APPLYING TERRAIN AND HYDROLOGICAL EDITING TO TANDEM-X DATA TO CREATE A CONSUMER-READY WORLDDEM PRODUCT

J. Collins, Dr. G. Riegler, H. Schrader, M. Tinz

Airbus Defence and Space, Geo-Intelligence (john.collins@airbus.com) - Claude-Dornier-Straße, 88090 Immenstaad, Germany

KEY WORDS: WorldDEM, TanDEM-X, TerraSAR-X, SAR, DSM, elevation model

ABSTRACT:

The Geo-intelligence division of Airbus Defence and Space and the German Aerospace Center (DLR) have partnered to produce the first fully global, high-accuracy Digital Surface Model (DSM) using SAR data from the twin satellite constellation: TerraSAR-X and TanDEM-X. The DLR is responsible for the processing and distribution of the TanDEM-X elevation model for the world's scientific community, while Airbus DS is responsible for the commercial production and distribution of the data, under the brand name WorldDEM™.

For the provision of a consumer-ready product, Airbus DS undertakes several steps to reduce the effect of radar-specific artifacts in the WorldDEM data. These artifacts can be divided into two categories: terrain and hydrological. Airbus DS has developed proprietary software and processes to detect and correct these artifacts in the most efficient manner. Some processes are fully-automatic, while others require manual or semi-automatic control by operators.

1. CREATING A CONSUMER-READY WORLDDEM PRODUCT

1.1 The TanDEM-X Mission and WorldDEM Product

Airbus Defence and Space and the German Aerospace Center (DLR) have partnered to produce the first truly global, high-accuracy Digital Surface Model (DSM) using the interferometric data from the twin satellite constellation: TerraSAR-X and TanDEM-X. The DLR is responsible for the processing and distribution of the TanDEM-X elevation model to the world's scientific community, while Airbus DS is responsible for the commercial production and distribution of the data, under the brand name WorldDEM™.

1.2 Radar Effects in Processed SAR Elevation Data

The WorldDEM DSM is derived from the phase information from two Synthetic Aperture Radar (SAR) signals using interferometric processing of X-band radar signals. The height information is obtained for any point on the earth's surface by calculating the phase difference received by the two receiving antennas on the mission's two satellites (TerraSAR-X or TanDEM-X). In order to accurately measure surface heights from space, it is essential that the orbits (i.e. height from the ground) of the satellites be precisely measured. In addition, the separation of the two receiving antennas must also be known extremely precisely. Fortunately, the precision of the two satellites' orbits is very well understood, as is the separation between them in flight. (DLR 2013)

The TanDEM-X mission recorded height information for the whole earth's surface at least two times. (DLR 2013) In more "challenging" areas (steep, sandy or heavily-forested areas) up to four acquisitions were made, in order to improve the resulting surface model. These additional acquisitions took advantage of different acquisition geometries: either looking at the earth's surface from another orbital direction, or by directly adjusting the angle of signal acquisition, with reference to the earth's surface. These techniques allowed the satellites to "see" more surfaces in challenging terrain and enabled the improved acquisition of surfaces that were more likely to deflect radar waves away from the receiving antennas on the satellites.

The acquisition strategies described above were undertaken to ensure a maximum coverage of the earth's surface, with a maximum vertical accuracy. Preliminary results indicate that the strategy appears to have been successful, with a minimum of areas with no data (voids) and a corresponding high accuracy. (Bachmann 2013) However, it was always anticipated that there would be certain areas of the earth that would not interact well with X-band radar signals and would therefore subject to some invalid and or inaccurate data being collected and processed.

The first surface that is highly susceptible to invalid height elevations in the DSM is waterbodies. Early-on during research on radar signal propagation, it was known that smoother surfaces, such as water, can be highly reflective to short-band radar waves (Davies 1954 and Grant, Yaplee 1957) As radar later became a tool for mapping with the start of widespread X-band SAR data collections (SRTM and commercial airborne radar) it became well understood that the relatively short wavelengths (~3cm) of X-band radar are frequently reflected by water surfaces, away from the receiving antennas (Farr et.al. 2007). The almost complete reflection of the radar signals by water surfaces means that the receiving antennas have little or no information about the surface heights of water. When the SAR data comes to be processed, there is no interferometric baseline for the processor to set an elevation over water: therefore the heights calculated for most water surfaces appear to be nothing more than random noise in the resulting DSM. In areas of more widespread water, the SAR processors are usually designed to detect the lack of signal coherence and assume that an area is water. In these larger areas, no surface elevation is calculated and the elevations are simply set using the "no data" flag value of -32767m.

While there are clearly difficulties with X-band radar SAR data over water, it should be clear that this is not an insoluble problem. While the elevations calculated over the surface of water are not accurate, the elevations calculated on the surfaces immediately adjacent to water (shorelines) are in fact coherent, accurate and consistent. This means that while the elevation of the water in a lake might not be correct in the TanDEM-X data, the elevation of the shoreline of the lake can be assumed to be correct, within the relative accuracy of the overall product. This important distinction between incoherent water and coherent

shorelines shall become clearer later, when this information can be used to edit the elevations of water in the WorldDEM data.

During SAR processing, surfaces other than water can be subject to the same signal reflection challenges as water. These include large areas of sand (deserts, beaches) dry and salt lakes, continuous areas of flat and smooth rock and finally ice and snow (fields and or ice-sheets). In the processed TanDEM-X data, many of these surfaces can exhibit some of the characteristics of water, such as extreme noise and/or voids in the resulting DSM, due to radar signal reflection. As noted above, the TanDEM-X mission planners worked hard to overcome the limitations posed by these smooth surfaces by customizing the acquisition process, where these features are widespread. However, the fact remains that some of these smooth features will not be accurately represented in the resulting TanDEM-X surface model.

A third type of elevation error occurs when the SAR processor has sufficient data to calculate the elevation over an area, but the uncertainty of the interferometric principle means that the processor generates a consistent surface, which does not have a correct elevation. This problem can occur when the processor is unable to accurately calculate the number of height offsets from the reference surface. This can lead to areas being represented as either too high, or too low in the resulting DSM. Islands, particularly in lakes and rivers, are most susceptible to this problem.

Very steep areas can also provide a challenge to the TanDEM-X SAR processor. Steep slopes that face directly into the radar during acquisition are generally well-portrayed in the DSM. However, steep slopes that faced away from the acquisition (and were not acquired from a different angle in a further acquisition) may be poorly represented in the DSM. This occurs as the SAR processor is unable to calculate an accurate surface, because the elevations change so rapidly in a short distance. It is possible for some of these steep areas to be completely void in the resulting DSM.

A final challenge that occurs in the TanDEM-X DSM are voids. During the course of more than three years of acquisitions, care was taken to acquire every square kilometer of the earth's surface. The reality is that due to any number of operational factors, some areas of the earth's surface were never acquired. These areas are portrayed as void in the resulting DSM and set to the null value of -32,787m, so that they can be easily identified.

1.3 Examples of Radar Effects in TanDEM-X Data

Shown below are some examples of X-band radar effects, that are apparent in the TanDEM-X data before editing is applied. In all images, the vertical exaggeration of the shaded relief images is 1x.
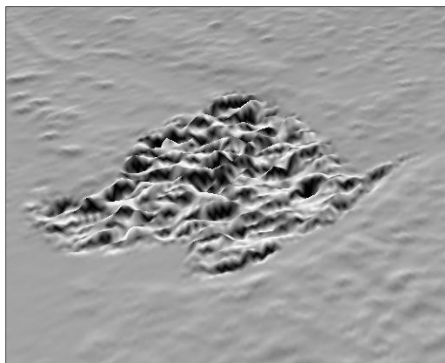

Figure 1: Characteristic roughness of the unedited WorldDEM DSM over a lake
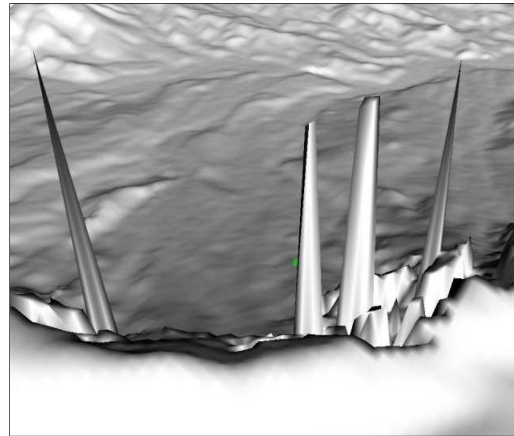

Figure 2: Elevation spikes on a steep mountain slope

1.4 Bespoke Production Software for the Production of WorldDEM Products

Airbus Defence and Space Geo-intelligence division provides WorldDEM data to clients in three forms: WorldDEM$_{core}$, WorldDEM and WorldDEM DTM. WorldDEM$_{core}$ is the unedited version of the WorldDEM. WorldDEM DTM is subject to additional processing, which removes surface features such as buildings and vegetation, to leave the elevations of the terrain in the model: however, these two additional products are beyond the scope of this paper.

The WorldDEM product is subject to terrain correction and hydrological editing, to provide clients with the most reasonable compromise of data quality, availability and price. In order to produce WorldDEM over large areas of the globe, a number of terrain and hydrological editing processes were developed by Airbus DS. In the initial stages, these processes were developed using commercial mapping and GIS software. However, it was always envisioned that the data volumes to be produced would exceed the reasonable capacities of commercial software. So Airbus DS Geo entered into a cooperation with a company that has significant experience with three-dimensional relief portrayal and data manipulation, in order to produce a software package to efficiently edit large volumes of WorldDEM data.

The resulting software package, called *Demes*, is designed to automate as much of the terrain and hydrological editing as possible, while still leaving the flexibility of human interaction with the data, when the automatic processes to not provide the intended results. The Demes editing software works on a standard PC workstation, and allows production staff to view, interact with and edit data directly. Demes server software manages the data workflow process: ensuring that data flows efficiently and consistently through the production stream, from initial ingest, through to final Quality Control (QC) and client delivery.

1.5 Editing Rules and Editing Guidance

In order to ensure consistency across the full edited WorldDEM dataset, a series of editing rules has been developed. These rules guide operators in their daily work to ensure that the correct terrain and hydrological features are being edited. The rules contain information about the identification and size of features to be edited and what software tools are acceptable for editing individual features.

Because the earth's surface is extremely variable from equator to the poles, further guidance is published to ensure that all operators are aware of special or single-instance surface features. For example, a mangrove area consists of significant

vegetation that stands in a continuously flooded part of the ocean shoreline. This could cause confusion among different operators, with one classifying mangroves as ocean, while another classifies them as land. When such uncertain features are found in the data, a decision is made on how to treat them. The decision and detailed technical guidance is posted on an internal Wiki, so that all operators have access to the same information.

2. TERRAIN EDITING

2.1 Spike and Well Removal

Spikes and wells are among the simplest of artefacts to identify and remove from the WorldDEM dataset. These are single pixel elevation errors, where the elevation of a point on the surface model is significantly different from the elevations of its neighbors. A spike is a positive elevation difference, while a well is a negative elevation difference to neighboring pixel. These spikes and wells are identified using a threshold difference elevation and are removed using a simple interpolation using the eight neighboring pixel elevations. The Demes software identifies and corrects spikes and wells in a single step. A QC tool is available to identify any spikes and wells that might have been missed by the tool, especially in steep terrain, or along Geocell edges. These spikes or wells may require individual manual editing to be completely removed.

2.2 Identification and Infilling of Voids

Another simple terrain artefact to identify and remove is areas that are void in the DSM. These voids are areas where either the data was not acquired, or the SAR processor was unable to calculate a valid elevation. Voids are identified in the data as elevations with the null data value of -32,767m.

2.2.1 Small voids
The production strategy with void data is to interpolate small voids, while larger voids are infilled (wherever possible) with an alternate elevation data source. The definition of a small void, is any void that is between one and sixteen contiguous pixels of void data. Any void larger than sixteen contiguous pixels is considered to be a large void. All calculations of void sized consider the eight-way connection of pixels (edge-adjacent and corner-adjacent).

Small voids are edited in Demes using *the Inverse Distance Weighted* (IDW) method of interpolation. IDW is based on the assumption that things that are close to one another are more alike than those that are farther apart from each other (ESRI 2007). Therefore any pixel that does not have a measurement should be interpolated (and its height calculated during interpolation) in reference to closer known points than to known points that are further away. The IDW process uses a standard weighting based on the distance to known points, providing a more plausible interpolated height than a simple linear interpolation would.

2.2.2 Large voids
The treatment of larger voids is somewhat different. While it is also possible to interpolate larger voids, their larger extent means that the likelihood of the interpolation providing an accurate portrayal of the earth's surface diminishes greatly as the void area becomes wider or longer. In such cases, the void area is infilled with an alternate DSM data source, such as SRTM or ASTER GDEM. While these alternate datasets might not have the same accuracy or temporal information as the surrounding WorldDEM data, it is understood that commercial

clients generally prefer to have some DSM information, rather than none, when WorldDEM is unable to provide any height information. For the benefit of WorldDEM users, areas of the DSM that are interpolated or infilled with another DSM source are identified in data masks that are delivered along-side of the WorldDEM elevation dataset.

For the filling of larger voids, the Delta Surface Fill (DSF) method of void filling is applied. First developed by the National Geo-Spatial Intelligence Agency (NGA) to infill voids in SRTM data, the Delta Surface method of infilling has the advantage of using the most accurate WorldDEM data at the edge of a void to seed the elevations of the less accurate fill source (Grohman 2006). Therefore a fill source that might have an overall acceptable portrayal of a surface feature, but might not be as accurate as WorldDEM data, will be manipulated to conform to the more accurate WorldDEM data at the edges of the void.

The Demes software allows the operator to use the DSF method of filling, either on a single void, multiple voids, or even every large void in a single Geocell. This flexibility allows the operator to efficiently fill a number of similar voids at the same time, while maintaining the flexibility to manually handle a single more complex void, which might require perhaps a different filling source than the other voids in the Geocell.

2.3 Identification and Correction of Systematic and Random Errors

While voids are generally easy to identify and treat in the WorldDEM data, there are two further categories of surface elevation artefacts that require additional editing to create a consumer-ready product: random and systematic errors.

2.3.1 Identification of systematic and random errors
The DLR provides additional data masks as part of the TanDEM-X data package. These masks are generated during the SAR processing and subsequent mosaicking of the data into a final DSM product. These masks are central to the identification of random and/or systematic errors in the DSM data.

The Height Error Mask (HEM) represents the height error of any pixel in the form of a standard deviation and is derived during the SAR processing from the interferometric coherence of every pixel measurement in the dataset.

The information derived from the HEM is used to identify areas in the DSM that are likely invalid, because the SAR processor only had weak or limited information about those areas. In addition, the height information contained in the pixels with poor values in the HEM are likely somewhat inconsistent with neighboring pixels (often visible as a certain "rough" texture in the DSM). The assumption is that in any DSM, neighboring elevation measurements should be somewhat consistent with each other. This combination of poor HEM values and a low consistency with neighboring pixels, provides a good indication of pixels/areas that likely have a high likelihood of random errors. Surfaces that are frequently subject to high random error include tropical forests and to a lesser extent, highly built-up urban areas.

A second mask provided by the DLR, the Consistency Mask (COM) indicates DSM pixels that have inconsistencies between the different acquisitions (known as a *Phase Unwrapping* error or a *Cycle Shift*). As noted previously, all areas of the earth's surface were collected at least twice during the TanDEM-X mission. A limited instance of the data collected have different elevations processed from the two (or more).
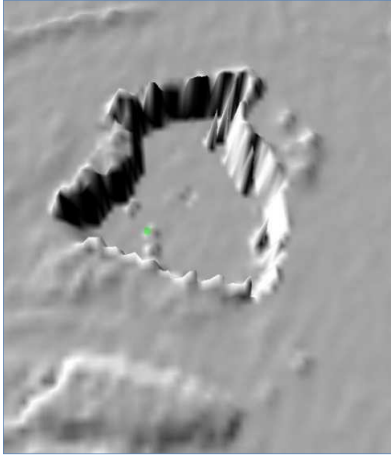
Figure 3: Example of a Phase Unwrapping error caused by the harvesting of trees in a forest between TanDEM-X acquisitions

Pixels that are identified as having two (or even more) elevations between acquisitions are identified in the COM. In the TanDEM-X SAR processing, only a single final height can be derived. The processor uses a number of tests to determine the best-fitting elevation, (such as number of measurements at the same height, pixel coherence from each acquisition, etc…) if there is a conflict between acquisitions. Whatever final height is derived, the fact that there was an inconsistency between acquisitions is noted in the COM mask. This allows the Demes software to identify areas with potential elevation inconsistencies and repair them during the terrain editing step.

2.4 Creation of the Reliability Mask

The potential DSM errors identified in the HEM and COM masks are somewhat different in their source, but in the end they both need to be treated during the terrain editing process. In order to make the terrain editing process as efficient as possible, the information from the HEM and COM masks is combined into a single mask, internal to the Demes software, which clearly identifies areas that require attention and/or repair. This mask is more than a simple combination of the HEM and COM data. Rather, it uses probability values from the HEM and the DSM in an attempt to determine what is actually an error in the DSM and what is likely legitimate noise in the DSM, perhaps caused by the legitimate roughness of a surface, such as the top of trees in a forest.

The amalgamation of COM, HEM and DSM error information results in the Reliability Mask (RLM). This is an internal Demes mask, which indicates the likelihood that each pixel in a DSM Geocell is reliable or not reliable. It further separates the unreliable pixels into ones that are likely unreliable because of random uncertainty in the DSM (noise) and pixels that are subject to offsets in the DSM due to differences between heights during acquisitions. The RLM does not discriminate between the size of erroneous data patches. Even areas of a single pixel (that is considered erroneous) are flagged in the RLM.
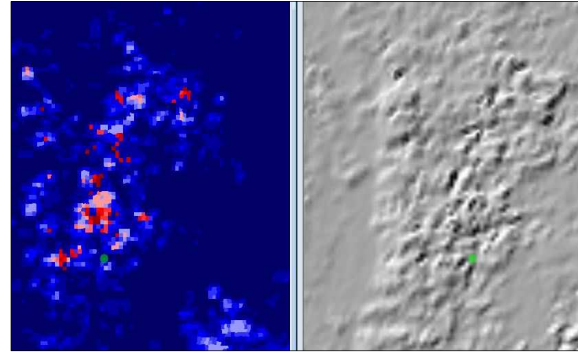

Figure 4: The RLM (left) indicating an area of forest where elevations are somewhat less reliable

2.5 Editing Terrain Using Demes

Before the terrain editing of artefacts identified in the RLM can begin, one more mask transformation is required. Because the RLM identifies each and every pixel that might have a random or systematic error, a further simplification is required. Generally a single pixel that displays the low to medium likelihood of error is not a major concern. Similarly, two or three pixels next to each other, showing only a slight likelihood of an error need not be edited. The goal of this simplification is to identify larger areas of concern, where higher error values are present. This is done by creating a new mask: the Feature Mask (FTM). In creating the FTM, smaller less error-prone pixels are removed, while larger error-prone areas have their edges regularized, in order to provide a more convenient edge for eventual Delta Surface Filling. In addition, areas that are likely to be cycle shifts, as identified in the COM, are given a special classification ("offset") in the FTM. This allows the operator to better understand the type of error that the FTM is identifying.

Because there is the possibility that some areas that are identified in the RLM as random errors, are in fact just the noise associated with the portrayal of water in the DSM, it is necessary to screen-out potential water areas from the resulting FTM. Water bodies will be identified, classified and edited in a separate step, outlined later. Editing water during the terrain editing step is not only inefficient, but there is a chance that infilling large areas of water with an external DSM source might adversely affect the shoreline elevations in the WorldDEM. Areas that are identified as possibly being water are flagged in the FTM as "water candidates". These areas will be ignored during the automatic terrain editing steps described next. The option does remain for an operator to manually override the water candidate class, if it has been incorrectly applied in the FTM.

In a final mask preparation step, the size of random errors is classified in the transition from RLM to FTM. This is because random errors of different extents will be treated differently during the editing of the DSM in Demes. Small artefacts are those between 2 and 16 neighboring pixels. Medium artefacts are those between 16 and 46 pixels. Any artefact 47 pixels and larger, is considered to be a Large artefact.
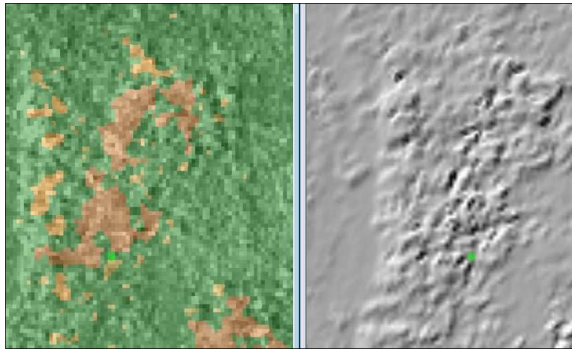
Figure 5: The FTM (left) demonstrates how the RLM is generalized to assist editing

During editing, all Small artefacts are voided and interpolated. All Medium artefacts are smoothed using a Gaussian smoothing approach. Large artefacts are voided and filled, using the DSF process, as described previously. Any area identified as an Offset in the FTM is also a candidate for filling using the DSF tool. However it is recommended that an operator visually identify Offset areas, to ensure that the filling solution is actually an improvement over the existing DSM data. There is the possibility that a portion of an area that is identified as an offset is actually accurate DSM data and infilling this area would only reduce the quality of the WorldDEM.

3. HYDROLOGICAL EDITING

In order to remove the characteristic noise that is present in many water bodies in the TanDEM-X data, a three-step process must be undertaken. These three steps are: the identification and delineation of all waterbodies, their division into the correct class of waterbody and finally, the application of a "flattening" algorithm to set the water body to a plausible elevation.

3.1 Identification and Delineation of Waterbodies

One advantage of editing water over terrain is that waterbodies are generally easier to identify because of the TanDEM-X ancillary image data sets (the amplitude image – AMP and the minimum amplitude image – AM2). Because of the lack of radar signal return, waterbodies have a distinctive black appearance, especially in the AM2 image. This consistency allows for a highly-automatic approach to the detection and delineation of water bodies.
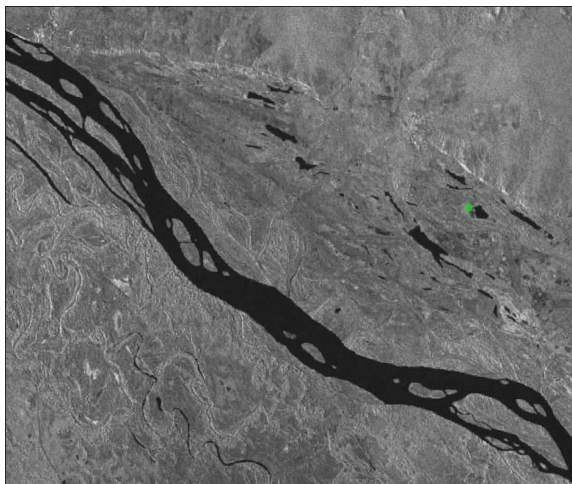

Figure 6: A river's course is easily identified using the AM2 amplitude image

The Demes software contains a number of tools to allow for the automatic, semi-automatic and manual delineation of waterbodies. The most important of these tools is called "Perception". The Perception tool is a type of trained classification tool: the operator finds a small area in a Geocell that is generally representative of the image tone for the full Geocell. The operator classifies this small subset area, by adjusting the sensitivity of the Perception tool to the image tone in the subset. When a plausible result is achieved in distinguishing land areas from water areas in the subset, the same Perception settings are applied to the whole Geocell, thereby identifying waterbodies throughout the whole Geocell simultaneously. Where a Geocell has areas that display widely divergent image tones, perhaps due to different data acquisition timings in the same Geocell, the Perception tool can be run on a subset of the Geocell only, to improve the local results of the waterbody identification process. After the Perception tool is used, a manual clean-up of some shorelines is usually required for some waterbodies. Demes provides the operator with a variety of drawing and selection tools to improve the delineation of shorelines.

3.2 Waterbody Classification

Three different types of waterbody class are used in the production of WorldDEM data. These three classes are applied once the operator has ensured that the shorelines of the automatically detected waterbodies are correct. The three classifications of waterbodies are: Ocean, Lake and River. The Ocean class is set for any saline waterbody whose geoidal reference elevation is 0m. The Lake classification is set for any fresh-waterbody that does not normally contain any flow. The River classification is set for drains that have a monotonic flow.

3.3 Setting the Elevation of Classified Waterbodies

Once classified, the elevations of the waterbodies are set in the Demes software. Oceans are universally set to 0m. In order to set the elevation of lakes, the elevation of all adjacent shoreline heights are measured. The elevation of a lake is fixed using the lowest 20% of the measured shoreline elevations. This calculation allows for the best elevation to be set where a lake is surrounded by surface features (such as a forest) as only the shoreline elevations that are most likely to be unobstructed are selected to set the lake elevation. River elevations are set in a similar manner as lakes, with the exception that rivers must flow downhill monotonically in 50cm steps from their point of first classification. Some operator interaction in the selection of river heights is required, as rivers generally encounter varying slopes throughout their courses.
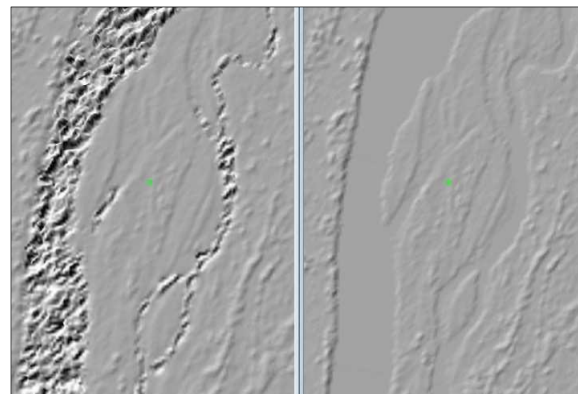

Figure 7: An unedited river (left) and the edited result (right)

In order to ensure the hydrological consistency of the WorldDEM dataset, a final automatic step is taken during the setting of waterbody elevations. Any shoreline pixel that would (for any reason) remain lower in elevation than the adjacent waterbody, is automatically raised 50cm above the surface of the waterbody. This ensures that all waterbodies are contained for hydrological modeling: water will not flow out of the edited hydrology in an unexpected manner.

4. QUALITY CONTROL PROCESSES FOR WORLDDEM PRODUCTION

Before the delivery of WorldDEM products to clients or the internal product library, QC checks are performed on every Geocell. These QC checks ensure that all WorldDEM data is produced in a consistent manner. The QC process is divided into two components: automatic tools that detect if each individual Geocell is consistently edited unto itself and its neighbors and a visual review of the edited data to ensure that the editing is consistent across the entirety of the WorldDEM dataset. Errors in the WorldDEM product are then marked interactively in the Demes software for review and correction.

A final QC step ensures that the complete WorldDEM package is correctly formatted and compiled and ready for delivery to customers or the internal product library.

5. CONCLUSION

The WorldDEM product, after terrain and hydrological editing, is a consumer-ready Digital Surface Model that is available on a global scale, with a consistent "look and feel" everywhere. WorldDEM follows-on from the great success of the original SRTM dataset by providing a higher resolution, greater vertical accuracy and surface information that is up to twelve years more current than SRTM.

REFERENCES

Bachmann, M. et. al. 2013. Acquisition Status and DEM Performance. *TerraSAR-X/TanDEM-X Science Team Meeting*, Oberpfaffenhofen, Germany. June 10 – 14.

Davies, H. 1954. The reflection of electromagnetic waves from a rough surface. *Proceedings of the IRE,* Volume 101, Issue 7, pp 209 – 214.

DLR. 2013. TanDEM-X Ground Segment: DEM Products Specification Document. Doc.: TD-GS-PS-0021, Issue 3. tandemx-science.dlr.de/pdfs/TD-GS-PS-0021_DEM-Product-Specification_v3.0.pdf

ESRI. 2007. How Inverse Distance Weighted (IDW) interpolation works. webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=How_Inverse_Distance_Weighted_%28IDW%29_interpolation_works

Farr, T. et. al. 2007. The Shuttle Radar Topography Mission. *Reviews of Geophysics*, Volume 45, Issue 2, pp 1 – 33

Grant, C., Yaplee, B., 1957. Back Scattering from Water and Land at Centimeter and Millimeter Wavelengths. *Proceedings of the IRE*, Volume 45, Issue 7, pp 976 – 982.

Grohman, G., Kroenung, G., Strebeck, J., 2006. Filling SRTM Voids: The Delta Surface Fill Method. *Photogrammetric Engineering & Remote Sensing,* Volume 72, Number 3, pp 213 – 216.

Slater, J., Garvey, G., 2006. The SRTM Data "Finishing" Process and Products. *Photogrammetric Engineering & Remote Sensing*, Volume 72, Number 3, pp 237 – 247.