Contents lists available at SciVerse ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Assessing the accuracy of blending Landsat–MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection

Irina V. Emelyanova [a],[*], Tim R. McVicar [b], Thomas G. Van Niel [a], Ling Tao Li [b], Albert I.J.M. van Dijk [b]

[a] CSIRO Land and Water, Private Bag No. 5, Wembley, 6913, WA, Australia
[b] CSIRO Land and Water, GPO Box 1666, Canberra, 2601, ACT, Australia

ABSTRACT

Blending algorithms model land cover change by using highly resolved spatial data from one sensor and highly resolved temporal data from another. Because the data are not usually observed concurrently, unaccounted spatial and temporal variances cause error in blending algorithms, yet, to date, there has been no definitive assessment of algorithm performance against spatial and temporal variances. Our objectives were to: (*i*) evaluate the accuracy of two advanced blending algorithms (STARFM and ESTARFM) and two simple benchmarking algorithms in two landscapes with contrasting spatial and temporal variances; and (*ii*) synthesise the spatial and temporal conditions under which the algorithms performed best. Landsat-like images were simulated on 27 dates in total using the nearest temporal cloud-free Landsat–MODIS pairs to the simulation date, one before and one after. RMSD, bias, and r^2 estimates between simulated and observed Landsat images were calculated, and overall variance of Landsat and MODIS datasets were partitioned into spatial and temporal components. Assessment was performed over the whole study site, and for specific land covers. Results addressing objective (*i*) were that: ESTARFM did not always produce lower errors than STARFM; STARFM and ESTARFM did not always produce lower errors than simple benchmarking algorithms; and land cover spatial and temporal variances were strongly associated with algorithm performance. Results addressing objective (*ii*) indicated ESTARFM was superior where/when spatial variance was dominant; and STARFM was superior where/when temporal variance was dominant. We proposed a framework for selecting blending algorithms based on partitioning variance into the spatial and temporal components and suggested that comparing Landsat and MODIS spatial and temporal variances was a practical method to determine if, and when, MODIS could add value for blending.

Crown Copyright © 2013 Published by Elsevier Inc. All rights reserved.

1. Introduction

Fundamental to all sensors are the inherent data characteristics in the spectral, radiometric, spatial and temporal domains. The data characteristics are defined by the extent, resolution and density in each of the four domains (Table 1), which can be specified for each remotely sensed data type providing a high-level summary. Ultimately all single-sensor monitoring systems are constrained by the domain characteristic elements that constitute this sensor-specific data framework (e.g., Ludwig et al., 2007; Phinn, 1998). To overcome this constraint, many monitoring systems combine: (*i*) remotely sensed data with dynamic physically-based models (model–data fusion; e.g., Heinsch et al., 2006; Renzullo et al., 2008; van Dijk & Renzullo, 2011); or (*ii*) two or more remotely sensed data sources with complementary data frameworks (data–data fusion, e.g., blending). The remainder of this paper focuses solely on the latter.

Capturing spatial and temporal dynamics is a key issue for many remotely sensed based monitoring systems, so a number of data blending algorithms have recently been developed that concentrate on these two domains. Examples include: (*i*) Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM, Gao et al., 2006); (*ii*) the Enhanced version of STARFM (ESTARFM, Zhu et al., 2010); (*iii*) Spatial Temporal Adaptive Algorithm for mapping Reflectance Change (STAARCH, Hilker et al., 2009a); (*iv*) a downscaling unmixing algorithm based on a linear mixing model to produce Landsat-like images from the Medium Resolution Imaging Spectrometer (Zurita-Milla et al., 2009); and (*v*) a semi-physical approach using a BRDF spectral model to fuse MODIS and Landsat data (Roy et al., 2008). These algorithms aim to blend high spatial resolution imagery (e.g., Landsat TM/ETM+) with high temporal density imagery (e.g., MODIS, AVHRR or MERIS) resulting in imagery with high spatial resolution and high temporal density characteristics to better capture spatio-temporal dynamics.

Here we primarily compare the STARFM and ESTARFM algorithms to determine if ESTARFM is always an improvement over STARFM. We chose STARFM and ESTARFM from the previously mentioned

---

Table 1
The data-framework comprised of the domain-characteristic elements; modified from (McVicar et al., 2002). EMS is an abbreviation for the electro-magnetic spectrum.

Domain	Characteristic		
	Extent	Resolution	Density
Spectral	Portion(s) of the EMS being sampled	Bandwidth(s)	Number of bands in a particular portion of the EMS[a]
Radiometric	Dynamic range of radiances (minimum and maximum radiance per band)	Change in radiance due to change by one digital number	Number of bits used across the dynamic range of radiances
Spatial	Area covered by the image	Pixel size acquired	Complete[b]
Temporal	Recording period over which the data are available[c]	Period of data acquisition[d]	Satellite repeat characteristics[e]

[a] For example, hyperspectral sensors (e.g., Hyperion) have higher spectral density than broadband instruments (e.g., Landsat TM/ETM+) though they sample similar EMS extents.

[b] This contrasts with the low spatial density of ground-based sampling, for example, meteorological stations.

[c] For some remotely sensed systems (e.g., AVHRR and Landsat TM) data have been recorded near-continuously for ~30 years.

[d] For remotely sensed images this is a matter of seconds, which contrasts with meteorological data such as the daily rainfall totals.

[e] For some applications using optical (i.e., reflective and thermal) data the availability of cloud-free images is an important consideration. Whereas the satellite repeat characteristics do not change, cloud cover will change the effective temporal density of a site over time.

blending algorithms due to: (*i*) their minimum input requirements (i.e., ancillary land cover data are not required); (*ii*) standard image pre-processing requirements (i.e., co-registration and atmospheric correction); (*iii*) availability of software; (*iv*) ease of implementation; (*v*) their ability to handle both homogeneous and heterogeneous landscapes; and (*vi*) their wide acceptance within the remote sensing community (see Section 4). We also compare STARFM and ESTARFM against two relatively simple (and computationally inexpensive) benchmarking algorithms (described in Emelyanova et al., 2012; their Appendix A). Comparing the four algorithms, in a context of quantifying the relative contributions of spatial and temporal variances helps identify conditions when a particular algorithm outperforms the rest.

Blending algorithms attempt to account for spatial and temporal changes in land cover that have not been observed concurrently in the imagery. This means that unresolved temporal and spatial variances will cause error in blending algorithms. Yet, to date, there has been no definitive assessment of algorithm performance against changing spatial and temporal variance. Our paper extends previous research by addressing this issue. Our specific objectives are to: (*i*) evaluate the accuracy of STARFM and ESTARFM and the two benchmarking algorithms in two landscapes with contrasting spatial and temporal variances; and (*ii*) synthesise the spatial and temporal conditions under which the algorithms perform best and develop a framework of why this is so.

2. Materials and methods

2.1. Study sites and data

The Coleambally Irrigation Area study site ('Coleambally' herein) is a rice based irrigation system located in southern New South Wales (NSW, Australia; 34.0034°E, 145.0675°S) which has been extensively used for time series remote sensing research (Van Niel & McVicar, 2003, 2004a,b; Van Niel et al., 2003, 2005). Over Coleambally 17 cloud-free Landsat–MODIS (L–M) pairs were available for the 2001–2002 austral (i.e., southern hemisphere) summer growing season (Table 2). Coleambally is 43 km in the north–south (N–S) direction and 51 km in the east–west (E–W) direction; an overall area of 2193 km^2 (1720 columns by 2040 lines at 25 m resolution). At Coleambally all Landsat imagery were ETM+ acquired by Landsat-7, and were atmospherically corrected using MODTRAN4 (Berk et al., 1999) as outlined in Van Niel and McVicar (2004a). Coleambally is entirely in the E–W overlap of two adjacent paths in the Landsat World Reference-2 system (i.e., Paths/Rows 92/84 and 93/84) allowing for an ~8-day repeat cycle (Van Niel & McVicar, 2004a).

The Lower Gwydir Catchment study site ('Gwydir' herein) is located in northern NSW (149.2815°E, 29.0855°S), where 14 cloud-free L–M pairs were available from April 2004 to April 2005. Gwydir is 80 km N–S and 68 km E–W covering 5440 km^2 (3200 columns by

2720 lines at 25 m resolution); all Landsat imagery were TM acquired by Landsat-5 and were atmospherically corrected using Li et al.'s (2010) algorithm.

For both study sites MODIS Terra MOD09GA Collection 5 data were used. After geometrically transforming the MODIS data, these were oversampled to 25 m resolution using a nearest neighbour algorithm to match the Landsat data resolution. To achieve sub-pixel accuracy prior to being input to the blending algorithms, each L–M pair was co-registered to within one 25 m pixel (0.05 of the MODIS 500 m pixel) by defining the optimal offset required to maximise the correlation function between the two images (using http://idlastro.gsfc.nasa.gov/ftp/pro/image/correl_optimize.pro accessed 28 September 2011). The resultant optimal offsets applied to the oversampled MODIS data in order to co-register them to Landsat are provided in Emelyanova et al. (2012; their Table 3).

The study sites have contrasting spatial and temporal dynamics. Fig. 1a–d shows that Coleambally exhibits temporal dynamics associated with crop phenology over a single growing season within the irrigation area as evident by the vegetation index and wetness index responses (the dashed green and blue lines, respectively in Fig. 1d). The surrounding dryland agricultural and woodlands are less variable through time, thus the whole study site summary of vegetation greenness and wetness is less variable through time (solid green and blue lines, respectively in Fig. 1d). Due to the relatively small field sizes in the irrigation area, Coleambally can be considered a more spatially heterogeneous site (Fig. 1a–c), as will be shown in detail later. The temporal extent of the Gwydir data is approximately 1 year, including

Table 2
Dates for the cloud-free Landsat–MODIS pairs for the two study sites. For Coleambally DSSD are the number of days since start of dataset (DSSD) from 30 September 2001, and for Gwydir this value represents the number of DSSD from 31 March 2004.

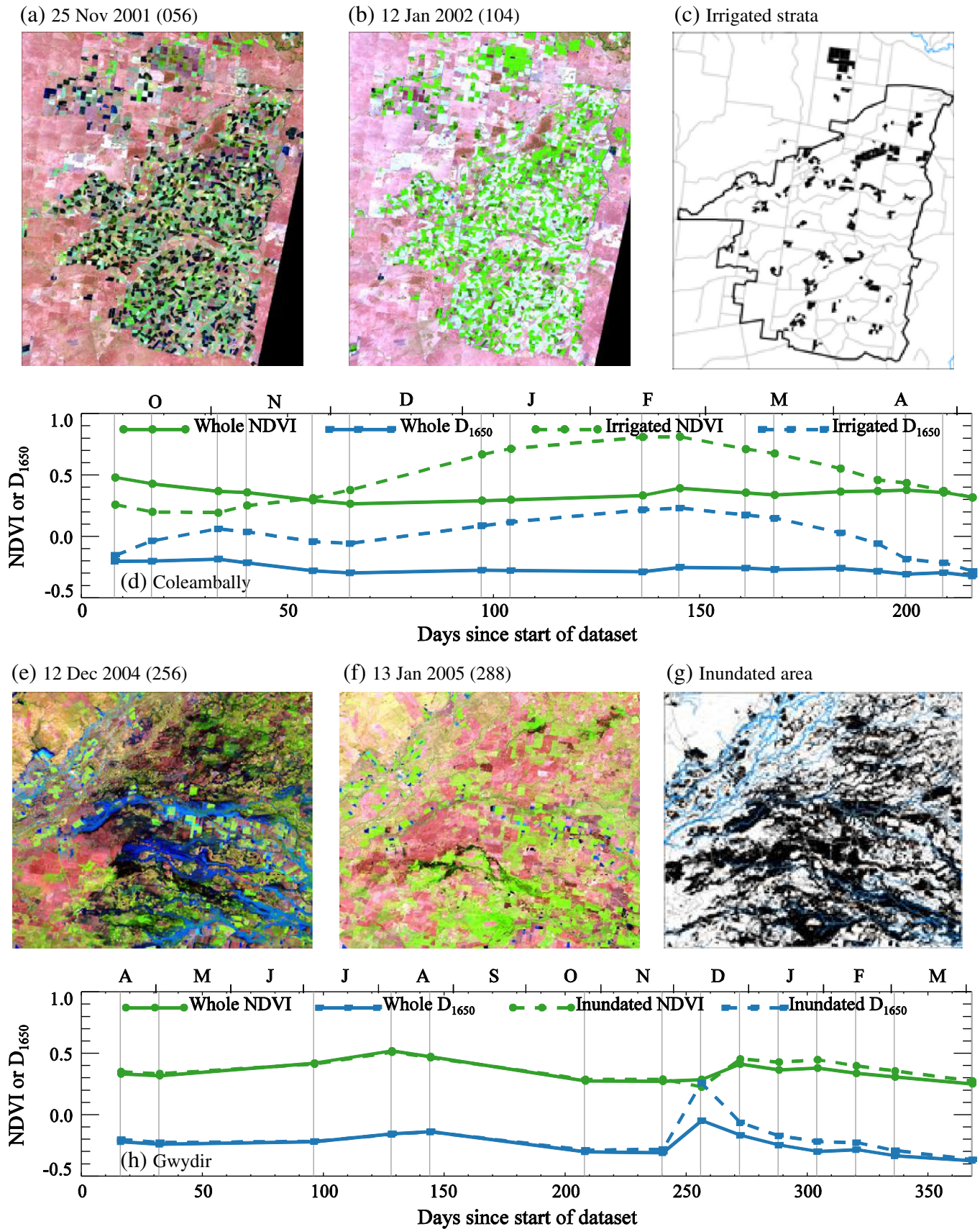| Image #	Coleambally		Gwydir	
	DSSD	Date	DSSD	Date
1	008	08 Oct 2001	016	16 Apr 2004
2	017	17 Oct 2001	032	02 May 2004
3	033	02 Nov 2001	096	05 Jul 2004
4	040	09 Nov 2001	128	06 Aug 2004
5	056	25 Nov 2001	144	22 Aug 2004
6	065	04 Dec 2001	208	25 Oct 2004
7	097	05 Jan 2002	240	26 Nov 2004
8	104	12 Jan 2002	256	12 Dec 2004
9	136	13 Feb 2002	272	28 Dec 2004
10	145	22 Feb 2002	288	13 Jan 2005
11	161	10 Mar 2002	304	29 Jan 2005
12	168	17 Mar 2002	320	14 Feb 2005
13	184	02 Apr 2002	336	02 Mar 2005
14	193	11 Apr 2002	368	03 Apr 2005
15	200	18 Apr 2002		
16	209	27 Apr 2002		
17	216	04 May 2002		

Fig. 1. Site characterisation for the Coleambally (a to d) and Gwydir (e to h) sites. Parts (a) and (b) are Landsat bands 5, 4, 3 displayed as RGB; (c) black shaded areas constitute the irrigated strata and the black line bounds the Coleambally Irrigation Area; (d) shows the Landsat time series mean response of both NDVI and depth at 1650 nm (D_{1650}) for both the whole site and irrigated strata. Parts (e to h) are provided in the same order for the Gwydir site; with (g) showing the inundated area shaded in black. For parts (a, b, e and f) the days since start of dataset (DSSD) are provided in parentheses following the date. For parts (d and h) the vertical grey lines indicate dates when cloud-free data are available and single-letter month abbreviations are provided at the top of those plots.

a winter and a summer crop growing season (Fig. 1e–h). A large flood occurred in mid December 2004, which caused the inundated vegetation and wetness indices to deviate from those of the whole study area (Fig. 1h). The flooding and subsequent inundation covered large areas (~44% see Fig. 1g) that resulted in a different spatial and temporal response compared to that at Coleambally (Fig. 1d). Due to the flooding event, Gwydir is considered a more temporally dynamic site, but as shown later, this is spectrally dependent.

2.2. Blending algorithms tested

Four blending models were evaluated: (*i*) LIM (linear interpolation model; Emelyanova et al., 2012); (*ii*) GEIFM (global empirical image fusion model; Emelyanova et al., 2012); (*iii*) STARFM (Gao et al., 2006); and (*iv*) ESTARFM (Zhu et al., 2010). A generic mathematical description of blending is provided in Emelyanova et al. (2012; their Section 2), with the four algorithms being explicitly defined in Emelyanova et al. (2012; their Appendix A). All four models use cloud-free fine resolution (i.e., Landsat) imagery before and after the date of simulation. GEIFM, STARFM and ESTARFM also use two observed cloud-free coarse resolution (i.e., MODIS) images on the same dates as the Landsat images (i.e., two L–M pairs), and one additional observed MODIS image on the date of simulation. While STARFM and GEIFM (not ESTARFM) are able to produce simulations based on a single L–M pair and one additional MODIS image on the date of simulation, we have not used them in this mode here to ensure that GEIFM, STARFM and ESTARFM all use the same input data. The basic assumptions with respect to temporal and spatial variability and possible limitations of the four algorithms follow.

All four algorithms assume some linear trend of surface reflectance change through time. ESTARFM and LIM assume a single rate of change over the entire time period between the two input Landsat acquisitions (Emelyanova et al., 2012; Zhu et al., 2010). STARFM and GEIFM assume two separate periods of linear change, being: (*i*) from the first Landsat acquisition date to the simulation date; and (*ii*) from the simulation date to the second Landsat acquisition date (Emelyanova et al., 2012; Gao et al., 2006). STARFM and GEIFM, therefore, can model a non-linear change over the entire period between the two input Landsat acquisitions and thus should model non-linear temporal variance better than ESTARFM and LIM. LIM should model non-linear temporal variability the worst as it simply linearly interpolates Landsat data between two dates without recourse to MODIS data.

LIM is calculated on a per-pixel basis, with no contextual neighbourhood assessment. This provides LIM with the best ability to model local linear spatial variability, but the worst ability to model non-linear temporal variability. ESTARFM and STARFM use neighbourhoods around each pixel to define locally influenced linear relationships between Landsat and MODIS, which results in non-linear spatial functions between Landsat and MODIS at the entire image scale. ESTARFM has been designed to model spatial variability better than STARFM by using a more sophisticated identification of pixel similarity within neighbourhoods (Zhu et al., 2010). GEIFM derives reflectance from a single global linear model (i.e., for the whole image extent), and thus should model spatial variability the worst of all four algorithms. In summary, LIM could model spatial variability the best of all four algorithms if there was either very low temporal variance, or if the rate of change was near-constant, but would be expected to model non-linear temporal variability the worst of all four algorithms. STARFM and GEIFM would be expected to model non-linear temporal variability better than ESTARFM, but ESTARFM should model spatial variability better. It is clear that the spatial and temporal variances of a site should directly relate to algorithm performance. However, since spatial and temporal variances are dynamic and likely differ by spectral band, spatial resolution, and surface characteristics, it is expected that no single algorithm will be best under all conditions.

2.3. Applying the blending algorithms

For both study sites the four blending algorithms were applied by simulating a Landsat-like image on some date using the two L–M pair dates that were the nearest temporal neighbours to the simulation date, one before and one after. On the simulation date, a cloud-free Landsat image was also acquired and this was only used for validation (i.e., it was not used as input for simulation). All possible combinations of simulations were processed at both sites for all four algorithms. For example using dates from the Coleambally series (Table 2) for the LIM algorithm, Landsat # 1 (08 October 2001) and Landsat # 3 (02 November 2001) were used to simulate Landsat # 2 (17 October 2001), then Landsat #2 (17 October 2001) and Landsat # 4 (09 November 2001) were used to simulate Landsat # 3 (02 November 2001), and so on for the entire series. The general process was the same for GEIFM, STARFM, and ESTARFM, except that L–M pairs were used on nearest temporal neighbour dates rather than Landsat alone, and the MODIS image on each simulation date was also input to the blending algorithms. This design meant that images were simulated using all four algorithms for the 15 central dates at Coleambally and for the 12 central dates at Gwydir (27 dates in total), and allowed for subsequent comparison between observed Landsat data and simulated Landsat-like output over the time series.

2.4. Defining land cover strata

At each site, subset areas exhibiting high spectral contrast (when compared to the entire study site) were defined to elucidate the influence of land cover dynamics on algorithm performance. The high spectral contrast subsets were primarily due to the addition of water and subsequent vegetation response at both sites. For Coleambally this was a selection of irrigated fields of various summer crops (116,898 pixels at 25 m resolution), including rice (flood irrigated), maize, sorghum and soybean (all furrow irrigated) fields identified in Van Niel and McVicar (2004a); see Fig. 1c. For Gwydir, the inundated area of a large flood event (3,856,288 pixels at 25 m resolution) was defined from the 12 December 2004 and 28 December 2004 Landsat imagery where the environmental moisture index (specifically the 'Depth at 1650 nm'; see Van Niel et al., 2003) exceeded 0.0 on either date (Fig. 1g).

2.5. Partitioning spatial and temporal variances

The overall variance of the Landsat and MODIS spatio-temporal datasets at each study site was partitioned into spatial and temporal variances using the approach of Sun et al. (2010; their Eq.(10)). Partitioning of variance was performed at each site at both Landsat and MODIS resolutions in two ways: (*i*) over the entire Landsat (MODIS) spatio-temporal data cube; and (*ii*) using all possible three sequential-date Landsat (MODIS) spatio-temporal data cubes that correspond to the application of the blending algorithms described above (i.e., the Landsat (MODIS) image of the simulation date and its two nearest temporal neighbours, one before and one after). The same partitioning of variance can be applied to any spatio-temporal dataset to assess blending algorithm performance, thus allowing future research to put their results into the context of our study.

2.6. Evaluating the blending algorithms

Accuracy estimation was performed by calculating the spatially-averaged (i.e., time series) root mean square difference (RMSD), bias (i.e., simulated minus observed), and r^2. Temporally-averaged RMSDs were calculated across the entire period for each pixel and band at both sites. The percentage area of the lowest RMSD for each algorithm was calculated, herein named 'percentage best algorithm'. All assessments were performed over both the whole site areas, and the

subset areas. Due to TM/ETM+ data dimensionality (Crist & Cicone, 1984), for select analyses we used Landsat bands 3, 4 and 5 only to assess the results. That is, we selected one band from the three highly correlated visual bands (i.e., B3 from B1 to B3), the near infra-red band (i.e., B4), and one band from the two highly correlated SWIR bands (i.e., B5 from B5 and B7). The associated MODIS bands are herein labelled M1, M2, and M6.

3. Results

Fig. 2 shows the spatial and temporal variance partitioning for the three key bands for both the Landsat (Fig. 2a–d) and MODIS (Fig. 2e–h) sequential-date spatio-temporal data cubes. The primary changes in spatial and temporal variances at both sites at Landsat resolution occurred in B4 and B5 due to the spectral response associated with water and subsequent plant growth. For the whole Coleambally site, the spatial variance of B5 was the dominant variance throughout the entire time period (Fig. 2a). This is because there was a high spectral contrast between the irrigated fields and the fallow fields, dryland pastures and woodlands (Fig. 1a–b). This spatial variability between irrigated and non-irrigated areas dominated the temporal variability, and since the main difference between the irrigated and non-irrigated regions was the presence or absence of water, B5 was most affected. The secondary source of variability at Landsat spatial resolution was the change in the vegetation response between the irrigated and non-irrigated portions of the site, and therefore, the spatial variance of B4 was the second highest variance overall (Fig. 2a), since B4 is most sensitive to greening vegetation. The spatial variance of B3 was the third highest overall influence (Fig. 2a), demonstrating that the complementary temporal variance for all three bands was of minor influence when the whole Coleambally site was considered using Landsat.

The influence of land cover was particularly evident when the whole study site Landsat variances (Fig. 2a) were compared to those of the irrigated subset (Fig. 2b). Rice fields were flood irrigated in October–November, which resulted in a high spatial variance in B5 during those 2 months (Fig. 2b). The large change in spectral reflectance between bare soil fields and flooded fields occurred to only some of the fields within the irrigated strata (since not all the fields were rice), which resulted in high spatial variability rather than high temporal variability. During the transition when open water or bare soil became replaced by green vegetation, the spatial variance of B5 decreased, and the temporal variance of B4 increased (see November–December, Fig. 2b). The reason why temporal variance was higher than spatial variance in B4 during this transition was because all fields comprising the irrigation strata were becoming greener (becoming less variant) during this time. This resulted in lower spatial variability and higher temporal variability. At Landsat resolution the increased importance of temporal variability is a key difference when comparing the irrigated strata to the whole site. When green crops were established as the dominant spectral signal (late December through mid-February, Fig. 2b), the spatial and temporal variances of B4 were approximately equal. Finally, from mid-February through April, the spatial variance of B4 was the highest variance, governed by the varying rates of crop senescence between the four crops (Fig. 2b).

The Landsat variances between the whole Gwydir (Fig. 2c) and the inundated subset (Fig. 2d) were not as different as was seen for the corresponding summaries at Coleambally. This was because almost half of Gwydir was inundated (Fig. 1g), which resulted in the whole site's variances being a slightly muted version of the inundated area variances. The partitioning of variance in the whole Gwydir and its inundated subset (Fig. 2c–d) showed high spatial variance for B4 from July–September (DSSD 96–144) due to the spatial variability of the winter crops. The flood caused many areas to become wetter (and thus less spatially variant) in a short period, thus the temporal

variance of B5 was very high during the December flood (DSSD 240–272, Fig. 2c–d). Due to increased water availability, the spatial variance of B4 became high, lagging the flood as the water signal receded being replaced by the spectral dominance of growing vegetation. At Coleambally, the highly managed flooding and planting occurred over homogeneous, fertilised, and flattened soil, whereas at Gwydir the natural flood covered heterogeneous pastures and riparian vegetation with variable elevations, nutrient levels, and prior vegetation amounts and types. Subsequently, the transition from water dominance to greening vegetation resulted in an increase in spatial variance of B4 at Gwydir, as plant growth responded variably across space due to a wide range in conditions (January–February, Fig. 2c–d). At Coleambally the whole site was dominated by spatial variance (Fig. 2a). However, this was not the case at Gwydir, where a single domain's variance was not as dominant through time, but rather alternated between high spatial and high temporal variances. This difference is important for interpretation of blending algorithm performance, and for subsequent development of the framework for algorithm selection.

When the variances where assessed using MODIS data, they were generally much lower at Coleambally (Fig. 2e–f), than the corresponding Landsat variances (Fig. 2a–b). This demonstrated that a good deal of the dynamics occurred at a finer spatial resolution than that of the MODIS data (because homogeneous patch sizes at Coleambally were small). There was no such decrease in the magnitude of MODIS variances at Gwydir (Fig. 2g–h) when compared to the corresponding Landsat variances (Fig. 2c–d), which indicated that much of the spatial and temporal dynamics were observable at the MODIS spatial resolution due to the larger size of homogeneous patches at Gwydir. At Coleambally, the relative dominance and general patterns were the same between Landsat and MODIS variances; just much reduced for the MODIS variances (Fig. 2a–b and e–f). At Gwydir, while the magnitude remained about the same, when comparing MODIS to Landsat, the relative dominance of spatial variance increased for all three key bands (Fig. 2c–d and g–h), most likely due to the higher proportion of 'mixels' at MODIS resolution. The important implication is that at Coleambally, because much of the spatial and temporal variances are not observable at MODIS resolution, it is unlikely that blending with MODIS data could make much improvement. Alternately, the patch size and land cover dynamics of Gwydir allowed for a meaningful summarisation of dynamics at both Landsat and MODIS resolutions, and therefore, there is a good possibility of improving simulations by inclusion of MODIS data. Of particular importance, are the times and spectral bands where Landsat and MODIS displayed similar variances (e.g., B4 and M2 spatial variance between DSSD 96–144 and B5 and M6 temporal variance between DSSD 240 and 272, Fig. 2d, h). These are the situations under which blending Landsat and MODIS should provide maximum benefit and when spatial and temporal variances should explain algorithm performance the best, as shown below.

Fig. 3a–b represents observed Landsat data and simulated Landsat-like images using the four blending algorithms for a key date at Coleambally and Gwydir, respectively. Insight can be gained by visual inspection of these selected dates. On DSSD 104, the spatial variances of B5 and M6 dominated all others at both Landsat and MODIS spatial resolutions at Coleambally (Fig. 2a, e). At Landsat spatial resolution, the spatial variances of B3 and B4 were also higher than their corresponding temporal variances (Fig. 2a). The simulated images were all able to capture the large spatial variance between the central irrigation area and the surrounding dryland agriculture on this day (Fig. 3a). When summarised from Landsat data, the whole of Gwydir on DSSD 256 had a higher temporal than spatial variance in B5, a slightly higher spatial than temporal variance in B4 and nearly equal proportions of spatial and temporal variances in B3 (Fig. 2c). From MODIS data, DSSD 256 had approximately equal (and high) values of spatial variance in M1, spatial variance
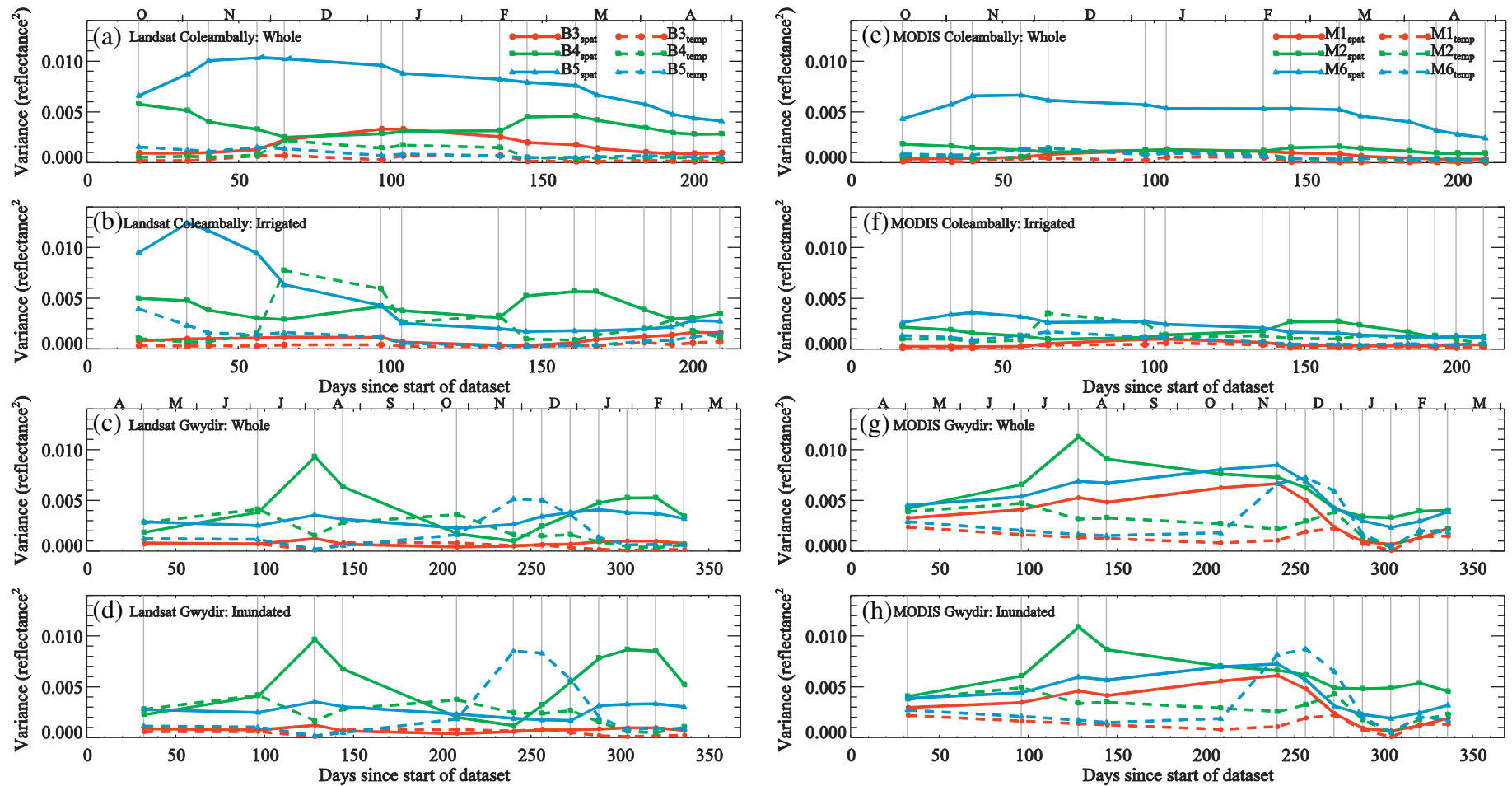
Fig. 2. Time series plots of the spatial (spat) and temporal (temp) variances at both Landsat and MODIS resolutions for both the Coleambally and Gwydir sites. Parts (a) and (b) are Landsat for the whole Coleambally site and irrigated strata, respectively, with parts (c) and (d) being Landsat for the whole Gwydir sites and inundated area, respectively. Parts (e) to (h) are MODIS data for the same areas as provided in (a) to (d). The spatial and temporal variances are calculated using all possible three sequential-date images with the results plotted at the date of the central image. Vertical grey lines indicate dates when cloud-free data are available and single-letter month abbreviations are provided at the top of the plots.
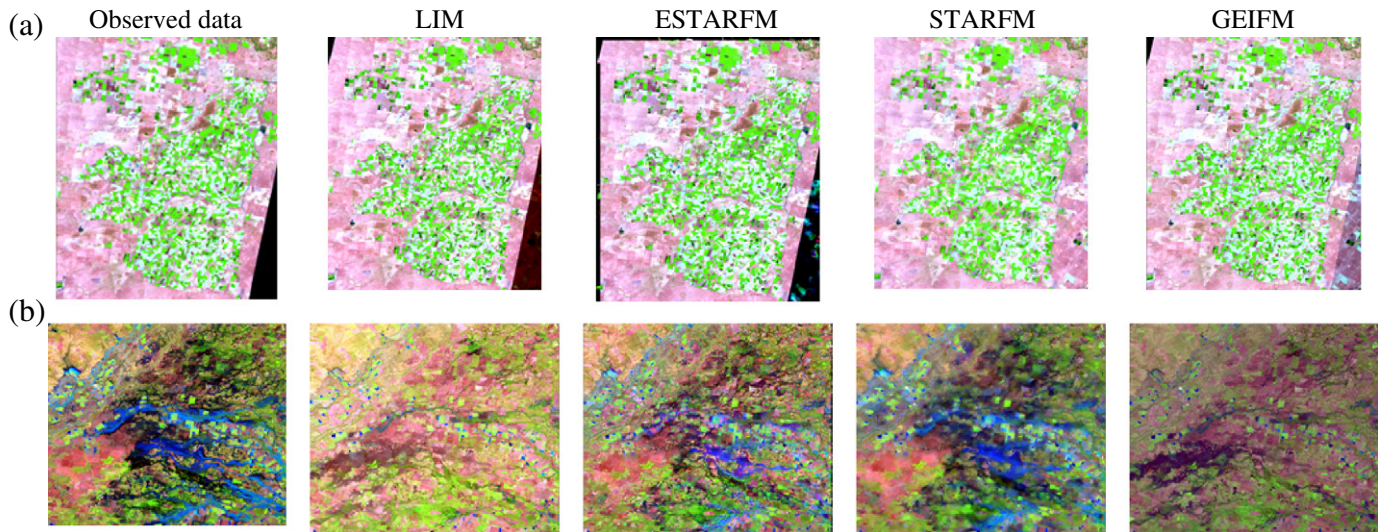
Fig. 3. Observed Landsat data and simulated Landsat-like images with the blending algorithm indicated above. Part (a), the top row, is Coleambally for 12 January 2002 (DSSD = 104) and part (b), the bottom row, is Gwydir for 12 December 2004 (DSSD = 256). All images are presented as Landsat bands 5, 4, 3 displayed as RGB with the same stretch applied to all images for each study site.

in M2, and spatial and temporal variances in M6 (Fig. 2g). This more complex mixture of domain variances at Gwydir caused the simulated images to have more variation across algorithms (Fig. 3b). As expected (Section 2.2), LIM and GEIFM simulations were not very similar to observed Landsat due to their overly simplified assumption of either time or space (Fig. 3b). ESTARFM and STARFM were more similar to observation (Fig. 3b). As mentioned earlier, ESTARFM estimates spatial variance better, but assumes a linear change over the entire period between the two input Landsat acquisitions. When run in two L–M pair mode (as we have done), STARFM assumes two linear changes in surface reflectance over this same period, thus should model non-linear temporal variance better. This is generally confirmed by Fig. 3b where STARFM simulations were visually most similar to the observed Landsat data. For DSSD 256, temporal variance of B5 was the dominant variance from the Landsat data (Fig. 2c) and was approximately equivalent to its spatial variance for MODIS (see M6, Fig. 2g). ESTARFM simulations matched slightly less well although the spatial representation of the output seems more visually appealing (Fig. 3b).

Many previous studies demonstrated accuracy assessment from scatter plots of observed versus simulated Landsat data (Section 4). Our study resulted in 1296 possible scatter plots: 720 for Coleambally (i.e., six bands by 15 dates by four algorithms by two summarisation areas) and 576 for Gwydir (i.e., six bands by 12 dates by four algorithms by two summarisation areas). A representation of such a large dataset required summarisation of scatter plot information. The 360 scatter plots for the whole Coleambally are summarised in Fig. 4. Each of the 12 scatter plot summarisations, herein called 'blocks', in Fig. 4a–l contains 90 cells (six bands by 15 dates). Each cell represents an error or accuracy metric (i.e., spatially-averaged bias, r^2, or RMSD) generated from a single scatter plot. For example, the bias, r^2, and RMSD value associated with the bold outlined cells highlighted in each of the three ESTARFM blocks (Fig. 4g–i) were calculated from the scatter plot shown in Fig. 4m. Subsequently, each row in any of the 12 blocks (Fig. 4a–l) is a time-series of the specified error or accuracy metric for a given algorithm. For example, the outlined rows associated with the four algorithm's r^2 image (Fig. 4b, e, h, and k) are shown in Fig. 4n. Finally, each column in any of the 12 blocks represents an inter-band comparison of the specified error or accuracy metric for a given date for a given algorithm. For example, Fig. 4o shows the spatially-averaged RMSDs associated

with the outlined columns of the four algorithm's RMSD blocks (Fig. 4c, f, i, l) for DSSD 104.

Summary of all 360 scatter plots in this manner allowed for a synoptic assessment of the algorithm performance over time, across bands, and between algorithms. When considering whole Coleambally spatially-averaged bias, LIM (Fig. 4j) was shown to have a larger error, evident by the presence of more cells having either large negative bias (i.e., red) or large positive bias (i.e., blue), than the other algorithms (Fig. 4a, d, g). The r^2 summaries revealed that all four algorithms had 'hot spots' of relatively low r^2 in the early and late parts of the time series for the three visible bands, evident by the lighter coloured areas centred around DSSD 040 and 184 for B1–B3 (Fig. 4b, e, h, k). GEIFM and ESTARFM had a much lower RMSD when averaged over the whole of Coleambally, and according to that error metric, ESTARFM would be considered most preferred, evident by the lowest errors (i.e., lightest colours) over the whole block, particularly for B4, B5, and B7 (Fig. 4i compared to Fig. 4c, f, and l).

Within the irrigated strata at Coleambally, the biases of GEIFM and STARFM increased substantially when compared to those of the whole study site (compare Fig. 5 to Fig. 4). The bias of ESTARFM was slightly larger for the irrigated strata when compared to the whole site (Fig. 5g compared to Fig. 4g). More red tones in the ESTARFM irrigated bias summary revealed that ESTARFM tended to underestimate reflectance in the irrigated area (Fig. 5g), whereas more blue tones revealed, that over the whole study site reflectance was more often overestimated by ESTARFM (Fig. 4g). The r^2 blocks showed lower correlations and the RMSDs were higher for all four algorithms within the irrigated strata (Fig. 5 compared to Fig. 4). The most obvious change was seen for the RMSD summaries for ESTARFM, where over the whole study site RMSDs were much lower (Fig. 4i) than for the irrigated strata (Fig. 5i). This verifies that ESTARFM is more sensitive to spatial variability, as spatial variance was dominant over the whole site across all times (Fig. 2a) whereas spatial variance was not as dominant within the irrigated strata (Fig. 2b).

The accuracy assessment at the whole Gwydir (Fig. 6) and that of the inundated subset (Fig. 7) revealed similar patterns. A group of low r^2 values was evident through the central part of the time series between DSSD 208–272 (Figs. 6b, e, h, k and 7b, e, h, k) when both spatial and temporal variances were comparatively high (Fig. 2c, d, g, h). It is also important to note, that during the flood (DSSD 256)
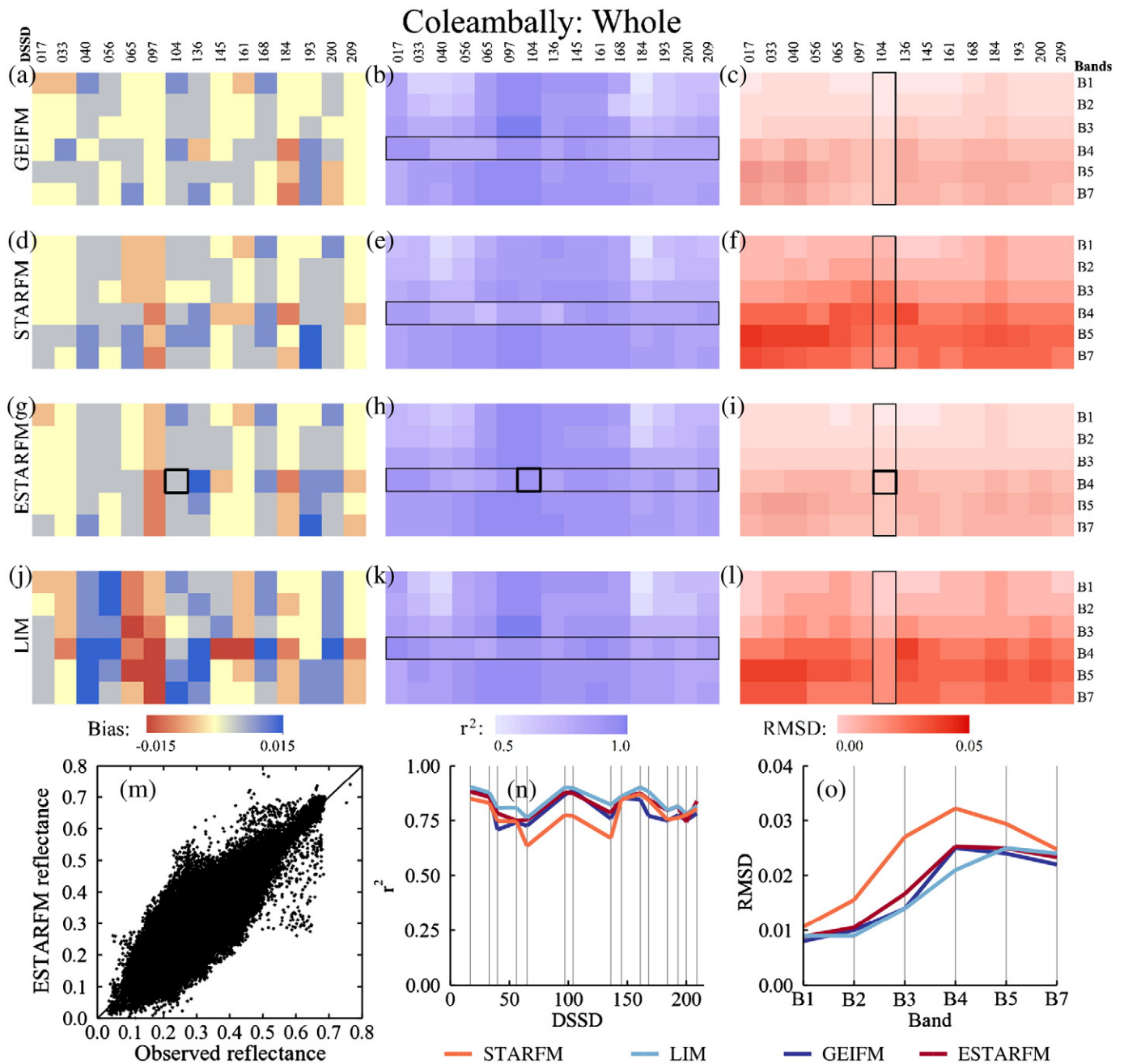
Fig. 4. Summary error and accuracy metrics for the whole Coleambally site. Parts (a) to (c) are bias, r^2 and RMSD for the GEIFM algorithm, respectively. Parts (d) to (l) are the same three statistics for the STARFM (d to f), ESTARFM (g to i) and LIM (j to l) algorithms, respectively. These 12 blocks are ordered by DSSD from left to right and by Landsat bands from top to bottom. Part (m) illustrates a single scatter plot, with the relevant statistics highlighted by the bold square. Part (n) is a time series of r^2 statistics for the four algorithms illustrated in the horizontal rectangles (shown directly above in parts b, e, h and k) with the vertical grey lines indicating dates when cloud-free data are available. Part (o) shows the band dependence of RSMD for the four algorithms illustrated in the vertical rectangles (shown directly above in parts c, f, i and l).

STARFM had the highest r^2 (Figs. 6n and 7n) and lowest RMSDs (Figs. 6o and 7o).

The expectation when comparing the advanced algorithms (STARFM and ESTARFM) to the simple algorithms (LIM and GEIFM) is that the advanced algorithms should have lower errors due to their increased complexity. From an operational standpoint, then, there is interest in defining when the extra complexity and associated computational cost of the advanced algorithms are not required. As examples, at Coleambally, STARFM and ESTARFM had a much larger bias than GEIFM for DSSD 097 across all bands (Fig. 4a, d, g). Also at Coleambally, LIM had a higher r^2 than the other algorithms for most of the time series

for B4 (Figs. 4n and 5n) and an RMSD that was lower or approximately equal to the other algorithms for DSSD 104 for all bands (Figs. 4o and 5o). These two examples showed that at a site where spatial variance is by far the dominant factor and where there is a high temporal density of imagery, even an algorithm as simple as LIM can outperform the advanced algorithms, given the right conditions. Since most of the variances at Coleambally occurred at a finer spatial resolution than that of MODIS (Fig. 2), there should be less difference between LIM and the two industry standard algorithms at Coleambally than at Gwydir.

The percentage best algorithm area estimates for the entire sites and for the subset areas (i.e., Coleambally irrigated strata and Gwydir
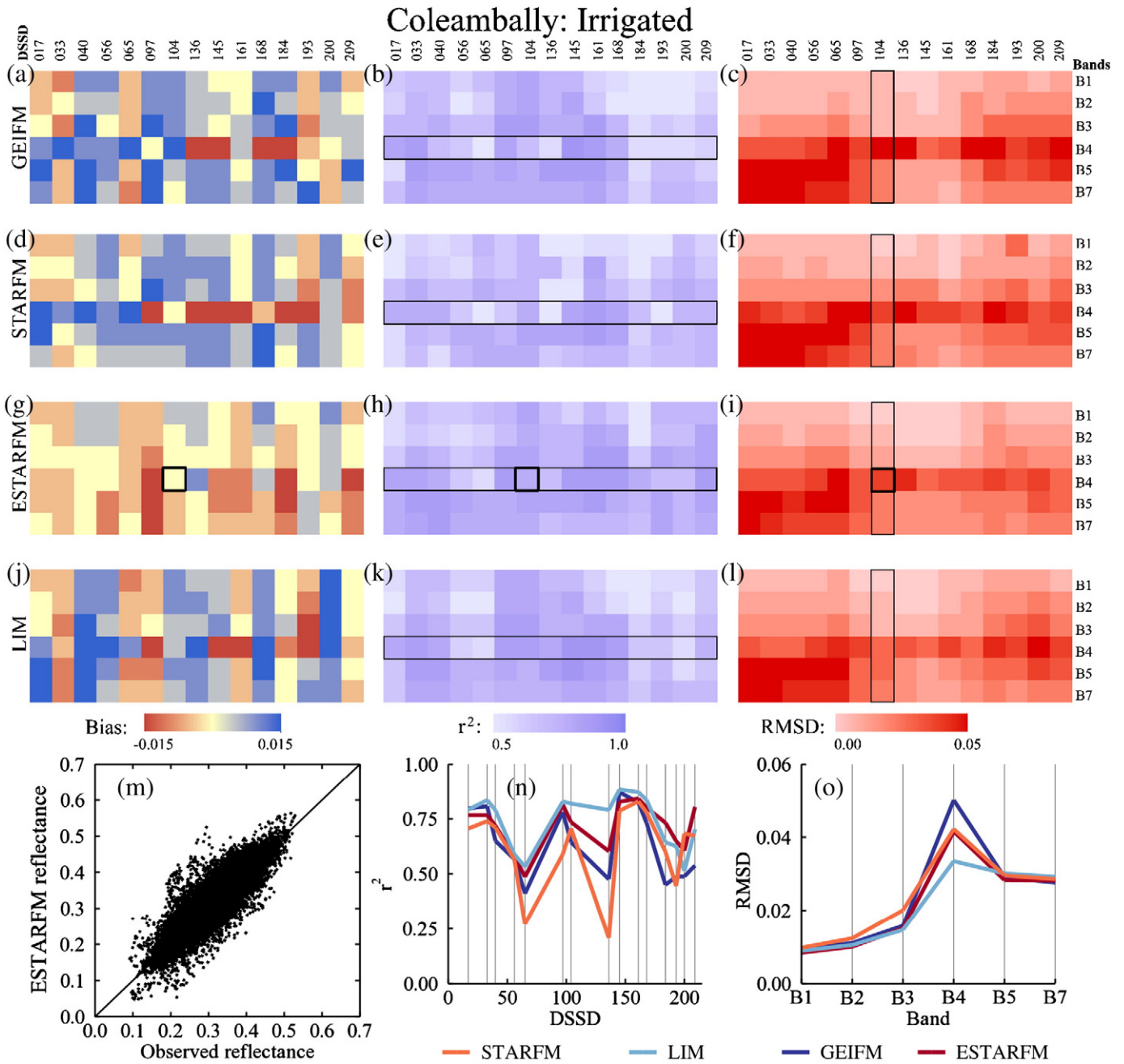
Fig. 5. Summary error and accuracy metrics for the Coleambally irrigation strata; see Fig. 4 for a detailed description of sub-plots.

inundated area) are shown in Fig. 8a–d. At Coleambally this land cover stratification clearly highlighted the higher performance of the advanced algorithms for the selected irrigated fields (compare Fig. 8a and b). The advanced algorithms produced the lowest RMSD for about 80% of the pixels in the irrigated strata (Fig. 8b), while over the whole Coleambally this advantage was not seen (see Fig. 8a where the combined advanced algorithms and the combined simple algorithms are both ~50% of the area). At Gwydir, there was little difference between the percentage best algorithms whether considering the entire site (Fig. 8c) or only the inundated area (Fig. 8d). As the dominant feature over the entire site and time series was the flood, as discussed above, this result was expected. STARFM outperformed ESTARFM for most bands with respect to percent area having the lower RMSD (Fig. 8c–d). At Gwydir (both the whole site

and the inundated area) the advanced algorithms clearly outperformed the simple algorithms. The percentage area, where the advanced algorithms were better, ranged between 71% (i.e., B4 in Fig. 8c) and 84% (i.e., B5 in Fig. 8d). Land cover explained the spatial pattern of which blending algorithm was best, with little variation in spatial pattern observed between bands for both sites (Emelyanova et al., 2012; their Fig. 8). Several other spectrally dependent error analyses for both specific dates and summarising the entire datasets (both for whole sites and the land cover strata) are provided in Emelyanova et al. (2012; their Tables 5 and 6, respectively).

The overall Landsat and MODIS variances over the entire dataset at each site was partitioned into spatial and temporal components for each spectral band (i.e., at Coleambally the overall variance was calculated using the entire image for the 17 dates and at Gwydir
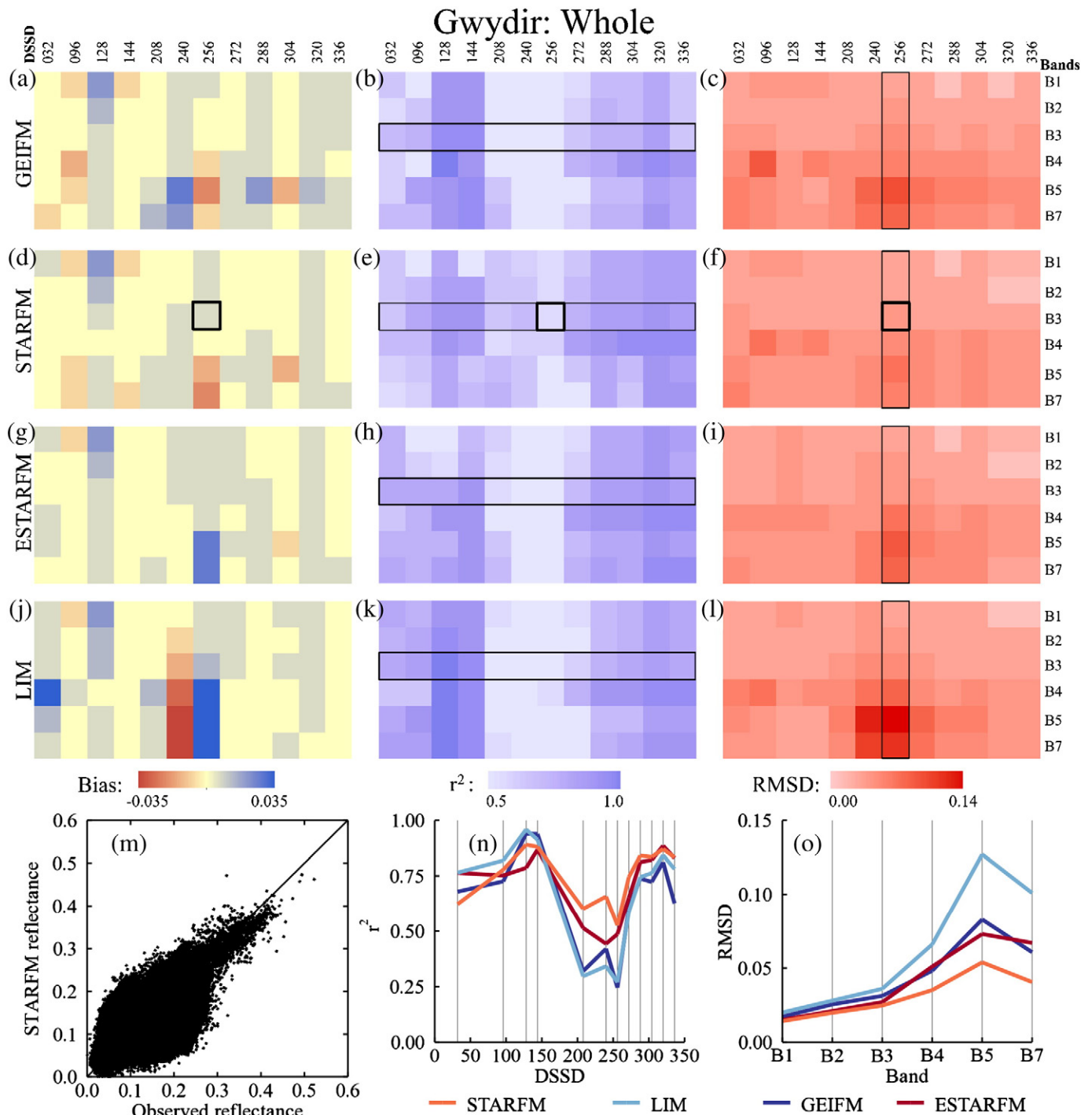
Gwydir: Whole

Fig. 6. Summary error and accuracy metrics for the whole Gwydir site; see Fig. 4 for a detailed description of sub-plots.

using the entire image for the 14 dates), see Fig. 9a–d. In Fig. 9, symbols are coloured by which algorithm between ESTARFM and STARFM (or LIM and GEIFM) had the lowest RMSD, and symbols differentiate study sites. Based on the lowest RMSD there was a strong association by study site and algorithm (i.e., ESTARFM with Coleambally and STARFM with Gwydir (Fig. 9a–b) or i.e., LIM with Coleambally and GEIFM with Gwydir (Fig. 9c–d)). From Fig. 2a–b and e–f, it was determined that the spatial and temporal variances at Coleambally primarily occurred at a finer spatial resolution than that of MODIS, essentially limiting the benefit from blending Landsat with MODIS there. Because of this, it is expected that spatial and temporal variances as a metric of algorithm performance should

have rather poor predictive power in this case. This is confirmed by Fig. 9a–b, where ESTARFM had the lowest overall RMSD for all spectral bands at Coleambally regardless of their spatial or temporal variance.

At Gwydir the land cover dynamics occurred at a spatial resolution observable with MODIS (Fig. 2). This means that the spatial and temporal variances should be more useful for predicting algorithm performance. This is corroborated using the Gwydir nearest temporal neighbour spatial and temporal variances as calculated for Fig. 2d and h for two selected time periods. At Gwydir there were two periods where the underlying spectral feature influenced both Landsat and MODIS variances proportionately. These are shown in Fig. 9e–f
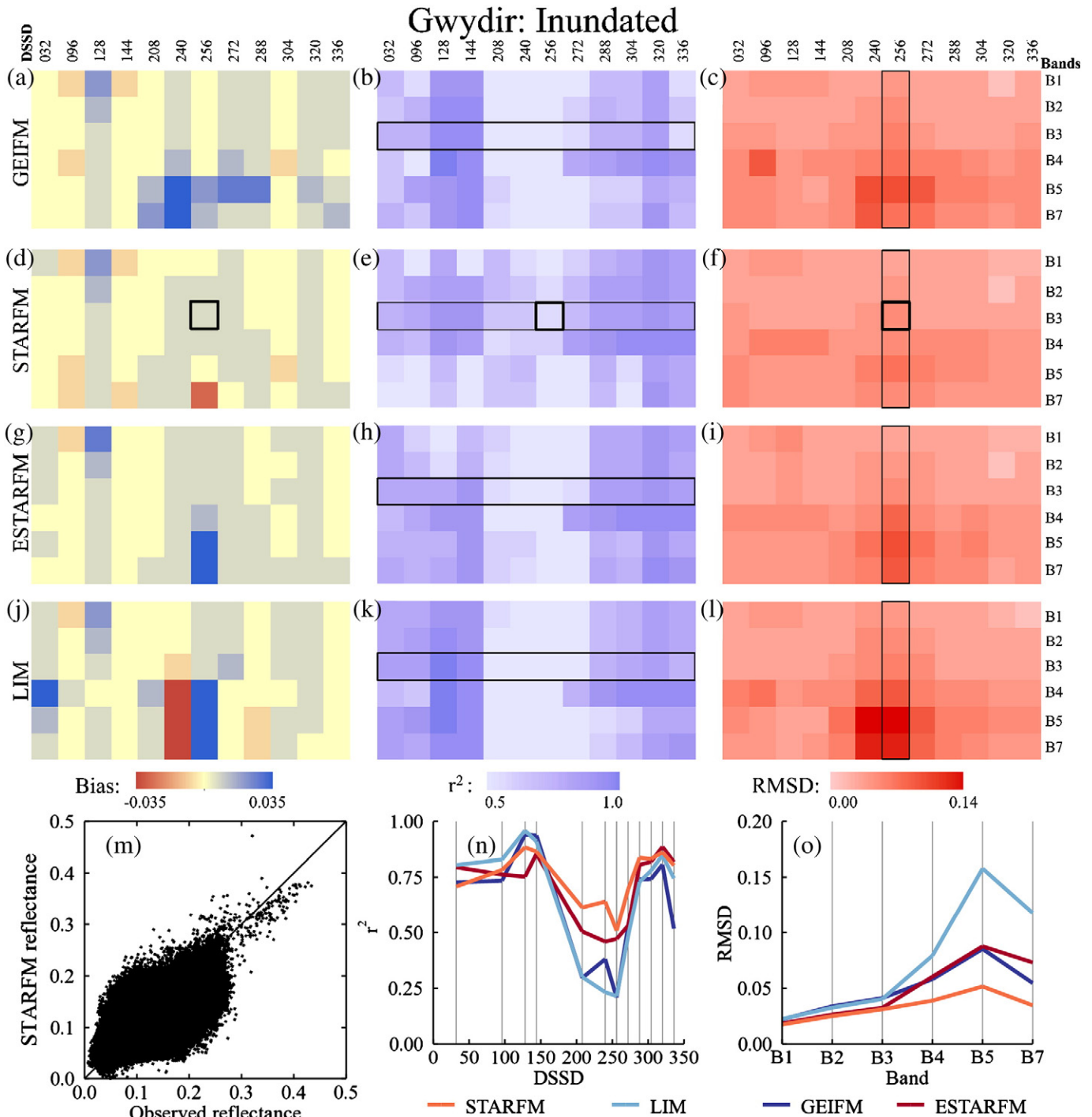
Fig. 7. Summary error and accuracy metrics for the Gwydir inundation area; see Fig. 4 for a detailed description of sub-plots.

for Landsat and MODIS variances, respectively. During these times and for these spectral bands (i.e., Landsat B4 and B5 and the MODIS equivalents), spatial and temporal variances have discriminating ability to predict which algorithm would simulate observed data better between both ESTARFM and STARFM (Fig. 9e–f). The implication is that when the land cover signal is strong enough and large enough that Landsat and MODIS both observe the same variance dynamics, then the spatial and temporal variances appear to have good predictive power (Fig. 9e–f). This assessment was performed to summarise our two sites in a manner useful for placing our results and those of future studies into context, and forms the basis of our framework for blending algorithm selection (discussed below).

Computing cost was evaluated by recording the time required to simulate a six band image at the Gwydir site. All four blending algorithms were run on a PC with an Intel(R) Core(TM) i7-2760M processor (2.4 GHz, 6 MB Cache) with Turbo Boost Technology and 4 GB 1333 MHz DDR3 SD RAM. No other applications were run during the testing and all data were stored and written to the local disk to avoid network issues. STARFM required ~6 h (21,600 s) to simulate a Landsat-like six band image; ESTARFM needed around twice as long (i.e., ~12 h or 45,000 s). Runtimes are dependent on the size of the moving window for both algorithms and these tests use the 'research code' (i.e., without any code parallelisation/optimisation). In comparison, the two simple algorithms, LIM and GEIFM, completed
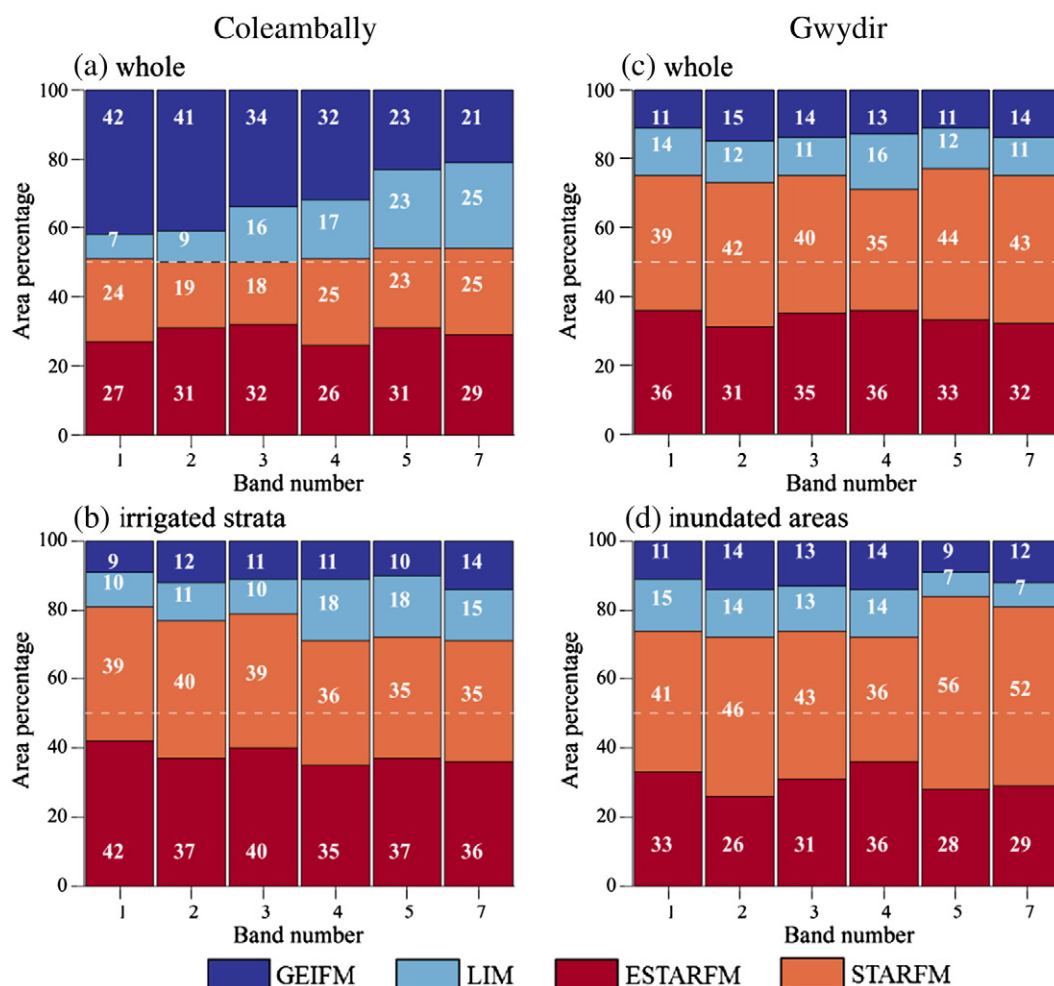
Fig. 8. Percent area for the six Landsat reflective bands with the smallest RMSD (i.e., greatest accuracy) for the four blending algorithms. Part (a) is Coleambally whole; (b) is Coleambally irrigated strata; (c) is Gwydir whole; and (d) is the Gwydir inundated areas. The numbers in white are the area percentages for each algorithm, and to aid in interpretation when comparing the two advanced algorithms (shown in red colours) with the two benchmarks (shown in blue colours) a white dashed line at 50% is overlaid.

the processing in less than 1 min (i.e., 25 s and 55 s, respectively). The trade-off between computational costs and accuracy for operational use, as a function of spatial heterogeneity and temporal dynamics, is discussed below.

4. Discussion

To place our results in context we reviewed studies which used STARFM and/or ESTARFM. Table 3 shows that since 2006, when STARFM was first published, the number of case studies using it has increased every year. STARFM is now widely used by the remote sensing research community, and the number of applications using ESTARFM has also grown since its publication; both algorithms have been applied across the globe. Despite STARFM being developed to simulate Landsat-like reflectances from L–M pairs, it has also been successfully used to simulate thermal imagery (study 10) and remotely sensed based grids of actual evapotranspiration (studies 12 and 19). While spatial extents of the study areas varied from hundreds to tens of thousands of square kilometres, temporal extents were mostly comparatively short (i.e., from 2 to 8 months), although for some studies much larger temporal extents were available: from 1 year up to 10 years (i.e., studies 4, 11, 13, 17, 19–22 and 24). One L–M pair per month was rarely available for the short-term studies and the number of L–M pairs per year did not exceed 8 for the long-term studies. In contrast with the other studies–except for

study 22 which often used two Landsat images per month–those conducted here (summarised as studies 23 and 24 in Table 3, respectively) used 17 L–M pairs over 8 months (Coleambally) and 14 L–M pairs over 12 months (Gwydir; see Table 2 for details). Moreover, our Gwydir site (study 24) is the first time where blending algorithms have been used in a highly temporally dynamic flooding and subsequent inundation environment.

Half of the studies applied STARFM and/or ESTARFM in landscapes where forestry was the dominant land cover with various temporal dynamics (studies 1, 4–6, 8–9, 11, 15, 18 and 21 in Table 3). Other studies have been conducted where the primary land cover type included woodland, agricultural, urban areas, patches of bare soil and open water. Some sites (studies 17 and 18) contained dryland forest with a moderately temporally dynamic understory. Even though previous studies encompass a range of land cover types, most performed a whole image assessment. Just a few (i.e., studies 5, 20, and 21, Table 3) presented specific land cover type accuracy assessment and none reported dependence between accuracy and patch size. We found a strong relationship between algorithm performance and land cover type at both of our sites (Fig. 8). Here, the advanced algorithms were considered the best models in a much higher percentage of area within the irrigated strata at Coleambally when compared to the entire study site (Fig. 8). Thus, land cover stratification provided a different interpretation of results than would have been made if accuracy was assessed without such stratification.

Before applying a computationally expensive blending algorithm in a heterogeneous and/or temporal dynamic region, the use of computationally cheaper and simpler algorithms should be considered. This is the first study that compared advanced and benchmarking algorithms. Such a comparison provided a critical context for accuracy assessment, and without it, our interpretation of accuracy metrics would have been less meaningful and potentially misguided. We

showed that very often, the simple algorithms had similar or lower errors than one or both of the advanced algorithms (e.g., Fig. 4o), with this effect varying across spectral, spatial, and temporal domains.

Zhu et al. (2010) tested STARFM and ESTARFM at two sites (Table 3), which is the only other study besides ours to test both algorithms. Their study was mainly represented by large forest areas with geographically specific growing seasons and temporal dynamics. The
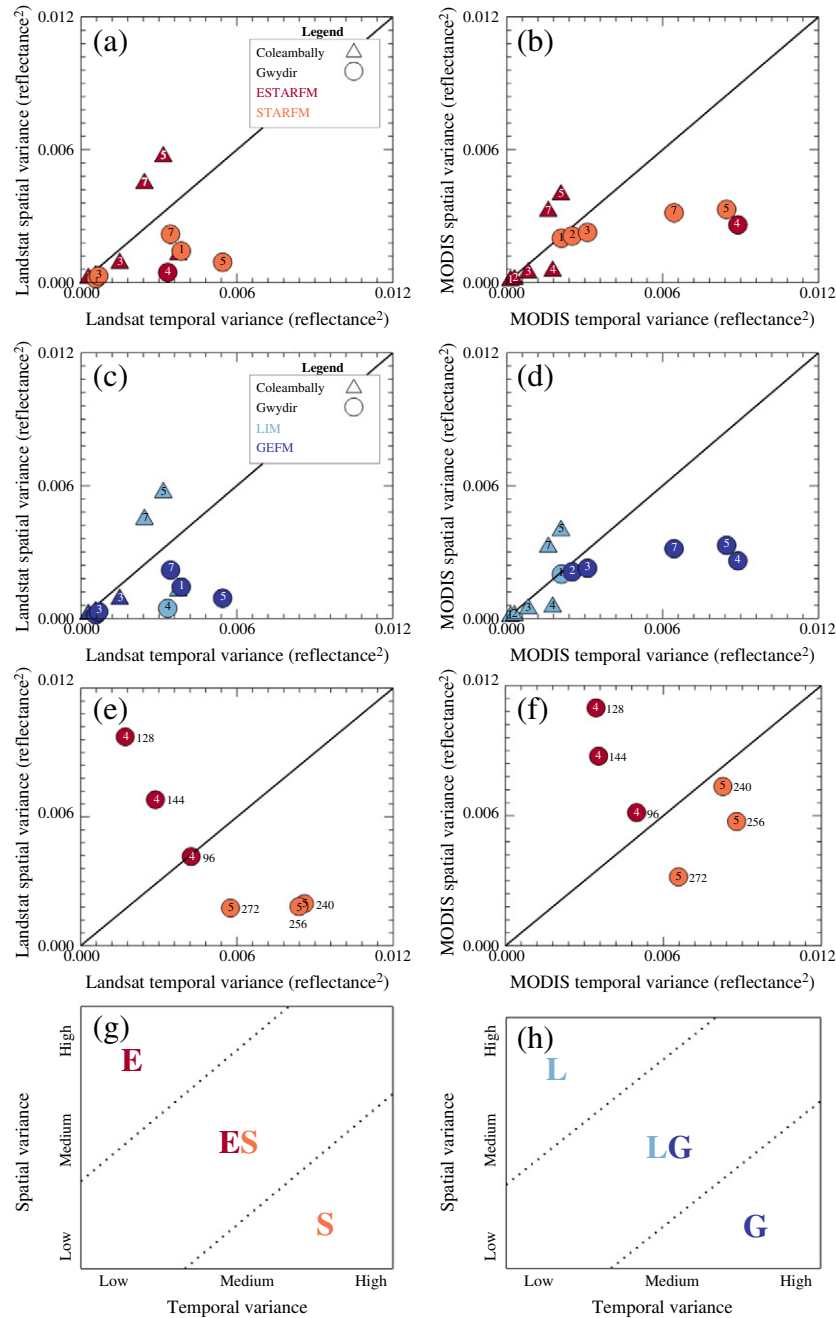

Fig. 9. Analysis of temporal variance and spatial variances. Part (a) is both whole sites calculated using the whole Landsat data cube coloured by lower RMSD between STARFM and ESTARFM; (b) as for (a), but with MODIS data; (c) as for (a), but coloured by lower RMSD between LIM and GEIFM; and (d) as for (b), but coloured by lower RMSD between LIM and GEIFM. Part (e) is calculated from Landsat data for selected bands of interest and times as defined in the main text (i.e., DSSD 96–144 for B4 and DSSD 240–272 for B5); (f) as for (e), but with MODIS data. Part (g) is a conceptual model for STARFM (S) and ESTARFM (E); and (h) is a conceptual model for LIM (L) and GEIFM (G). In parts (a) to (f) all internal symbols are labelled by Landsat band numbers for ease of comparison. In parts (e) and (f) numbers proximally located external to each symbol denote the DSSD. The legend in (a) applies to (b), (e) and (f), and the legend in (c) applies to (d).

Table 3

Summary of applications using the STARFM and/or ESTARFM algorithms. Under the 'Remote sensing inputs' column the following abbreviations are used: (*i*) LMP = Landsat–MODIS image pair; (*ii*) AMP = ASTER–MODIS image pair; (*iii*) 8-MODIS = an eight-day MODIS composite; and (*iv*) 16-MODIS = a sixteen-day MODIS composite. Under the 'Accuracy metric(s) of simulated reflectances' column abbreviations are defined as follows: (*i*) VA = Visual Analysis between observed and simulated Landsat images; (*ii*) SPO = Scatter Plot between Original observed and simulated Landsat band reflectances, (*iii*) SPN = Scatter Plot between observed and simulated reflectances Normalised to the maximum band reflectance; (*iv*) SPD = Scatter Plot of Differences between the simulated and observed images versus differences between the observed images; (*v*) SPFPC = Scatter Plot between ground measured FPCs (Foliage Projective Cover; Bhandari et al., 2012) and derived from the simulated images; (*vi*) MD and MAD = Mean Difference and Mean Absolute Difference between observed and simulated images, respectively; (*vii*) r² denotes the r-squared regression coefficient estimates; (*viii*) T presents a T-test of whether the mean difference between observed and simulated data varied significantly from zero (i.e., P-value ≤ 0.01); (*ix*) AEI = Absolute Error Image; and (*x*) TRMSD and SRMSD stand for Temporal and Spatial RMSD assessment, respectively, as described in Emelyanova et al. (2012). Studies are ordered chronologically then alphabetically.

N	Reference	Algorithm(s)	Region	Landscape specifics	Remote sensing inputs	Study period	Study area (km²)	Accuracy metric(s) of simulated reflectances
1	Gao et al. (2006)	STARFM	Boreal Ecosystem–Atmosphere Study (BOREAS) Southern study area, Saskatchewan, Canada	Forest and some sparse vegetation; large patches, extreme phenology changes	4 LMPs	May–Sep 2001	1296	VA, SPO, MD, MAD
2	Gao et al. (2006)	STARFM	Western Iowa, USA	Cropland; small patches; distinct temporal patterns based on planting/harvest activities	1 LMP	Jul–Aug 2001	289	VA, SPD
3	Gao et al. (2006)	STARFM	Eastern Virginia, USA	Complex mixture of landcover: mixed forest and some cropland	3 LMPs	Feb–Jun 2001	225	VA, MD at MODIS resolution
4	Hilker et al. (2009a)	STARFM[a]	Western–Central Alberta, Canada	Mainly forest with herbal and shrub vegetation and patches of water and rocks; large land cover patches	3 LMPs, 110 8-MODIS	Mar–Oct 2002-2005	34,225	VA, VA of STARFM and STAARCH simulations
5	Hilker et al. (2009b)	STARFM	Central British Columbia, Canada	Mainly coniferous forest with subsidiary herbal and shrub vegetation and patches of water and rocks; large land cover patches	5 LMPs, 19 8-MODIS	May–Oct 2001	34,225	VA, SPN, r², T (whole image and land cover specific)
6	Chen et al. (2010)	ESTARFM	Qian-Yinzhou, Zhejiang, China	Mainly forest; patches of shrub and bare soil	7 LMPs, 33 8-MODIS	Oct–Nov 2004	900	r², T
7	Liu and Wang (2010)	STARFM[b]	Beijing, China	Winter wheat	3LMPs	Apr–Jun 2009	16,808	AEI
8	Zhu et al. (2010)	STARFM and ESTARFM	Central Virginia, USA	Forest, bare soil, water and urban land	3 LMPs	Jan–May 2002	~180[c]	VA, SPO, MD, MAD
9	Zhu et al. (2010)[d]	STARFM and ESTARFM	Boreal Ecosystem–Atmosphere Study (BOREAS) Southern study area, Saskatchewan, Canada	Forest and some sparse vegetation; large patches, extreme phenology changes	4 LMPs	May–Sep 2001	1296	VA, SPO, MD, MAD
10	Anderson et al. (2011)	STARFM	Orlando, Florida, USA	Irrigated fields, urban and wetland areas	2 LMPs, 9 daily MODIS thermal infrared data	Nov–Dec 2002	8432	Not performed[e]
11	Gaulton et al. (2011)	STARFM[f]	Rocky Mountains and foothills, Alberta, Canada	Mainly forest; various density road network	Unspecified LMPs; 8- MODIS	Landsat: summers of 2001, 2004, 2008; MODIS: 2001-2009	19,396 and 18,646	Not performed[g]
12	Liu et al. (2011)	STARFM	North-East of Beijing, China	Woodland, arable land, construction land, open water areas	1 LMP, 9 daily MODIS	May 2007	2230	Not performed[h]
13	Singh (2011)	STARFM	Meerut district, Uttar Pradesh, India	Mainly wheat and sugarcane crop; non-crop land including the Ganges river and villages	20 LMPs (2 per year); 8- MODIS	2000–2009	1250	r², T

14	Watts et al. (2011)	STARFM	North Central Montana, USA	Mainly spring and winter wheat and some barley fields	5LMPs, 26 daily MODIS-	May–Sep 2009	14,685	VA, r^2, SPO, T
15	Coops et al. (2012)	STARFM	Rocky Mountains and foothills, Alberta, Canada	Coniferous and mixed forest	2 LMPs, 32 8-MODIS	Feb–Sep 2009	34,225	Not performed [i]
16	Liu and Weng (2012)	STARFM	Los Angeles, California, USA	Mainly urban areas; flat and hilly terrains	3 AMPs	Jul–Dec 2007	3059	VA, TRMSD, MAD
17	Schmidt et al. (2012)	STARFM	Charters Towers, Queensland, Australia	Savanna	90 LMPs; 333 16-MODIS	Feb 2000–Sept 2007	120	VA, SPO, r^2
18	Walker et al. (2012)	STARFM	Flagstaff, Arizona, USA	Dryland forest, woodland	6 LMPs; 20 each daily, 8- and 16-MODIS	Apr–Oct 2006	13,949	VA, MAD, SPO, r^2
19	Wu et al. (2012)	STARFM	Hai Basin, Northern China	Wheat, maize, cotton fields	Landsat, MODIS[j]	2004–2008	320,000	Not performed [h]
20	Singh (2012)	STARFM	Uttar Pradesh, India	Mainly wheat, rice and sugarcane crop, patches of trees and bushes	16 LMPs; 46 8-MODIS	Mar 2002–Sep 2009	1250	SPO, r^2, T (whole image and land cover specific)
21	Bhandari et al. (2012)	STARFM	Queensland, Australia	Mainly forest	38 LMPs; 16-MODIS[k]	Jul 2003- Apr 2008	2600	SPO, SPFPC, r^2 (whole image and land cover specific)
22	Zhong et al. (2012)	STARFM	California, USA	Mainly irrigated areas of crops and pasture	72 LMPs; daily MODIS	2004 and 2010	3923	Not performed [i]
23	Current study	STARFM and ESTARFM	Coleambally, New South Wales, Australia	Irrigated fields, woodland, dryland agriculture	17 LMPs	Oct 2001–May 2002	2193	VA, MD, SPO, SRMSD, TRMSD[l] (whole image and land cover specific)
24	Current study	STARFM and ESTARFM	Gwydir, New South Wales, Australia	Irrigated fields, woodland, dryland agriculture, flood areas	14 LMPs	Apr 2004–Apr 2005	5440	VA, MD, SPO, SRMSD, TRMSD [l] (whole image and land cover specific)

[a] STARFM is used by the STAARCH model of Hilker et al. (2009a) which produces an image time series of disturbance events and requires Landsat–MODIS simulations.

[b] STARFM is used by the DASTARF model of Liu and Wang (2010) which optimizes STARFM outputs by assimilating observations into a dynamic system model.

[c] As no information was available about the exact extent in km^2 in (Zhu et al., 2010) we estimated the site size using the 17 May 2002 Landsat data downloaded from the GLOVIS website (http://glovis.usgs.gov).

[d] This study used the same data as in study 1.

[e] The study only used thermal imagery, which is not reflective.

[f] This study used the same model STAARCH as in study 4.

[g] Only validation of predicted disturbance patterns was described.

[h] Validation was performed on actual evapotranspiration, not reflectances.

[i] Validation was performed on derived data, not reflectances.

[j] No details about the Landsat and MODIS datasets were presented in this study.

[k] No exact number of MODIS 16-day composites used for simulating Landsat data was specified.

[l] SRMSD images and TRMSD plots are discussed in detail in Emelyanova et al. (2012).

periods of monitoring for both sites were 5 months with only 4 L–M pairs used for simulation. The area of the largest site was about 1300 km². Zhu et al. (2010) concluded that ESTARFM outperformed STARFM when applied to two heterogeneous landscapes by: (*i*) visual assessment; (*ii*) comparing observed vs. simulated reflectances on scatter plots; and (*iii*) calculating the mean difference and mean absolute difference between observed and simulated images (see Table 3, studies 8 and 9). Considering all the error metrics we reported (Figs. 4–9), ESTARFM clearly outperformed STARFM at Coleambally, while STARFM clearly outperformed ESTARFM at Gwydir. Because the spatial and temporal variances of each site fundamentally influence blending algorithm performance, and on the basis of our results we conclude that a universally best algorithm probably does not exist. That is, although ESTARFM is arguably in many ways an 'enhanced' version of STARFM, it does not always perform better. In particular, STARFM had a lower RMSD at Gwydir during the flood (e.g., Figs. 7o and 9e–f). Our study demonstrated that consideration of the spatial and temporal variances of imagery acquired for a site was needed to ensure optimum blending algorithm selection.

The results were very different at our two sites, which we attributed to both the effective patch size and the spatial and temporal variances (Fig. 2 and Section 2.2). The dominance of spatial variance at Coleambally reduced STARFM accuracy when compared to ESTARFM due to specific algorithm assumptions (Fig. 4). We showed that most of the variances at Coleambally occurred below the spatial resolution of MODIS (Fig. 2a, b, e, f), which meant that the inclusion of MODIS data could not greatly improve the blending accuracy. Since the MODIS data are meant to improve the modelling of the temporal dynamics, this particularly limited STARFM (compared to ESTARFM) as the main algorithm design advantage of STARFM (over ESTARFM) is its non-linear temporal assumption. The limited ability of MODIS to observe variance at Coleambally also meant that there was less separation between the four algorithms at Coleambally than at Gwydir (e.g., Fig. 3). This suggests that at sites where MODIS does not adequately sample the spatial variance, the extra effort of the more complicated blending algorithms have diminishing returns. In the case of a highly dense time series of L–M pairs (i.e., up to 3 a month) the simple algorithms like GEIFM and LIM could be an appropriate choice. However, the GEIFM-used global relationship between Landsat and MODIS should be less relevant as the spatial extent of a site increases. Our two sites were not very large, so the difference between GEIFM and STARFM or ESTARFM might not be as obvious as it would be if the sites were larger (e.g., spatial extent of an entire Landsat scene).

It does not seem commonplace to question whether a computationally expensive blending algorithm provides benefit over a simple one, but rather just presume that it does. Our results suggested that whether a site can substantially benefit from the inclusion of MODIS data can be determined by comparing the spatial and temporal variances through time at both Landsat and MODIS spatial resolutions (Fig. 2). When variances from the MODIS spatial resolution are substantially less than that of the Landsat spatial resolution, MODIS will likely not improve the simulation much (this should be confirmed at other sites).

We found that the alternating dominance of spatial and temporal variances at Gwydir across spectral bands made it a more challenging test for the blending algorithms. When green vegetation amount was maximal (DSSD 128), the NIR bands' spatial variance was dominant at both Landsat and MODIS spatial resolutions, and ESTARFM was more accurate than STARFM (Fig. 9e–f). During the flood (DSSD 256) the temporal variance of the SWIR bands was dominant at both Landsat and MODIS spatial resolutions (Fig. 2d and h) and STARFM was more accurate than ESTARFM (Fig. 9e–f). This result fits well with known weaknesses and strengths of these two algorithms (Gao et al., 2006; Zhu et al., 2010).

To date, a gap in blending research is that there has been no common metric to quantify the spatial and temporal dynamics of L–M pair time series in order to assess the suitability of blending algorithms. Here, we used the method of Sun et al. (2010) and determined that spatial and temporal variances as a metric, while very useful, did not have complete predictive power regarding algorithm performance. We attributed this largely to the effective domain characteristics of each site. However, if the actual temporal change really is linear and/or the time interval between images is short enough (i.e., the temporal density of images is high), then the single rate of change assumption would not reduce the accuracy of ESTARFM output much. STARFM is expected to outperform ESTARFM when temporal variance is relatively large and observed at both Landsat and MODIS spatial resolutions. Our results indicated that spatial and temporal variances were only relevant for predicting algorithm performance between ESTARFM and STARFM when Landsat and MODIS variances corresponded (Fig. 9e–f). Based on this result, we propose the two conceptual models presented in Fig. 9g–h where spatial and temporal variances might be used to guide algorithm selection. As discussed above, the relationship proposed in Fig. 9g–h will not hold under all situations. For example, if a site is similar to Coleambally, then STARFM may never outperform ESTARFM because the temporal dynamics modelled by MODIS are negligible due to the patch size of the site, thus removing STARFM's design advantage.

We encourage future comparisons of blending algorithms to: (*i*) report their study sites' spatial and temporal variances in order to put their results into a meaningful context; and (*ii*) further populate the data-space of the plots presented in Fig. 9 to test the framework we proposed. If spatial and temporal variances were reported for all studies as part of a 'site characterisation' exercise, comparing studies as in Table 3 would be easier. In a similar vein to assessing dataset complexity, currently there is no consistent manner of measuring spectrally dependent accuracy of the simulated reflectances (see Table 3). If there was an agreed standard used by the blending community this would be advantageous. As results can be site specific, to eradicate this difference when assessing new blending algorithms, the two L–M pair databases developed here are made freely available for further research.

5. Conclusion

We conclusively showed that ESTARFM did not always produce lower errors than STARFM, and that the advanced algorithms did not always produce lower errors than the simple algorithms, which address our first objective. Land cover was shown to be strongly associated with algorithm performance and was intricately linked with site domain characteristics, spatial and temporal variances, and spectral band. To address the second objective, we proposed a framework for blending algorithm selection based on how the overall variance is partitioned into the spatial and temporal components at both Landsat and MODIS spatial resolutions. Our results indicated that ESTARFM should, in general, be a superior blending algorithm for sites and/or times and for those spectral bands where spatial variance is dominant. Conversely, STARFM should generally be preferable when a given spectral band is dominated by temporal variance. This relationship was shown to not hold when the magnitudes of MODIS variances were much less than those of Landsat. Comparing Landsat resolution and MODIS resolution variances was suggested as a practical method to determine if and when MODIS could add much value to a blending exercise. If most of the variances occur at a finer spatial resolution than MODIS, then a researcher or manager may determine the site unsuitable for blending Landsat and MODIS, or may decide that complicated and computationally expensive blending algorithms should be replaced by simpler and computationally cheaper algorithms.

References

Anderson, M. C., Kustas, W. P., Norman, J. M., Hain, C. R., Mecikalski, J. R., Schultz, L., et al. (2011). Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery. *Hydrology and Earth System Sciences*, 15(1), 223–239.

Berk, A., Anderson, G. P., Bernstein, L. S., Acharya, P. K., Dothe, H., Matthew, M. W., et al. (1999). MODTRAN4 radiative transfer modeling for atmospheric correction. *Proceedings of SPIE, optical spectroscopic techniques and instrumentation for atmospheric and space research III, July 1999* (http://www.spectral.com/pdf/sr116.pdf)

Bhandari, S., Phinn, S., & Gill, T. (2012). Preparing Landsat image time series (LITS) for monitoring changes in vegetation phenology in Queensland, Australia. *Remote Sensing*, 4(6), 1856–1886.

Chen, B., Ge, Q., Fu, D., Yu, G., Sun, X., Wang, S., et al. (2010). A data-model fusion approach for upscaling gross ecosystem productivity to the landscape scale based on remote sensing and flux footprint modelling. *Biogeosciences*, 7(9), 2943–2958.

Coops, N. C., Hilker, T., Bater, C. W., Wulder, M. A., Nielsen, S. E., McDermid, G., et al. (2012). Linking ground-based to satellite-derived phenological metrics in support of habitat assessment. *Remote Sensing Letters*, 3(3), 191–200.

Crist, E. P., & Cicone, R. C. (1984). Comparisons of the dimensionality and features of simulated Landsat-4 MSS and TM data. *Remote Sensing of Environment*, 14(1–3), 235–246.

Emelyanova, I. V., McVicar, T. R., Van Niel, T. G., Li, L. T., & van Dijk, A. I. J. M. (2012). On blending Landsat–MODIS surface reflectances in two landscapes with contrasting spectral, spatial and temporal dynamics. *WIRADA project 3.4: Technical report.* Australia: CSIRO Water for a Healthy Country Flagship (72 pp. https://publications.csiro.au/rpr/pub?list=SEA&pid=csiro:EP128838)

Gao, F., Masek, J., Schwaller, M., & Hall, F. (2006). On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance. *IEEE Transactions on Geoscience and Remote Sensing*, 44(8), 2207–2218.

Gaulton, R., Hilker, T., Wulder, M. A., Coops, N. C., & Stenhouse, G. (2011). Characterizing stand-replacing disturbance in western Alberta grizzly bear habitat, using a satellite-derived high temporal and spatial resolution change sequence. *Forest Ecology and Management*, 261(4), 865–877.

Heinsch, F. A., Zhao, M., Running, S. W., Kimball, J. S., & Nemani, R. R. (2006). Evaluation of remote sensing based terrestrial productivity from MODIS using regional tower eddy flux network observations. *IEEE Transactions on Geoscience and Remote Sensing*, 44(7), 1908–1925.

Hilker, T., Wulder, M. A., Coops, N. C., Linke, J., McDermid, G., Masek, J. G., et al. (2009a). A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. *Remote Sensing of Environment*, 113(8), 1613–1627.

Hilker, T., Wulder, M. A., Coops, N. C., Seitz, N., White, J. C., Gao, F., et al. (2009b). Generation of dense time series synthetic Landsat data through data blending with MODIS using a spatial and temporal adaptive reflectance fusion model. *Remote Sensing of Environment*, 113(9), 1988–1999.

Li, F., Jupp, D. L. B., Reddy, S., Lymburner, L., Mueller, N., Tan, P., et al. (2010). An evaluation of the use of atmospheric and BRDF correction to standardize Landsat data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3(3), 257–270.

Liu, F., & Wang, Z. (2010). Synthetic Landsat data through data assimilation for winter wheat yield estimation. *18th international conference on geoinformatics, Peking University, Beijing, June 18–20, 2010* (http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5567524)

Liu, H., & Weng, Q. H. (2012). Enhancing temporal resolution of satellite imagery for public health studies: A case study of West Nile virus outbreak in Los Angeles in 2007. *Remote Sensing of Environment*, 117, 57–71.

Liu, S. F., Xiong, J., & Wu, B. F. (2011). ETWatch: A method of multi-resolution ET data fusion. *Journal of Remote Sensing*, 15(2), 255–269.

Ludwig, J. A., Bastin, G. N., Wallace, J. F., & McVicar, T. R. (2007). Assessing landscape health by scaling with remote sensing: When is it not enough? *Landscape Ecology*, 22(2), 163–169.

McVicar, T. R., Davies, P. J., Qinke, Y., & Zhang, G. (2002). An introduction to temporal-geographic information systems (TGIS) for assessing, monitoring and modelling regional water and soil processes. In T. R. McVicar, L. Rui, J. Walker, R. W. Fitzpatrick, & L. Changming (Eds.), *Regional water and soil assessment for managing sustainable agriculture in China and Australia* (pp. 205–223). Canberra: ACIAR Monograph 84 (http://www.eoc.csiro.au/aciar/book/PDF/Monograph_84_Chapter_16.pdf)

Phinn, S. R. (1998). A framework for selecting appropriate remotely sensed data dimensions for environmental monitoring and management. *International Journal of Remote Sensing*, 19(17), 3457–3463.

Renzullo, L. J., Barrett, D. J., Marks, A. S., Hill, M. J., & Guerschman, J. P. (2008). Multi-sensor model–data fusion for estimation of hydrologic and energy flux parameters. *Remote Sensing of Environment*, 112(4), 1306–1319.

Roy, D. P., Ju, J., Lewis, P., Schaaf, C., Gao, F., Hansen, M., et al. (2008). Multi-temporal MODIS–Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. *Remote Sensing of Environment*, 112(6), 3112–3130.

Schmidt, M., Udelhoven, T., Gill, T., & Roder, A. (2012). Long term data fusion for a dense time series analysis with MODIS and Landsat imagery in an Australian Savanna. *Journal of Applied Remote Sensing*, 6, 063512. http://dx.doi.org/10.1117/1.jrs.6.063512.

Singh, D. (2011). Generation and evaluation of gross primary productivity using Landsat data through blending with MODIS data. *International Journal of Applied Earth Observation and Geoinformation*, 13(1), 59–69.

Singh, D. (2012). Evaluation of long-term NDVI time series derived from Landsat data through blending with MODIS data. *Atmosfera*, 25(1), 43–63.

Sun, F., Roderick, M. L., Farquhar, G. D., Lim, W. H., Zhang, Y., Bennett, N., et al. (2010). Partitioning the variance between space and time. *Geophysical Research Letters*, 37, L12704. http://dx.doi.org/10.1029/2010gl043323.

van Dijk, A. I. J. M., & Renzullo, L. J. (2011). Water resource monitoring systems and the role of satellite observations. *Hydrology and Earth System Sciences*, 15(1), 39–55.

Van Niel, T. G., & McVicar, T. R. (2003). A simple method to improve field-level rice identification: Toward operational monitoring with satellite remote sensing. *Australian Journal of Experimental Agriculture*, 43(4), 379–387.

Van Niel, T. G., & McVicar, T. R. (2004a). Determining temporal windows of crop discrimination with remote sensing: A case study in south-eastern Australia. *Computers and Electronics in Agriculture*, 45(1–3), 91–108.

Van Niel, T. G., & McVicar, T. R. (2004b). Current and potential uses of optical remote sensing in rice-based irrigation systems: A review. *Australian Journal of Agricultural Research*, 55(2), 155–185.

Van Niel, T. G., McVicar, T. R., & Datt, B. (2005). On the relationship between training sample size and data dimensionality: Monte Carlo analysis of broadband multi-temporal classification. *Remote Sensing of Environment*, 98(4), 468–480.

Van Niel, T. G., McVicar, T. R., Fang, H., & Liang, S. (2003). Calculating environmental moisture for per-field discrimination of rice crops. *International Journal of Remote Sensing*, 24(4), 885–890.

Walker, J. J., de Beurs, K. M., Wynne, R. H., & Gao, F. (2012). Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology. *Remote Sensing of Environment*, 117, 381–393.

Watts, J. D., Powell, S. L., Lawrence, R. L., & Hilker, T. (2011). Improved classification of conservation tillage adoption using high temporal and synthetic satellite imagery. *Remote Sensing of Environment*, 115(1), 66–75.

Wu, B. F., Yan, N. N., Xiong, J., Bastiaanssen, W. G. M., Zhu, W. W., & Stein, A. (2012). Validation of ETWatch using field measurements at diverse landscapes: A case study in Hai Basin of China. *Journal of Hydrology*, 436, 67–80.

Zhong, L. H., Gong, P., & Biging, G. S. (2012). Phenology-based crop classification algorithm and its implications on agricultural water use assessments in California's Central Valley. *Photogrammetric Engineering and Remote Sensing*, 78(8), 799–813.

Zhu, X. L., Chen, J., Gao, F., Chen, X. H., & Masek, J. G. (2010). An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sensing of Environment*, 114(11), 2610–2623.

Zurita-Milla, R., Kaiser, G., Clevers, J. G. P. W., Schneider, W., & Schaepman, M. E. (2009). Downscaling time series of MERIS full resolution data to monitor vegetation seasonal dynamics. *Remote Sensing of Environment*, 113(9), 1874–1885.