestimates \(T_{\text{Logger}}\) for all sites except for the fully resolved Landsat ETM+ data from the Columbia River study site, which almost exactly predicts \(T_{\text{Logger}}\).

Summary statistics for all groups (A, B, C) are shown in Table 2. For group-A, \((T_{\text{Best}} - T_{\text{Logger}})\) ranges from –0.51 to 2.18 °C, with an average of 1.15 °C (SE = 0.16 °C). For group-B, the systematic relationship between \(T_{\text{Logger}}\) and \(T_{\text{Best}}\) becomes weaker, and \((T_{\text{Best}} - T_{\text{Logger}})\) decreases with the number of pixels across the stream. This is not unexpected since these study sites are more likely to contained mixed pixels, and for group-C, most pixels are mixed and the relationship breaks down entirely.

A similar pattern and grouping is displayed for our uncertainty estimate, \(U_{\text{Best}}\) (Figure 6b, Figure 5a). For group-A, \(U_{\text{Best}}\) ranges from 0.19 to 0.33 °C with an average of 0.27 °C (SE = 0.03). Summary statistics for all groups are shown in Table 2. For group-C there are some satellite-based sites where \(U_{\text{Best}}\) actually decreases, which is an artifact of the small number of pixels that were identified as stream and their correspondingly low within-band standard deviation. Sites in groups B and C may still have values of \((T_{\text{Best}} - T_{\text{Logger}})\) and \(U_{\text{Best}}\) as low as sites in group-A, which suggests they are from sites where the image processing and classification procedure identified fewer mixed pixels. The MASTER data are more likely to fall into this category, since these pixels are smaller than those of the satellite images.

No obvious patterns were apparent when data from Figure 6a were examined by logger elevation or location. This rules out the different temperatures from the three different eco-climates as a significant influence on the results. There were no discernable patterns with the sensor observation angle, which is important since it was possible that the higher observation angles from some of the MASTER data would influence the accuracy of remotely sensed Tr.
4 DISCUSSION

4.1 Accuracy assessment

These results quantify the expected relationship between stream temperature and pixel size. As the number of pixels across the stream decreases (groups B, C), the ability to resolve the stream diminishes, and the uncertainty associated with the measurements increases compared to those from fully resolved stream pixels (group-A). When the stream is resolved by fewer than three pixels (groups B, C), both the accuracy and uncertainty associated with these measurements decreased until it was not possible to resolve the stream. Although it is possible to use techniques such as spectral mixture analysis (Gillespie, 1992; Gustafson et al., 2003) to extract stream temperatures for cases where fewer than three pixels resolve the stream, there is a limit beyond which it is not possible to resolve the stream at all (Figure 7).

At almost all study sites, \((T_{\text{Best}} - T_{\text{Logger}})\) is > 0, even when the stream was well resolved. The most likely explanation is thermal stratification between warmer surface water and cooler water deeper in the stream, which is likely during warm summer conditions with low water levels. This hypothesis is supported by \(T_k\) measurements made in the near-surface layer at a limited number of locations at the three study sites, which were on average 0.47 °C ± 0.23 higher than \(T_{\text{Logger}}\). Hook et al. (2003) investigated the difference between \(T_k\) and the \(T_r\) in the skin layer using four monitoring stations permanently moored on Lake Tahoe, California–Nevada. Hook et al. (2003) found that there is a ‘skin effect,’ or a difference between \(T_k\) and \(T_r\) in the skin layer, which varies over the diurnal cycle which is attributed to solar heating and lower wind speeds in the morning. Daytime measurements of the \(T_r\) in the skin layer were on average 0.11°C lower than \(T_k\) (S.E. 0.40 °C), compared to 0.46°C cooler for night time measurements, (S.E. of 0.18 °C).
Other sources of this difference between $T_{\text{Logger}}$ and $T_r$ include the mixing of warm sub-pixel objects such as in-stream rocks that were not identified during image processing. We also expect that the skin layer would be cooler than the near-surface water layer because of evaporative cooling, but we do not see this effect in our data, likely because it is much smaller than the magnitude of $(T_{\text{Best}} - T_{\text{Logger}})$.

Although the atmospheric correction is a source of absolute error in thermal remote sensing, the atmospheric correction we used was developed for a range of sensors and study sites and days, and tested independently (Kay, 2002). Because the atmospheric correction is applied uniformly across each image, its quality only affects the absolute accuracy and not the uncertainty of remotely sensed stream-temperature measurements. This makes it possible to use non-atmospherically corrected stream temperatures to assess relative spatial patterns within a single image, assuming that the stream is resolved by a sufficient number of pixels and that the temperature range across the image is not large.

**Figure 7 about here**

### 4.2 Sources of Uncertainty

As our results show, our uncertainty measurement, $U_{\text{Best}}$, varies with the number of pixels in a similar pattern as values of $(T_{\text{Best}} - T_{\text{Logger}})$. Some of $U_{\text{Best}}$ is due to $T_{\text{Best}}$ being an aggregate of multiple pixels within a 1-km$^2$ subset, whereas $T_{\text{Logger}}$ is measured at the single location but is a temporal aggregate over the 15-30 minute recording interval of the logger. Although sources of this uncertainty include the sensitivity, and precision of the imager and the ground instruments, as well as natural variation in $T_r$, the influence of the bank environment is seen in the increase in $U_{\text{Best}}$ when the stream is resolved by fewer pixels. Additional sources of this uncertainty, which we will discuss, include thermal scattering from the near-bank environment, the role of shade, surface effects, and observation angle, and problems with mixed stream pixels.

The near-bank environment includes objects with a wide variety of temperatures and emissivities, including bark, branches, dead grasses, and leaves, and soil, sand, and rocks which are commonly much
warmer than the water and exposed during low summer water levels. Some of the TIR radiation emitted by objects in the near-bank environment will reach the stream surface, either directly or through multiple scattering from objects in the scene which warms the near-surface layer of the stream. TIR radiation emitted from near-bank objects may also pass directly into the path of the sensor, or it may be multiply scattered or reflected from other surfaces into the sensor, both of which will increase the measured temperature. However, this effect is small compared to the temperature of the water, and it is not discernable at the temperatures of our sites and the sensitivity of the instruments. For example, with a typical summer temperature of 25 °C for a tree at the Green River study area, and assuming emissivities at 10 µm for pine (0.9738) and distilled water (0.9885) (http://speclib.jpl.nasa.gov/), the 9.50 Wm⁻²µm⁻¹sr⁻¹ emitted by the tree has a lambertian reflection from a placid water surface of 0.04 Wm⁻²µm⁻¹sr⁻¹, or 0.2 °C (at 25 °C).

When bank objects shade the water surface, shaded regions that are observed by the TIR sensor will be cooler than areas where the skin layer is warmed by solar irradiance. Examination of airborne data of lake and streams from the MASTER sensor show that it is not possible to discern shadows in the TIR data at the temperature of the scene and the sensitivity of the instruments, although these shadow are visible in the VNIR data.

The skin layer of the water surface can be disrupted to expose the surface water, as has been seen in studies of ocean waves (Jessup et al., 1997). The emissivity of the stream will also change in response to features such as riffles and foam, which roughen the water surface. For high observation angles (> 70°) rough water will have a higher emissivity than placid water at the same observation angle, and will therefore appear warmer (Masuda et al., 1988). For low observation angles up to 30° from nadir, the decrease in spectrally variable emissivity and temperatures is small (< 0.1 °C at 10 µm for distilled water at 17.2 °C and look angles between 0 and 30°), but for very high observation angles (> 70°) emissivity and temperature are significantly lower due to Fresnel's reflection (Masuda et al., 1988; Ishiyama et al., 1995). At our study sites the observation angle is not a major source of variability since all but four of the image
subsets has sensor observation angles of less than 30° from nadir, and those four were all less than 38° from nadir.

The low observation angles used for our image subsets also mean that the viewing geometry makes it unlikely that emitted TIR radiation from warm bank objects such as vegetation will be reflected from into the sensor. Such reflections can be seen in the ground-based FLIR image taken at the Green River study site, which was observed with a high observation angle (Figure 8b). The reflection of TIR radiation emitted from bank vegetation is clearly visible in the image as apparently higher temperatures on the stream surface in a spatial pattern that mirrors the shape of the tree silhouette. This pattern is not apparent for near-nadir viewing geometry (not shown), although this image has a narrow field of view.

Individual sensor characteristics are an additional source of uncertainty. For example, the adjacently effect noted for ASTER (F.D. Palluconi, *Pers. Comm*, 2003.) occurs across strong thermal contrasts such as found between stream and bank, and can result in thermal contamination between adjacent pixels. Investigation of ASTER images of beaches suggests that for fully resolved water this is in the order of 1-2 °C. However, for the ASTER images used in the present study, all streams are un-resolved at the 90-m pixel size, and this adjacency effect is small compared to the larger contribution of bank mixing. In the MASTER data, visual inspection rules this out as a major source of uncertainty.

**Figure 8 about here**

### 4.3 Issues of spatial scale

The spatial-, temporal-, and spectral scales of the images, and the spatial- and temporal- scales of the observed object (i.e., stream) all have an effect on our ability to use remotely sensed data to estimate physical properties on the ground. Since a remotely sensed image is a data snapshot of a time-dependent variable such as temperature, the infrequency of repeat measurements is a practical limit. To establish a base line for the variability, ground validation data should be measured with greater frequency during the time that the remotely sensed measurements are obtained. The spectral resolution of TIR data is also
important since the single TIR value measured for a spectral band will be the result of wavelength
dependant factors such as atmospheric and emissivity effects that must be applied using the sensor filter
functions. Landsat ETM+ for example has only one wide thermal band (10.4 – 12.5 µm), which unlike
ASTER and MASTER gives no repeat measurements to constrain the temperature estimates.

Although temporal and spectral resolution are important in making remotely sensed TIR stream
temperature measurements, we have seen that the pixel size of the images and the stream width limit how
well we can determine stream temperatures. Even if the stream is large and completely resolved in the
TIR data, with coarser pixel sizes the measured variability among multiple pixels is expected to decrease
due to aggregation, which will make measurements approach a mean value. This loss of information is not
important if the goal of the stream study is to generate a generalized along-stream temperature profile, but
features such as the location of thermal pollution sources may be lost.

Larger pixels are also more likely to be mixtures of different surface materials with different
temperatures and emissivities. If the pixel is assumed to be pure water and one emissivity and temperature
is assumed for the entire mixed pixel, this can result in inaccurate water temperatures due to both the
unknown radiance contributions from the material, and its unknown emissivity. The difference in
emissivities will vary by wavelength (Figure 9a). For example, at 10 µm, dry grass has a lower emissivity
than water, but this is reversed at 13 µm. If it is suspected that the pixels identified as stream will be
mixed with a known object such as sandbars or boulders in the stream, it is possible to minimize the
difference between the emissivity of water and the mixed constituent by choosing an appropriate TIR
band where the emissivity of distilled water and grass are similar, such as MASTER band 48 (11.01-11.76
µm) or ASTER Band 14 (10.95 - 11.65 µm). The more mixed the pixel (i.e., lower percentage of water),
the greater the error in the estimated radiance (Figure 9b).

If the water pixel is very mixed and contains a large temperature range, then non-linearity in
Planck’s Law will result in a spectrally variable error in the calculated temperature. For the general case
this concept is well described by Rastetter (1992). For a small temperature range this effect is small, but the temperature difference can be much larger during high-summer conditions when the water levels are low, and sand and rocks are exposed at the stream edge. For example, for blackbody emittance at 10 µm from an object that is 50% at 15 °C and 50% at a higher temperature, the temperature difference within the pixel must exceed 16 °C before the averaged temperature is too low by 0.3 °C. Such temperature contrasts are only likely where the pixel is large enough to contain multiple objects with different temperature, in which case mixing will dominate.

Figure 9 about here

5 SUMMARY AND CONCLUSIONS

We used TIR measurements of stream temperatures from four airborne and satellite sensors with a range of pixel sizes (5 – 1000 m) to observe a range of stream widths in the Pacific Northwest (10 to 500 m), and quantify the uncertainties involved in the remote sensing of the complex stream environment.

When the stream was resolved by more than three pixels, the difference between the image and validation data (T_{Best} - T_{Logger}) ranged from –0.51 to 2.18 °C, with an average of 1.15 °C, and U_{Best} (spatially weighted standard deviation in the image subset), ranged from 0.19 to 0.33 °C with an average of 0.27 °C. For more than three pixels across the stream the systematic relationship between T_{Logger} and T_{Best}, allows T_{Best} to be used as a surrogate for T_{Logger} for relative assessment of changes in stream temperature within a single image. For fewer than three pixels across the stream, mixed pixels limited the ability to resolve the stream, and the corresponding accuracy decreased and uncertainty increased.

In addition to mixing with the stream bank, sources of this uncertainty include the sensor sensitivities, the precision of instruments used for ground validation, and natural variation in T_k. The complexity of the stream environment, particularly vegetation, affects both the accuracy and uncertainty of TIR stream temperature measurements. Choosing a near-nadir viewing geometry for the sensor will reduce many of these sources of uncertainty.
Fine pixel-size TIR data are necessary to determine stream temperatures for smaller streams, which are often the streams that have management significance. Satellite TIR measurements of stream temperatures are limited to temperature recovery for large rivers. For smaller rivers it is necessary to use finer pixel sizes from airborne sensors, but the expense of calibrating and processing these images is prohibitive over large areas. When the streams are very narrow or overhung by vegetation as is typical in the Pacific Northwest, discriminating the stream surface is very difficult. While expensive to collect and process, the strength of fine pixel-size TIR data may be in qualitative assessment such as monitoring thermal pollution in high-priority areas, and in limited surveys, which locate areas for more intensive monitoring using in-stream temperature gages and repeated TIR imaging.

The EPA defines "thermally impaired" water segments (including lakes, reservoirs and rivers) based on their Total Maximum Daily Load (TMDL), which is the pollution load (i.e., temperature) that a water body can carry without violating water quality standards (http://oaspub.epa.gov/waters/). Across the USA ~ 4% of thermal TMDL sites, and for EPA Region 10 (OR, WA, ID, AK) ~ 2% of thermal TMDL sites, are wide enough to span three 90-m ASTER pixels (~ 270-m stream).

Our findings offer guidelines for the size of streams that are suitable for monitoring using remotely sensed TIR data and the expected uncertainty that can be achieved for different sensors. These results show that satellite TIR measurements of stream temperatures is limited to temperature recovery for rivers that are large enough to be resolved by at least three pixels, (currently, ~180-m with Landsat ETM+).

6 ACKNOWLEDGEMENTS

EPA Science to Achieve Results (STAR) R827675-01-0

Field assistance from University of Washington volunteers, and NASA-JPL
7 REFERENCES CITED


### Table 1  Sensor and image details. The number of scenes used is noted in brackets for each image date.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>MODIS</th>
<th>ASTER</th>
<th>Landsat ETM+</th>
<th>MASTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal pixel size of TIR bands</td>
<td>1000 m</td>
<td>90 m</td>
<td>60 m</td>
<td>5-15 m</td>
</tr>
<tr>
<td>Number of TIR bands (bands used)</td>
<td>10 (1)</td>
<td>5 (B11, B13)</td>
<td>1 (high gain)</td>
<td>10 (B43, B44, B46, B47, B48)</td>
</tr>
<tr>
<td>Detector sensitivity (NE∆T)</td>
<td>0.05 °C at 27 °C</td>
<td>0.3 °C at 27 °C</td>
<td>0.22 °C at 7 °C</td>
<td>0.46 – 0.71 °C, for the bands used</td>
</tr>
<tr>
<td>Study area</td>
<td>Columbia River</td>
<td>Green, Yakima</td>
<td>Green, Yakima, Colombia</td>
<td>Green, Yakima</td>
</tr>
<tr>
<td>Number of images</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>24</td>
</tr>
<tr>
<td>Number of 1-km² image subsets</td>
<td>3</td>
<td>8</td>
<td>14</td>
<td>42</td>
</tr>
</tbody>
</table>

1 [http://modis.gsfc.nasa.gov/about/specs.html](http://modis.gsfc.nasa.gov/about/specs.html)

2 Gillespie et al., 1998

Table 2: Parameters describing the distribution of the accuracy and uncertainty associated with temperature measurements at the logger locations, from Figure 6. Results are grouped by the number of pixels across the stream (A, B, C), for a, b) \( T_{\text{Best}} - T_{\text{Logger}} \) (\(^\circ\text{C}\)) and c, d) \( U_{\text{Best}} \). Data are further grouped by whether all radiometric corrections have been applied (a, c), or no radiometric corrections have been applied (b, c).

<table>
<thead>
<tr>
<th>Number of pixels across stream</th>
<th>( T_{\text{Best}} - T_{\text{Logger}} ) ((^\circ\text{C}))</th>
<th>( U_{\text{Best}} ) ((^\circ\text{C}))</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>σ</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>a)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fraction of 1</td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>σ</td>
<td>SE</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.65</td>
<td>0.16</td>
<td>22.16</td>
<td>5.60</td>
<td>1.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.25</td>
<td>1.04</td>
<td>6.85</td>
<td>1.36</td>
<td>0.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.15</td>
<td>-0.51</td>
<td>2.18</td>
<td>0.72</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>b)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>σ</td>
<td>SE</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.54</td>
<td>-0.03</td>
<td>16.43</td>
<td>4.08</td>
<td>0.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.06</td>
<td>-0.17</td>
<td>4.73</td>
<td>1.19</td>
<td>0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.27</td>
<td>-5.02</td>
<td>2.08</td>
<td>1.95</td>
<td>0.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>c)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>σ</td>
<td>SE</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.54</td>
<td>0.00</td>
<td>1.14</td>
<td>0.63</td>
<td>0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.39</td>
<td>0.20</td>
<td>0.76</td>
<td>0.35</td>
<td>0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.27</td>
<td>0.19</td>
<td>0.33</td>
<td>0.13</td>
<td>0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>σ</td>
<td>SE</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.42</td>
<td>0.00</td>
<td>0.87</td>
<td>0.47</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.31</td>
<td>0.17</td>
<td>0.63</td>
<td>0.28</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.22</td>
<td>0.15</td>
<td>0.28</td>
<td>0.11</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
9 FIGURE HEADINGS

Figure 1  Overview map showing the location of the three study areas in Washington State, USA.

Figure 2  Flowchart of image-processing approach.

Figure 3  Examples of 1-km² image subsets at three study sites and a range of TIR pixel sizes from multiple sensors.

Figure 4  $T_{\text{Best}}$ vs.$T_{\text{Logger}}$ for all sites, symbolized by the type of image sensor and the study site. a) With no radiometric corrections applied, b) with all radiometric corrections applied.

Figure 5  $T_{\text{Best}}$ vs.$T_{\text{Logger}}$, grouped by the number of pixels across the stream. a) For all sites and with error bars of $\pm U_{\text{Best}}$, b) for sites with > 3 pixels across the stream, c) for sites with 1 to 3 pixels across the stream, d) for sites with a fraction of 1 pixel across the stream.

Figure 6  a) $(T_{\text{Best}} - T_{\text{Logger}})$ vs the number of pixels across the stream, b) $T_{\text{Uncertainty}}$ versus the number of pixels across the stream. Both graphs are symbolized by the type of image sensor.

Figure 7  Schematic showing the relationship between stream width, pixel size and the number of pixels across the stream. When we can resolve the stream by many pixels, it is possible to investigate spatial patterns. When we can resolve the stream by only a few pixels, sub-pixel analysis is necessary to extract stream temperatures.

Figure 8  a) Photograph of the Green River study site looking downstream from a bridge across the river, taken at a high observation angle, b) FLIR TIR image of the same scene and viewing geometry, showing reflected thermal emissions from bank vegetation, c) TIR image of the same scene, with a near-nadir viewing geometry.
Figure 9  a) Emissivity in the 8-14 µm TIR region for distilled water, dry grass, and basalt (http://speclib.jpl.nasa.gov/). b) Modeling the effect on temperature from changing the proportion of bank material in a pixel (30 °C) mixed with distilled water (15 °C) while assuming a single emissivity of distilled water. At 11.7 µm, dry-grass and distilled water have similar emissivities, and the water component is correctly estimated. At 12.5 µm dry-grass has a higher emissivity than distilled water, so the temperature of the water component is over-predicted. At 10.0 µm basalt has a lower emissivity than distilled water, so the temperature of the water component is under-predicted.
Yakima River study sites (●)

Green River study sites (●)

Columbia River study sites (▲)
FLIR
Pixel size 0.003 m
Ground-based

MASTER
Pixel size 5 & 15m
Aircraft

Landsat ETM+
Pixel size 60 m
Satellite

ASTER
Pixel size 90 m
Satellite

MODIS
Pixel size 1000 m
Satellite

Radiance at Sensor
(Wm⁻²·sr⁻¹·µm⁻¹)

Atmospheric Correction

Radiance at Surface
(Wm⁻²·sr⁻¹·µm⁻¹)

Emissivity Correction

Corrected Radiance
(Wm⁻²·sr⁻¹·µm⁻¹)

Calculate temperature using Planck's Law

MODTRAN
Radiative Transfer Model

Atmosphere Profiles

MM5
Atmospheric Model

ε
Distilled Water

Auxiliary & VNIR data

Selection of 1-km study sites

Automatic and Manual Classification

Radiant Temperature
for centre stream pixels (°C)

Calculate class statistics

Best estimate of Radiant Temperature for each 1km study site

Figure 2
Gr een River  
(YR6)

Y a k im a  
River  
(YAK4)

C o l u mb ia  
River  
(COL2)

<table>
<thead>
<tr>
<th>5m</th>
<th>15m</th>
<th>60m</th>
<th>90m</th>
<th>1000m</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="GR6_5m" /></td>
<td><img src="image2" alt="GR6_15m" /></td>
<td><img src="image3" alt="GR6_60m" /></td>
<td><img src="image4" alt="GR6_90m" /></td>
<td><img src="image5" alt="GR6_1000m" /></td>
</tr>
<tr>
<td><img src="image6" alt="YAK4_5m" /></td>
<td><img src="image7" alt="YAK4_15m" /></td>
<td><img src="image8" alt="YAK4_60m" /></td>
<td><img src="image9" alt="YAK4_90m" /></td>
<td><img src="image10" alt="YAK4_1000m" /></td>
</tr>
<tr>
<td><img src="image11" alt="COL2_5m" /></td>
<td><img src="image12" alt="COL2_15m" /></td>
<td>Landsat</td>
<td><img src="image13" alt="COL2_90m" /></td>
<td><img src="image14" alt="COL2_1000m" /></td>
</tr>
</tbody>
</table>

Figure 3
Figure 4

(a) $T_{	ext{Logger}}$ (no radiometric correction) vs. $T_r$ (°C)

(b) $T_{	ext{Logger}}$ vs. $T_{\text{Best}}$ (°C)

Legend:
- Circle: Green River study site
- Square: Yakima River study site
- Triangle: Columbia River study site
Figure 5
Figure 6

a)  

\[ T_{\text{Best}} \ (\degree C) \]

Number of pixels across the stream

b)  

\[ U_{\text{Best}} \ (\degree C) \]

Number of pixels across the stream

- MASTER 5m
- MASTER 15m
- Landsat ETM+
- ASTER
- MODIS
Figure 9

(a) Emissivity (0 to 1) vs. Wavelength (um)

- Distilled water
- Dry grass
- Basalt

(b) Water Temperature (°C) vs. % Water

- Dry-grass (12.5 µm)
- Basalt (10.0 µm)
- Dry-grass (11.7 µm)