Improving Wide-Area DEMs Through Data Fusion – Chances and Limits

KONRAD SCHINDLER, HARIS PAPASAIKA-HANUSCH, STEFAN SCHÜTZ, EMMANUEL BALTSAVIAS, Zürich

ABSTRACT

For almost any region of the world multiple wide-area DEMs are available nowadays. Thus, the question arises whether this redundancy can be used to fuse two (or more) DEMs and generate a new one that is more accurate than any of the inputs. We first give a brief overview over some popular DEMs covering large areas. We then review two possible fusion strategies, as well as methods for deriving point-wise accuracies/weights for the fusion process, since in most cases they are not supplied with the DEMs. In the last part of the paper we present experimental results for DEM fusion and a discussion of its benefits and limitations.

1. INTRODUCTION

Digital elevation models (DEMs) are one of the core products of photogrammetry and remote sensing, and one of the fundamental layers of geodata: they form the basis for orthophoto generation, mapping, and geo-visualisation, and they are a necessary ingredient for 3D planning and analysis in diverse application fields, including geology, hydrology, environmental modelling and urban planning. Furthermore, they are by their nature dense, thus airborne and satellite remote sensing are the only viable technologies to acquire large-area DEMs over large areas. This article is concerned with *wide-area* DEMs, by which we mean that they cover entire states or even continents. Such DEMs are usually supplied in the standard 2.5-dimensional raster representation.

In Section 2 we briefly repeat the main acquisition technologies for wide-area DEMs: matching of optical images, SAR interferometry, and airborne LiDAR. In Section 3 we give an overview and comparison of some popular wide-area DEMs. In Sections 4-5 review two methods for fusing DEMs from different data sources in order to obtain a new model with better accuracy and reliability. Section 6 is dedicated to an experimental evaluation of DEM fusion.

2. ACQUISITION TECHNOLOGY

The base data for DEM estimation are terrain points. Historically, these points were acquired manually through ground-based surveying or stereo measurement of contour lines or terrain points. These methods have now been superseded by less labour-intensive alternatives. Nowadays, there are three main technologies for acquiring 3D points over large areas.

Automatic image matching is the direct successor of manual stereo-photogrammetric measurements. The technology is nowadays mature. Since the advent of digital cameras, image blocks are routinely recorded with large overlaps and thus high redundancy, which allows one to generate DEMs with an accuracy comparable to airborne LiDAR (Leberl et al. 2010). The main disadvantages of the technology is that it can only deliver DSMs, and that, being a passive technique, it cannot cope well with shadows and untextured regions. When applied to optical satellite imagery, automatic matching can cover very large areas, making it suitable for wide-area DEM generation – albeit often with lower reliability, because only two or three scans are available.

Synthetic Aperture Radar (SAR) interferometry is a widespread technology in satellite remote sensing. It has the important advantage that data acquisition is independent of daylight, and mostly also atmospheric conditions. Depending on the wavelength, SAR in principle allows to generate both DSMs (X- and C-bands) and DTMs (L- and P-bands), however most satellite sensors operate in the short wavelengths and deliver DSMs. Airborne SAR is less widespread, but has also been used to generate wide-area DEMs, employing usually X- or C- band and P-band.

Airborne laser scanning has over the last decade become the dominant technology for smaller DEMs. The technology offers high density and accuracy with little processing overhead. A further advantage is that in vegetation areas the laser pulse partially penetrates the canopy, which allows one to acquire both a DSM and a DTM from the same sensor data. However, the technology is difficult to scale up to wide areas, because the costs for data capture increases considerably. At this point, LiDAR DEMs are not available for most of the world: complete coverage is only available for a few small countries like for example the Netherlands; for several other countries, especially in Europe, significant parts are covered (e.g. Switzerland, UK, Germany).

3. WIDE-AREA DEMS

In the following, we present important wide-area DEMs that are currently available, or may be available soon. DEMs which have been created by stitching together parts from different data sources (e.g. GTOPO30) are not discussed. Such DEMs generated by varying and time-wise different sources have very inhomogeneous data quality – essentially their properties in any given region are identical to those of the input used for that region.

SRTM (Shuttle Radar Topography Mission) acquired DEMs in February 2000 by single-pass SAR interferometry in the C- and X-bands (Farr et al. 2007). Here, only the C-band will be treated (the X-band data has large gaps due to the limited strip width). Thee are several versions, with or without holes, patched with data from other sources, etc. Here we discuss the "classical" product. SRTM-C covers the landmasses between 60°N and 56°S, but has local holes. The grid spacing is 30 m, but outside USA only a 90 m grid is publicly available. The specified accuracy is on the order of 15-20 m (90% confidence interval, see Table 1), and empirically the data meets these specifications (Rodriguez et al. 2006).

ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) GDEM was created from data acquired between 1999 and 2009 with stereo matching of image data in the visible and near-infrared range. It covers the landmasses between 83°N and 83°S at ~30 m grid spacing, with some small holes. The accuracy (95% confidence) is 20 m. Empirical valuations have shown that ASTER has somewhat inhomogeneous quality – while in most tiles the specifications (see Table 1) are met, there are regions with a significant amount of blunders as well as systematic artifacts (Reuter et al. 2009).

Reference3D is produced by stereo matching of optical images from the French SPOT-5 satellite acquired since 2002. The DEM product currently covers large parts of Europe, Africa and Central America, and the coastal regions of China and Australia, as well as some smaller regions. However, data is available for almost the entire landmass of the earth, and can be expected to be processed in the future. Reference3D is delivered with ~30 m grid spacing. There are no holes in the available tiles. The accuracy is 10-30 m depending on the terrain slope, with relative accuracies about a factor two higher. These accuracies have been confirmed empirically (Bouillon et al. 2006).

NEXTmap. The company Intermap offers a DEM of the USA and Western Europe, acquired with airborne SAR interferometry. The DEM has a grid spacing of 5 m, and a nominal accuracy of 2 m for the planimetry and 1m for the height, which is reached in open terrain, but not necessarily in vegetation areas (Dowman et al. 2003).

Other DEMs. Several other satellites not designed for DEM generation have nevertheless been used for that purpose. Examples include tandem interferometry with ERS-1 and ERS-2, and repeat-pass interferometry with ALOS PALSAR. These projects have resulted in elevation models at the regional and national level (e.g. DEMs computed from ERS data exist for Switzerland, Czech Republic, the UK, Egypt, and possibly other countries), however there are no systematic efforts to create DEMs from this data, and the data coverage is patchy. High spatial resolution satellites have also been used for DEM generation using optical sensors, but due to data availability and costs only for rather small areas.

Future projects. Several projects are currently underway to generate new, more precise wide-area DEMs. A notable example is the **TanDEM-X** mission, which uses a pair of X-band SAR satellites in close formation flight (Krieger et al. 2007). The mission is currently in its operational phase, and is expected, over the next 2.5 years, to cover all land masses of the earth at ~12 m grid spacing at an accuracy of about 10 m, but with relative accuracies <4 m. Another example is Microsoft's **Global Ortho** project, which aims at creating high-quality orthophotos of the USA and Europe from airborne imagery. In the process, a DEM will be created by multi-image matching, which can be expected to be accurate enough for orthophoto production at 30 cm GSD. Specifications are not available, and it is not known whether a lower-resolution version might be released to the public.

	SRTM-C	**ASTER GDEM**	**Reference3D**	**Intermap**	**TanDEM-X**
data capture	02/2000	1999-2009	Since 2002	Since 2004	2011-2013
technology	InSAR	matching	matching	airborne InSAR	inSAR
coverage	60°N-56°S	83°N-83°S	30% land masses	USA, W-Europe	all landmasses
grid resolution	90 m / 30 m	30 m	30 m	5 m	12 m
abs. Z-error @90%	16 m	17 m	10-30 m	1 m	<10 m
rel. Z-error@90%	10 m	–	5-28 m	1 m	<4 m
abs. XY-error@90%	20 m	25 m	15 m	2 m	<20 m
rel. XY-error@90%	15 m	–	8 m	2 m	<3 m
Cost	free	free	~10 US$ / km^2	~30 US$ / km^2	unknown

Table 1: Characteristics of important wide-area DEMs.

4. DEM FUSION

The overview emphasises that nowadays multiple DEMs of comparable accuracy exist already for almost any region of the earth. The situation gives rise to the natural question, whether this redundancy can be exploited to create a DEM of higher accuracy than any of the available ones. The optimal solution from the point of view of estimation theory would of course be to obtain the raw measurements and sensor models from all sensors, and fit a single DEM to the entire set of heterogeneous observations in one shot. Unfortunately this is usually not practically viable. A more realistic alternative is to fuse DEMs from different sources into a higher-quality product (and in the process estimate its quality from the redundancy). For completeness we point out that there is also an intermediate solution, namely to regard the existing DEMs as "raw measurements" of 3D points and fit a DEM to them. This approach however has the – at least conceptual – weakness that one

cannot know whether the employed fitting method is compatible with the one used to generate the input DEMs.

Here, we concentrate on the real fusion case, i.e. the existing DEMs are regarded as (usually 2.5-dimensional) non-parametric surfaces represented by samples on regular grids, and the goal is to obtain a new surface in the form of new height values for the denser one of the two grids. NB: filling the holes in a digital elevation model by stitching in data from another one is also sometimes called "fusion" in the literature, or at least seen as one step of fusion (e.g. Constantini et al. 2006, Rao et al. 2003, Slatton et al. 2002). Hole-filling is not investigated here, neither is mosaicking of DEMs to increase coverage, which is also sometimes referred to as "fusion".

The crucial technical difficulty of DEM fusion is that it requires weights to quantify the influence of the inputs at every surface location. These weights are a function of the relative height accuracy, and typically vary significantly across each DEM, due to the sensor technology, scene characteristics and method used to generate it (e.g. models based on SAR interferometry have larger errors in urban areas). These weights (sometimes referred to as "height error maps") should be a natural by-product of DEM generation, but are nevertheless often not available, in which case one can try to estimate them statistically from local surface or scene properties.

In the following, we briefly review two exemplary methods for DEM fusion, and present an empirical investigation of how to derive fusion weights from a given DEM. We then go on to experimentally evaluate DEM fusion for a test site in central Switzerland, and discuss the findings of that study.

Preprocessing. Experience shows that systematic differences between the geo-referencing of two independent DEMs are the norm rather than the exception, both in planimetry and height. We therefore follow the common practice (e.g. Grün and Akca 2005, Constantini et al. 2006) and co-register the two DEMs before fusion. We note that errors in height can in principle also result from different geoid corrections. This effect is not of relevance to the tests presented here, but may have to be taken into account when working over very large areas.

DEM fusion by weighted averaging. Probably the most obvious fusion algorithm is to simply resample the two input DEMs to the same grid and compute an output value at each grid node by weighted averaging of the inputs. This approach is used frequently (e.g. Constantini et al. 2006, Reinartz et al. 2005, Schultz et al. 1999, Xu et al. 2010). In practice it works surprisingly well, which confirms the intuition that the correct choice of weights has a far greater influence on the results than the mathematical recipe for fusing the inputs. Its main weaknesses from an algorithmic point of view are on one hand that it looks at each grid point independently, disregarding the resulting surface shape; and on the other hand that the up-sampling of one of the input DEMs could potentially introduce further artefacts – which however is a largely theoretical concern with little practical relevance.

DEM fusion with sparse representations. To mitigate the first weakness of weighted averaging, it has been proposed to exploit the framework of sparse representations for DEM fusion (Papasaika et al. 2011). In a nutshell, the rationale is the following: to be more robust to blunders it would be desirable to include prior knowledge about plausible surface shapes, such that improbable local surface geometries caused by blunders are suppressed.

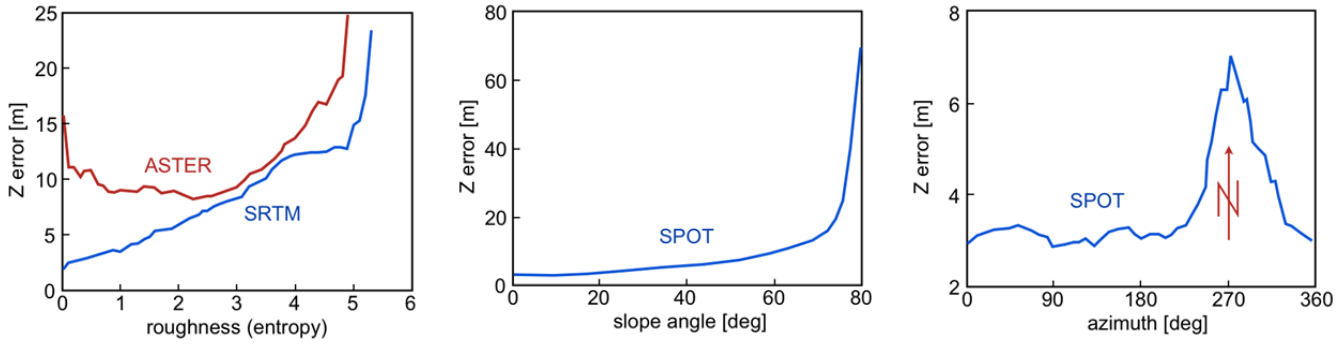
Such a fusion scheme can be designed by representing a local DEM patch as a sparse combination of basis patches, such that the basis can (theoretically) represent all local terrain shapes. The use of

patches instead of single height values, together with the sparsity constraint (i.e. only very few basis patches shall be combined to generate the output) regularises the output to probable surface shapes; generating the generic patch basis by down-sampling patches from a DEM with much higher grid spacing avoids sampling artefacts. The fused DEM is estimated by minimising the deviations from the two inputs while enforcing that both be represented by the same sparse combination of basis patches. It is well-known that sparsity of a linear combination can be achieved by minimising the L_1-norm of the coefficient vector (e.g. Tibshirani 1996). Overall, the estimation thus amounts to solving a L_1-regularised least-squares system, for which efficient solvers exist (e.g. Mallat 1998).

Fusion weights. For both described methods (and most other conceivable fusion schemes) weights are required, which govern the relative influence of the two input DEMs at a given raster location. These weights are critical for proper DEM fusion. Under the simplifying assumption that there are no blunders, these weights can be derived by standard error propagation: it is easy to confirm that given two direct observations (z_1, z_2) with standard deviations (s_1, s_2), the weights for a weighted average are $z=w_1z_1+w_2z_2$ are $w_1=s_2^2/(s_1^2+s_2^2)$ and $w_2=s_1^2/(s_1^2+s_2^2)$.

If such uncertainties s_i are indeed available for every single raster point (sometimes called "height error maps", e.g. Knöpfle et al. 1998, Reinartz et al. 2006), computing the weights is thus straightforward. Unfortunately, this is often not the case. In many cases only a global uncertainty for the entire DEM is reported, while local variations of the errors are lost. Simply using this global number everywhere is unsatisfactory, because the physical characteristics of different sensors, data acquisition conditions, processing methods and scene characteristics (especially geomorphology and land-cover) imply strongly varying accuracy. Some DEM providers take this into account in a crude fashion by reporting several standard deviations, e.g. for different classes of slope or land-cover, e.g. vegetation and buildings.

If adequate error measures are not provided, it is natural to ask whether the missing accuracies can be estimated directly from the DEM. To investigate this question we have analysed the residuals of different DEMs w.r.t. LiDAR ground truth. It turns out that in fact clear correlations exist between height errors and certain surface properties. In Figure 1 we show examples for surface roughness, surface slope, and aspect angle.



(a) (b) (c)

Figure 1: Relation between surface properties and DEM errors: (a) The errors grow with increasing terrain roughness for SRTM, whereas they initially decrease for ASTER – very low roughness typically means a lack of texture for image matching; (b) The errors of SPOT show a high correlation with the slope ; (c) SPOT flies over Switzerland around 10-10.30am, so depending on the sun elevation its images have shadows on north slopes, leading to matching errors.

5. EXPERIMENTS

Test site. Our test site is located in the area of Thun in central Switzerland. It features a varied geomorphology (small lakes, flat river valley, hills, steep mountain slopes), and variable land-cover (cropland, denser and sparser urban areas, forest, alpine rock and gravel). The site covers approximately 12 x 17.5 km^2 with a height range of about 1600 m.

Test data. Three different DEMs are available for pair-wise fusion, and one significantly more accurate one as ground truth. All DEMs were co-registered. No measures were taken to account for errors due to different acquisition times and different penetration depth of the sensor into the vegetation.

SPOT Reference3D with 30 m grid spacing, produced by *SpotImage* by stereo matching of images (acquisition date 30.09.2002). The specified absolute elevation accuracy is 10 m for flat or rolling terrain (slope ≤20%), 18 m for hilly terrain (20% ≤ slope ≤ 40%), and 30 m for mountainous terrain (slope >40%). For our test site, the overall empirical accuracy (root mean square error w.r.t. ground truth) is 15 m. Since the specification is too coarse for DEM fusion, we estimated a per-point accuracy from the slope (see Papasaika et al. 2011).
ALOS/PALSAR **DEM** with 15 m grid spacing, produced by *sarmap S.A.* by repeat-pass interferometry with the L-band sensor of the ALOS satellite (image acquisition date: master 19.06.2006, slave 04.08.2006). The DEM comes with a full height accuracy map estimated from the signal coherence, which however in our experience is not always reliable. The overall empirical accuracy w.r.t ground truth is 20 m.
ERS-1/2 Tandem DEM with 25 m grid spacing, also produced by *sarmap S.A.* by repeat-pass interferometry with the C-band sensors of the ERS-1 and ERS-2 satellites (several passes, acquisition 1995-1998). Again, a full height error map is supplied. The overall empirical accuracy w.r.t. ground truth is 11 m.

As ***ground truth*** we use a LiDAR DEM provided by *swisstopo*, with 2 m grid spacing. The airborne LiDAR data was acquired in 2000 with a mean density of 1-2 point per m^2 and with first and last pulse recorded. The specified accuracies (1 std. dev.) is 0.5 m in open country and 1.5 m in vegetation.

Fusion ALOS – SPOT

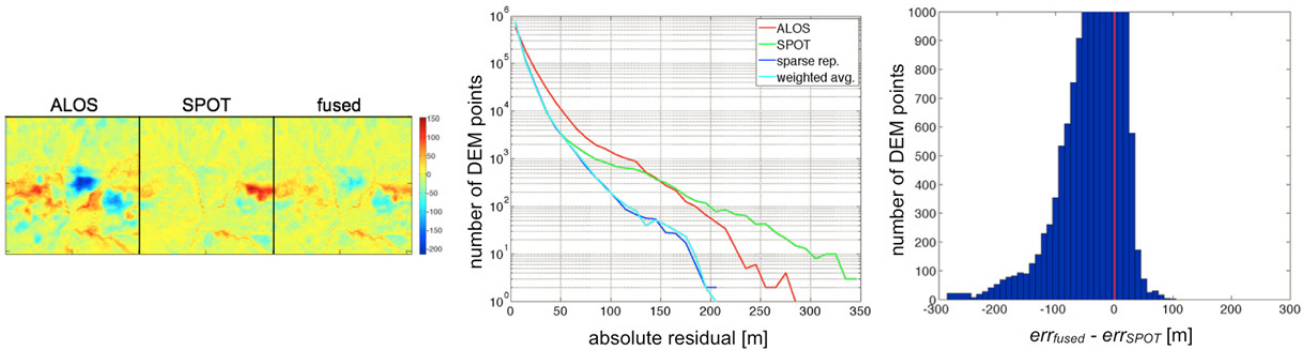
The fused DEM has a size of 800×1167 postings at a grid spacing of 15 m. Quantitative results are given in Table 2.The following values are given: the mean deviation (*mean*), in order to estimate the remaining bias in the co-registration, the familiar root mean square error (*RMSE*), the median absolute deviation from the median (*MAD*) as a robust measure of the accuracy without gross errors, and the maximum absolute deviation (*ABSMAX*) to quantify the size of the outliers.

The fusion result corresponds to a 44% improvement over ALOS, while maintaining the resolution, respectively a 29% improvement over SPOT, while doubling the resolution to 15 m. Note, error propagation with the global RMSE of the inputs (19.4 m and 15.3 m) yields a theoretical expected RMSE for the output of 12.1 m, so using the proper weights does significantly improve the fusion.
Figure 2(a) demonstrates the effect of DEM fusion with a qualitative example showing the reduction of blunders in the fused DEM. Figure 2(b) shows the errors histogram of the ALOS and SPOT DEMs as well as the fusion results with both weighted averaging and sparse representation. The analysis shows that the main improvement is due to the removal of gross errors. The fusion results have a similar distribution as SPOT below 50 m, i.e. fusion does not reduce small random

errors. However, both inputs contain a significant number of errors >50 m – ALOS has more blunders in total, but SPOT has a higher number of extremely high errors >200 m. Note, the *y*-axis is logarithmic to make the plot more readable. Figure 2(c) shows how the error distribution changes due to the fusion, compared to the more accurate input SPOT. The *x*-axis shows the error difference (err_{fused} - err_{SPOT}), between the absolute height errors, so negative values mean that fusion brought an improvement. The overall distribution is significantly skewed towards negative values, meaning that the heights of many more points were improved than deteriorated by the fusion process.

	MEAN	**RMSE**	**MAD**	**ABSMAX**
ALOS	-1.0	19.3	6.6	280.6
SPOT	-1.6	15.4	4.4	349.1
Weighted avg.	-1.0	10.9	4.1	202.8
Sparse rep.	-1.0	10.9	4.2	205.1

Table 2: Results of fusing the ALOS and SPOT DEMs. All errors are measured w.r.t. the LiDAR data as ground truth. All values are in (m).



(a) (b) (c)

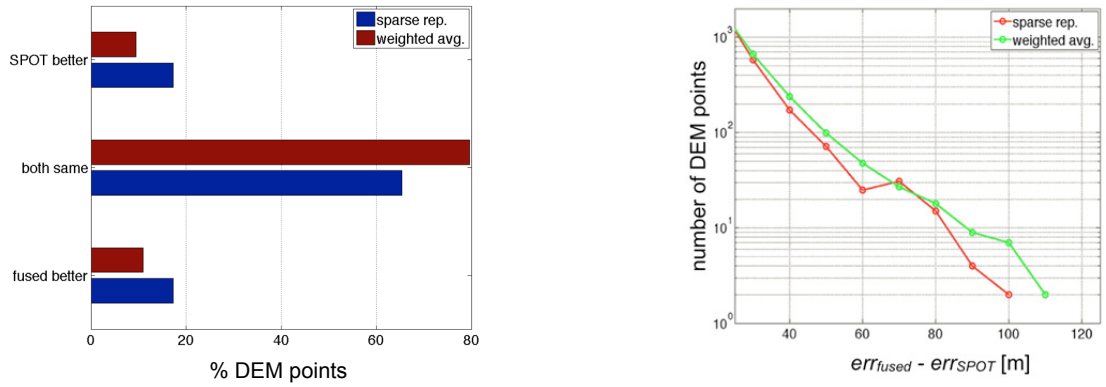
Figure 2: Results of fusing the ALOS and SPOT DEMs. (a) Example how fusion reduces gross errors – note, deviations from ground truth range from -200 m to 150 m; (b) histogram of absolute errors of different DEMs); (c) differences between the absolute residuals of fused DEM and SPOT DEM (negative values mean fused is better).

A comparison of the weighted average and sparse representation methods shows that overall their performance is very similar – the regularisation built into the sparse representation framework does not seem to make a big difference, see Figure 2(b). This is in agreement with the intuition that the crucial factor is not the mathematical recipe for fusion, but rather the correct determination of the fusion weights (respectively relative accuracies). However, the two methods do behave differently. Compared to the more accurate input SPOT, weighted averaging changes fewer points, i.e. more of the output is identical to the SPOT DEM – see Figure 3(a). On the contrary, the sparse representation method is more aggressive and deviates from SPOT more often when using the identical weights, but not always to the better. More errors are reduced, but also more are aggravated. Putatively, this is due to the additional constraints on the terrain shape imposed by the regularisation. A further observation is that weighted averaging, although on average staying closer to SPOT, makes the bigger mistakes – it *increases* noticeably more height errors than the sparse representation method. We point out that the experience with the sparse representation method is preliminary and more research is needed how to optimally adapt it to the fusion problem.

Finally we also note that the influence of DEM fusion depends on the local characteristics of the terrain. In our experience, larger improvements are made in areas of high slope and high roughness, which is not too surprising, since the inputs tend to be more inaccurate in these regions.

Fusion ERS-SPOT

The fused DEM has a size of 480×700 postings at a grid spacing of 25 m. Quantitative results are given in Table 3. The fusion result corresponds to a 37% improvement over SPOT, while increasing the resolution from 30 to 25 m, respectively an 11% improvement over ERS, at the same resolution. Error propagation with the global accuracies of the inputs leads to a theoretical expected RMSE of 8.8 m, which is not reached – most likely due to inaccurate weights.



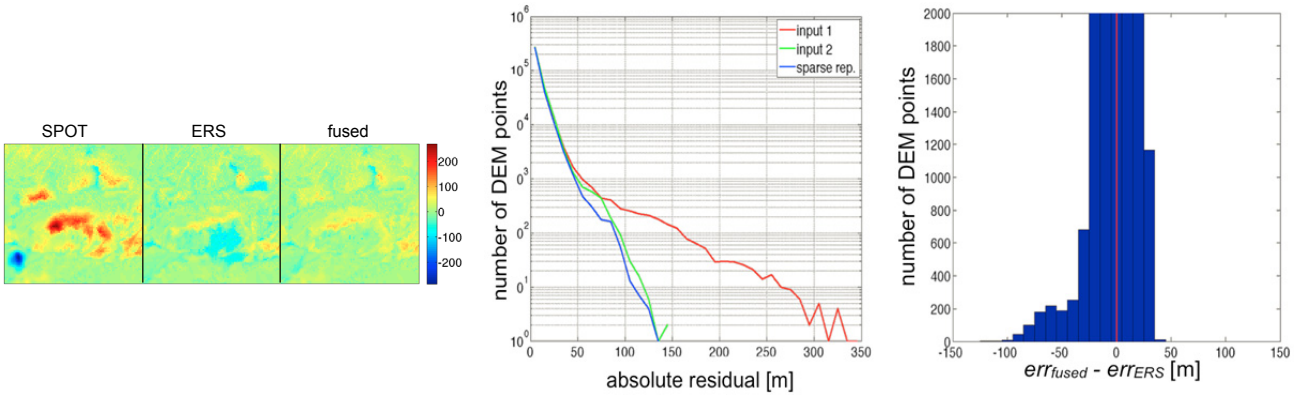
(a) (b)

Figure 3: Comparison between weighted averaging and sparse representation. (a) The sparse representation method more aggressively changes height values. Weighted averaging improves fewer points, but also makes fewer mistakes; (b) detail from the plot of changes w.r.t. SPOT, c.f. Figure 2(b); generally both methods perform similarly, however weighted averaging exhibits a tendency towards big mistakes, where the error increases by >30 m.

Figure 4(a) demonstrates the effect of DEM fusion on a qualitative example. Figure 4(b) shows the histogram of the errors of the ERS and SPOT DEMs as well as the fusion result. The main improvement again comes from removing gross errors >50 m. Note, the *y*-axis is again logarithmic. Figure 4(c) shows how the errors change due to the fusion, compared to the more accurate input ERS. The heights of many more points were improved than deteriorated by the fusion process. There is practically no difference between the two fusion algorithms, although weighted averaging still has a slight tendency to greatly increase the error for a small number of points.

	MEAN	**RMSE**	**MAD**	**ABSMAX**
ERS	0.1	10.8	3.1	144.6
SPOT	-1.6	15.3	4.5	345.3
Weighted avg.	-1.1	9.6	2.7	131.4
Sparse rep.	-1.1	9.6	2.7	131.3

Table 3: Results of fusing the ERS and SPOT DEMs. All errors are measured w.r.t. the LiDAR data as ground truth. All values are in (m).



(a) (b) (c)

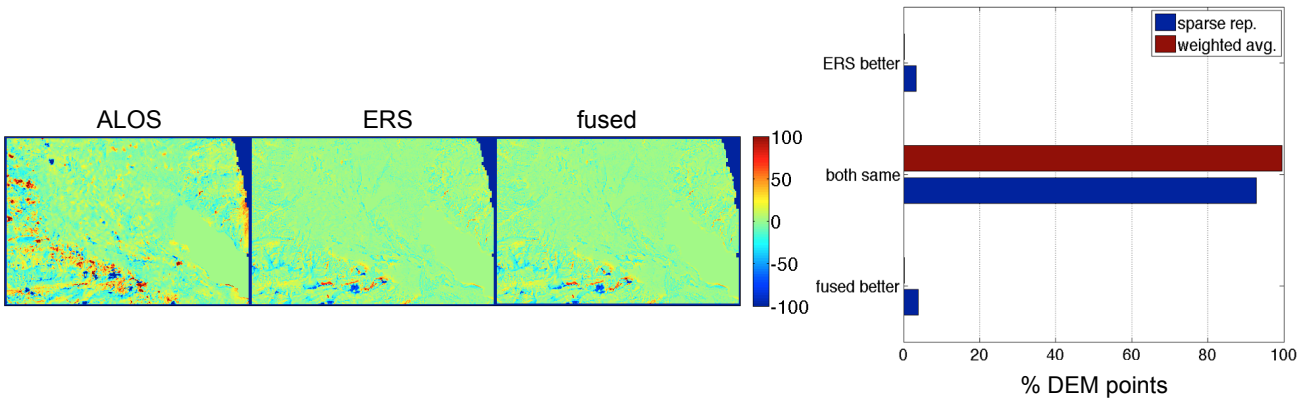
Figure 4: Results of fusing the ERS and SPOT DEMs. (a) Example how fusion reduces gross errors; (b) histogram of absolute errors of different DEMs; (c) differences between the absolute residuals of fused DEM and ERS DEM (negative values mean fused is better).

Fusion ALOS – ERS

The fused DEM has a size of 800×1167 postings at a grid spacing of 15 m. This experiment shall illustrate the limits of DEM fusion. Firstly, the two inputs have very different accuracy (almost a factor of 2), and secondly they have both been generated with the same technology (interferometric SAR), such that we expect them to be less complementary. The quantitative results in Table 4 show that fusion does not improve the ERS data significantly. Error propagation with the global accuracies of the inputs leads to a theoretical expected RMSE of 9.4 m, which is not reached. A closer analysis reveals that even in areas of high slope or roughness, where the ERS DEM is known to contain some blunders, the fused DEM did not improved at all, meaning that ALOS can not contribute any useful extra information. Figure 5(a) shows the error maps for both input and the fused DEM – the output is essentially identical to ERS. The quantitative analysis in Figure 5(b) further confirms that weighted averaging almost does not change the input, and the sparse representation approach only alters <10% of all grid points, without making an improvement.

	MEAN	**RMSE**	**MAD**	**ABSMAX**
ALOS	-1.0	19.3	6.6	280.6
ERS	0.1	10.8	3.1	159.2
Weighted avg.	0.1	10.7	3.1	159.3
Sparse rep.	0.1	10.7	3.1	156.8

Table 4: Results of fusing the ALOS and ERS DEMs. All errors are measured w.r.t. the LiDAR data as ground truth. All values are in (m).



(a) (b)

Figure 5: Fusion of ALOS and ERS DEMs. (a) No improvement is possible, the output is practically identical to the more accurate input ERS over the entire test site; (b) the quantitative analysis confirms that ERS prevails almost unchanged.

6. DISCUSSION

Chances. The experiments confirm that significant improvements are possible by fusion of existing DEMs – in the ALOS+SPOT case the RMSE was reduced by 29% compared to the more accurate input. Most of the gain is due to the reduction of gross errors, of which there are quite many in large-area DEMs. The combination of complementary DEM generation technologies, such as for example interferometric SAR and optical image matching, has the biggest potential, because the different recording and processing principles introduce different blunders, which can be at least partially remedied with more correct data from the other technique.

Although different fusion algorithms are available, the choice which mathematical toolbox to employ makes little difference. More important is to use the correct influence weights for the inputs. Fusion requires fine-grained knowledge of the local uncertainty to determine the appropriate fusion weights. If these are not supplied together with a DEM, it is nevertheless feasible to determine them from local geomorphological characteristics and land-cover.

Limits. Models generated with similar technology and/or very different accuracy are less suitable for fusion, since they tend to contain blunders in the same areas, and the weaker DEM will not be able to contribute any important information. Furthermore, one needs to be aware that even in the case of successful fusion, the errors are increased for some points.

Regarding the data-driven estimation of fusion weights, the method requires access to training data consisting of both estimated heights and residuals, thus it can only be done if DEM data as well as sufficiently accurate ground truth is available, such that the dependency between surface properties and height uncertainties can be estimated for the specific DEM generation technology.

Open questions. Here, we have only investigated the fusion of two input DEMs. While the described methods can trivially be extended to more inputs, an empirical investigation is still missing. The use of more than two DEMs can also support blunder detection. A further interesting domain is the fusion of airborne photogrammetry and airborne LiDAR, which could potentially be useful for high-accuracy applications. The methods we have investigated are generic, but again the application to high-density airborne data remains to be verified experimentally.

So far we have only looked at the statistics of height errors w.r.t ground truth, an error measure that is independent of any particular application. For certain problems, DEM quality may not be primarily the per-point uncertainty, e.g. for hydrological applications it may be more important to correctly delineate watersheds, even at the cost of higher overall errors. We have not investigated this issue yet, and at least in principle it is quite possible that altering heights to reduce their individual residuals significantly changes gradient directions and thus run-off patterns, especially in flat regions.

7. REFERENCES

Bouillon, A., Bernard, M., Gigord, P., Orsoni, A., Rudowski, V., Baudoin, A. (2006): SPOT 5 HRS geometric performances: using block adjustment as a key issue to improve quality of DEM generation. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 60, Issue 3, pp. 134-146.

Costantini, M., Malvarosa, F., Minati, F., Zappitelli, E. (2006): A data fusion algorithm for DEM mosaicking: building a global DEM with SRTM-X and ERS data. In: IEEE International Geoscience and Remote Sensing Symposium, Vol. 8, pp. 3861-3864.

Dowman, I., Balan, P., Renner, K., Fischer, P. (2003): An evaluation of NEXTMAP terrain data in the context of UK national datasets. Report to Getmapping 19. University College London, UK.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimanda, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D. (2007): The Shuttle Radar Topography Mission. Reviews of Geophysics, Vol. 45, Issue 2.

Grün, A., Akca, D. (2005): Least squares 3D surface and curve matching. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 59, Issue 3, pp 151-174.

Knöpfle, W., Strunz, G., Roth, A. (1998): Mosaicking of digital elevation models derived by SAR interferometry. International Archives of Photogrammetry and Remote Sensing, Vol. 32, Part 4, pp. 306-313.

Krieger, G., Moreira, A., Fiedler, H., Hajnsek, I., Werner, M., Younis, M., Zink, M. (2007): TanDEM-X: A satellite formation for high-resolution SAR interferometry. IEEE Transactions on Geoscience and Remote Sensing, Vol. 45, Issue 11, pp. 3317-3341.

Leberl, F., Irschara, A., Pock, T., Meixner, P., Gruber, M., Scholz, S., Wiechert, A. (2010): Point clouds: Lidar versus 3D vision. Photogrammetric Engineering & Remote Sensing, Vol. 76, Issue 10, pp. 1123-1134.

Mallat, S. (1998): A Wavelet tour of signal processing, 2nd edition. Academic Press.

Papasaika, H. Kokiopoulou, E., Baltsavias, E., Schindler, K., Kressner, D. (2011): Fusion of digital elevation models using sparse representations. Photogrammetric Image Analysis, Munich, Germany, to appear.

Rao, Y. S., Rao, K. S., Venkataraman, G., Khare, M., Reddy, C.D. (2003): Comparison and fusion of DEMs derived from InSAR and optical stereo techniques. Proc. 3rd ESA International Workshop on ERS SAR Interferometry, 1-5 December, Frascati, Italy.

Reinartz, P., Müller, R., Hoja, D., Lehner, M., Schröder, M. (2005): Comparison and fusion of DEM derived from Spot HRS stereo data and SRTM data and estimation of forest heights. EARSeL Workshop on 3D Remote Sensing, 10-11 June, Porto, Portugal (on CD-ROM).

Reuter, H. I., Nelson, A., Strobl, P., Mehl, W., Jarvis, A. (2009): A first assessment of ASTER GDEM tiles for absolute accuracy, relative accuracy and terrain parameters. International Geoscience and Remote Sensing Symposium, Vol. 5, pp. 240-243.

Rodriguez, E., Morris, C. S., Belz, J. E. (2006): A global assessment of the SRTM performance. Photogrammetric Engineering & Remote Sensing, Vol. 72, No. 3, pp. 249-260.

Schultz, H., Riseman, E.M., Stolle, F.R., Woo, D.-M. (1999): Error detection and DEM fusion using self-consistency. IEEE International Conference on Computer Vision, Vol. 2, pp. 1174-1181.

Slatton, K.C., Teng, S., Crawford, M. (2002): Multiscale fusion of InSAR data for hydrological applications. Proc. of 2002 AIRSAR Earth Science and Application Workshop, available at http://airsar.jpl.nasa.gov/documents/workshop2002/papers/S8.pdf (accessed 29th June, 2011).

Tibshirani, R. (1996): Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society B, Vol. 58, Issue 1, pp. 267-288.

Xu, C., Wei, M., Griffiths, S., Mercer, B., Abdoullaev, R. (2010): Hybrid DEM generation and evaluation from spaceborne radargrammetric and optical stereoscopic DEM. Proc. of Canadian Geomatics Conference, available at http://www.isprs.org/proceedings/XXXVIII/part1/11/11_02_Paper_123.pdf (accessed 29th June, 2011).