GMES

TERRAFIRMA

ESRIN/Contract no. 19366/05/I-E

Validation of existing processing chains in TerraFirma stage 2

Product validation:

Validation in the Amsterdam and Alkmaar area

26th May 2008

Draft version 3

R.F. Hanssen, F.J. van Leijen, G.J. van Zwieten (TU Delft),

C. Bremmer, S. Dortland, M. Kleuskens (TNO)

INDEX

1 INTRODUCTION

The Product Validation of the "Validation of existing processing chains in Terrafirma stage 2" project (TerraFirma, 2007) is focused on the validation of the processing chains of the four Operational Service Providers (OSPs) of TerraFirma with respect to ground truth data. Two test sites are chosen in the Netherlands: (i) the region around Alkmaar, affected by the withdrawal of natural gas and (ii) an area in the city of Amsterdam, monitored extensively in relation to the planned construction of a metro tunnel, the so-called North/South-line (N/S-line).

These different driving mechanisms result in different expected deformation characteristics. The natural gas withdrawal in the Alkmaar region will result in a spatially correlated deformation field, leading to area-based comparison. The maximum deformation rate is expected to be in the order of a few millimeters per year. Displacements in the N/S-line region in Amsterdam are caused by geotechnical instability and localized construction work. These displacements can be spatially variable, leading to point-wise analyses.

The second cause of difference between the two study areas is the available ground truth. For the Alkmaar area sparsely distributed (in space and time) levelling data are available. The results in the N/S-line region in Amsterdam will be validated against 3D displacement measurements obtained by automatically operated tachymeters, forming a very dense spatial network and having a high temporal sampling. The current measurement frequency is about 1 measurement every 4 hours. This offers the possibility of comparing PSI-data closely co-registered in time and space with actual in-situ data.

Finally, the two areas have different characteristics in terms of their geography. The area influenced by subsidence near Alkmaar consists of a mixture of forest, dunes, beach, and small villages, whereas the Amsterdam city area is completely urbanized, leading to different characteristics in their radar reflectivity behavior.

These differences between the two data sets require a tuned product validation procedure, which is discussed in the following chapters together with the validation results. First, the characteristics of the validation test sites are discussed in more detail in chapter 2. The PSI results of the different teams are shown in chapter 3. Chapter 4 and 5 describe the validation procedure and results for the Amsterdam and Alkmaar site, respectively. A new independent quality parameter, the spatio-temporal consistency (STC), is described and applied in chapter 6. Finally, chapter 7 contains the conclusions.

Ground truth data for the Alkmaar test case were obtained from RWS-DID (Survey department of Ministry of Transport & Public Works). Data for the Amsterdam test case were obtained from the North/South-line project office. Both organizations are kindly thanked for their contribution in this project

2 BACKGROUND

2.1 Amsterdam

The N/S-line is a 9.5 km long metro line which is currently being built and runs through the city of Amsterdam, see Figure 2.1. About 3.8 km of this line will be constructed by a tunnel boring machine and will contain three stations (Rokin, Vijzelgracht and Ceintuurbaan) which will be constructed by cut-and-cover. The construction of the stations has started in 2003 while boring of the tunnel is expected to start in 2008. The sensitive conditions in Amsterdam place high demands on both settlement control and monitoring of structures which could potentially be affected by the works.

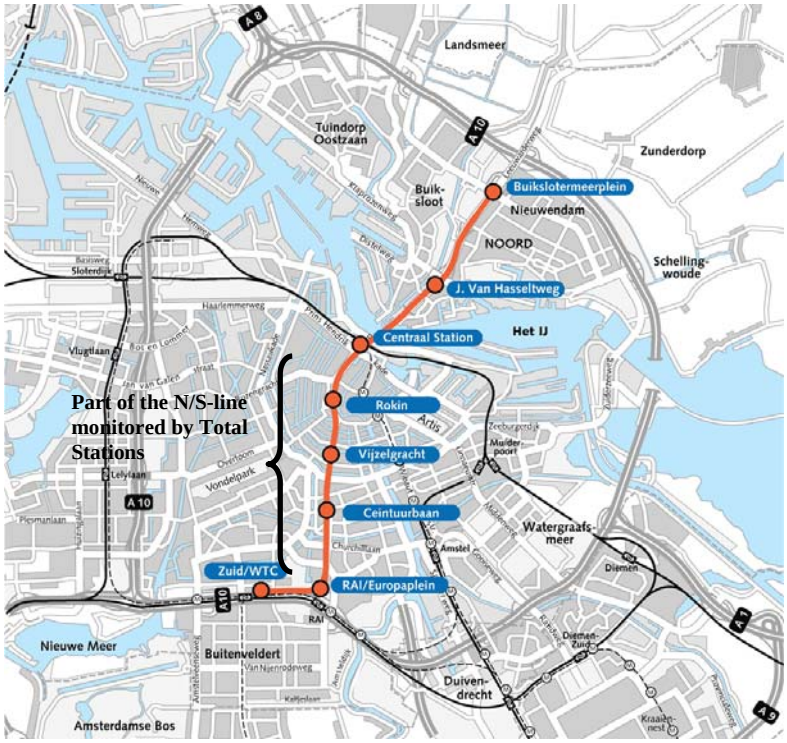

Figure 2.1: The trajectory of the planned North/South metro line.

The soft soil, high groundwater levels and historic nature of many of the buildings (17th – 19th century) make it of paramount importance to have a proper settlement-monitoring system installed which gives both timely and accurate information of settlement of individual constructions along the transect of tunnel boring. To this end an extensive monitoring system has been set up and installed in 2001 along the 3.8 km transect of the proposed bored tunnel. The monitoring system started 1 November 2001 and consists of three components:

1) a fully automated system consisting of 74 robotic tachymeters (total stations) aimed at 5350 prisms on 1500 constructions along the 3,8 km transect measuring individual prisms in (x,y,z).
2) Traditional precise levelling of reference objects along the tunnel transect in order to reference the local system of total stations.
3) Sub-surface monitoring of ground movement by a network of inclinometers, extensometers and piezometers.

The on-line monitoring system comprises about 74 Total Stations, computer controlled theodolites (Figure 2.2), which monitor some 5350 prisms in a continuous operation.

Figure 2.2: Example of a Total Station.

The prisms (about 4 per building) and the Total Stations are installed in the front and side façade walls of the buildings in the area of influence (Figure 2.3). The locations have been selected along the entire route so that each prism can be monitored by at least one Total Station (Figure 2.4). The prisms are measured within a range of 75 meters to obtain the required degree of accuracy. The system used, Cyclops of SolData, is reported to have an accuracy of +/- 0,5 mm over a distance of 60 meters. For the North/South metro line between 50 and 100 prisms are monitored by one Total Station. They are linked to a data logger which collects the data and transfers it via radio link to the central monitoring office.

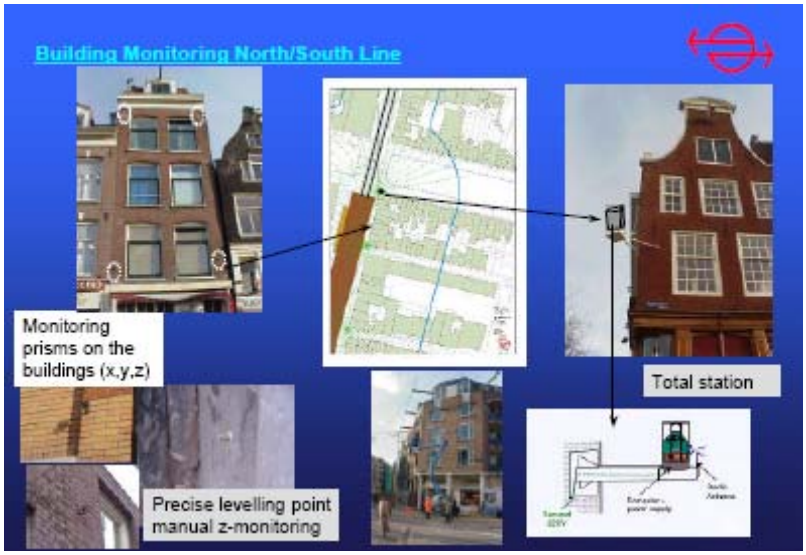

Figure 2.3: Monitoring system North/South metro line.

The Total Stations are split into groups of up to five instruments to form local geodetic networks (see Figure 2.5 as an example). In total, there are 23 networks of 2 to 5 total stations. The on-line monitoring system determines the x,y,z-deformations of the prisms. Apart from translation, tilting or rotating movements are therefore included in the observations. The current frequency is about 1 measurement every 4 hours. For each network of Total Stations it is important that stable reference targets are available. These are defined outside the zone of influence of the construction works.

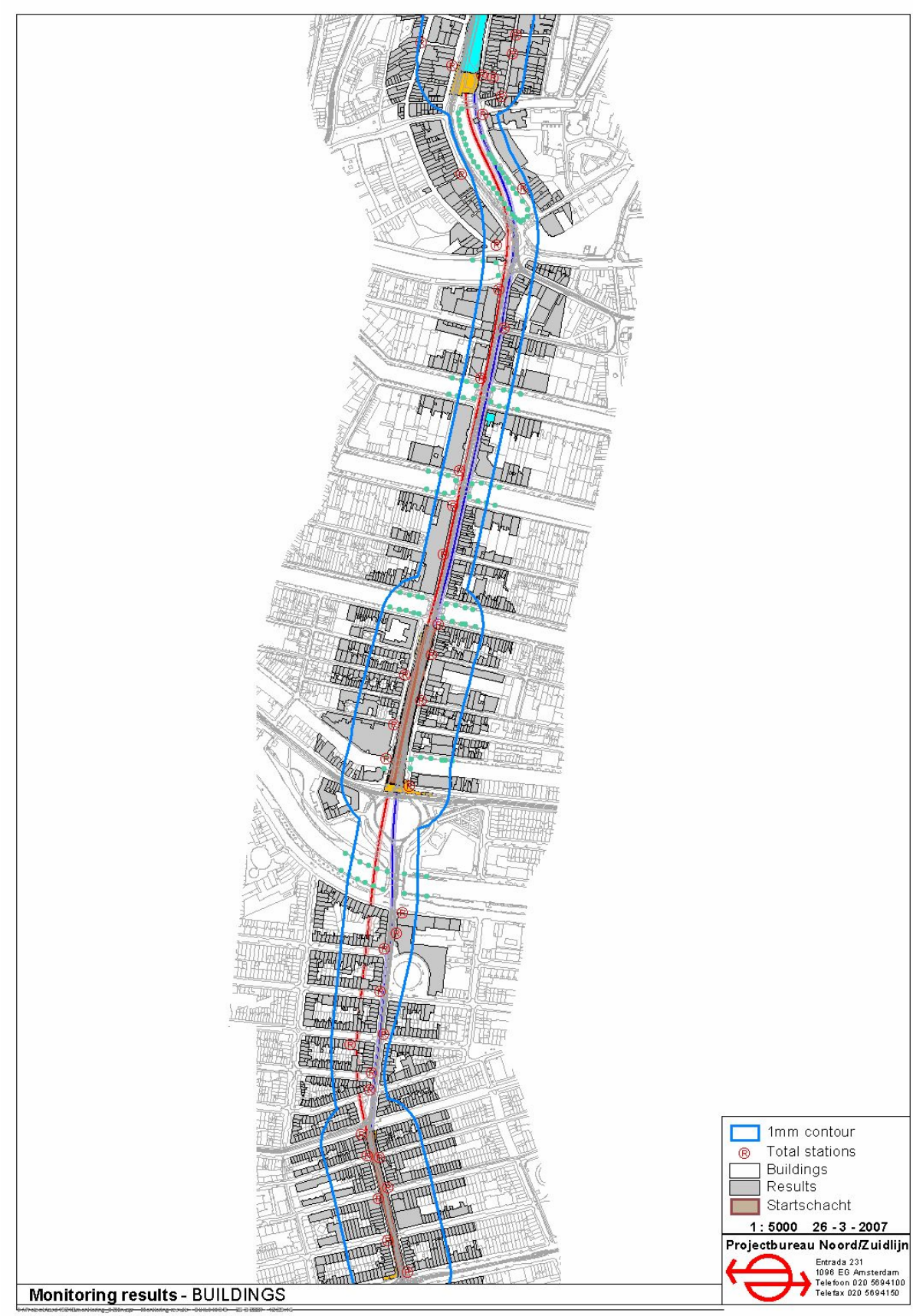

Figure 2.4: Set-up of part of the monitoring system along the route of the N/S-metro line. Highlighted buildings contain prisms. The blue contour indicates the region where the deformation is expected to be 1 mm or more. Outside this contour the expected total deformation is less than 1 mm.

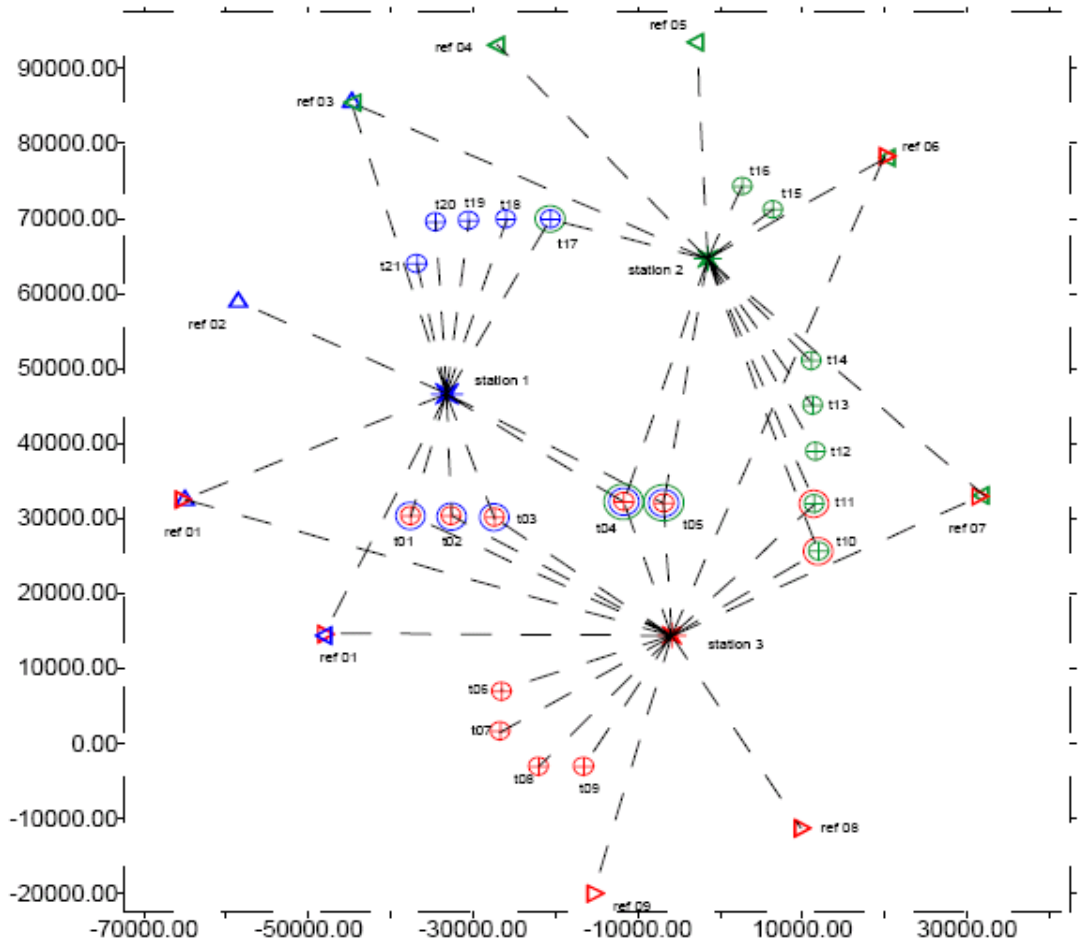

Figure 2.5: Example of local geodetic network of Total Stations (stars) ,prisms (crosses) and presumed stable reference points (triangles).

Monitoring using the Total Stations started in 2001[1]. The data are stored in a database/GIS-system which enables the selection of time series of x,y,z,-deformation for individual locations. The system and data are owned by the North/South-line project office of the city of Amsterdam.

First analysis shows that movements of the constructions in Amsterdam prior to the tunnel drilling show very little spatial correlation. Most of the movements are related to the combined effects of ground conditions and foundation characteristics. However, as a result of the local ground conditions the rates of deformation in the area of interest is about 1 to 3 mm/yr. However, as the monitoring results of the N/S-line showed, sudden movements have been recorded at various locations either due to e.g. foundation repair or to ground movement caused by the activities related to the construction of the stations. Figure 2.6 is an example of the output of the monitoring system.

[1] Note that the construction of the stations started in 2003 and the actual drilling of the tunnel is expected to start in 2008.

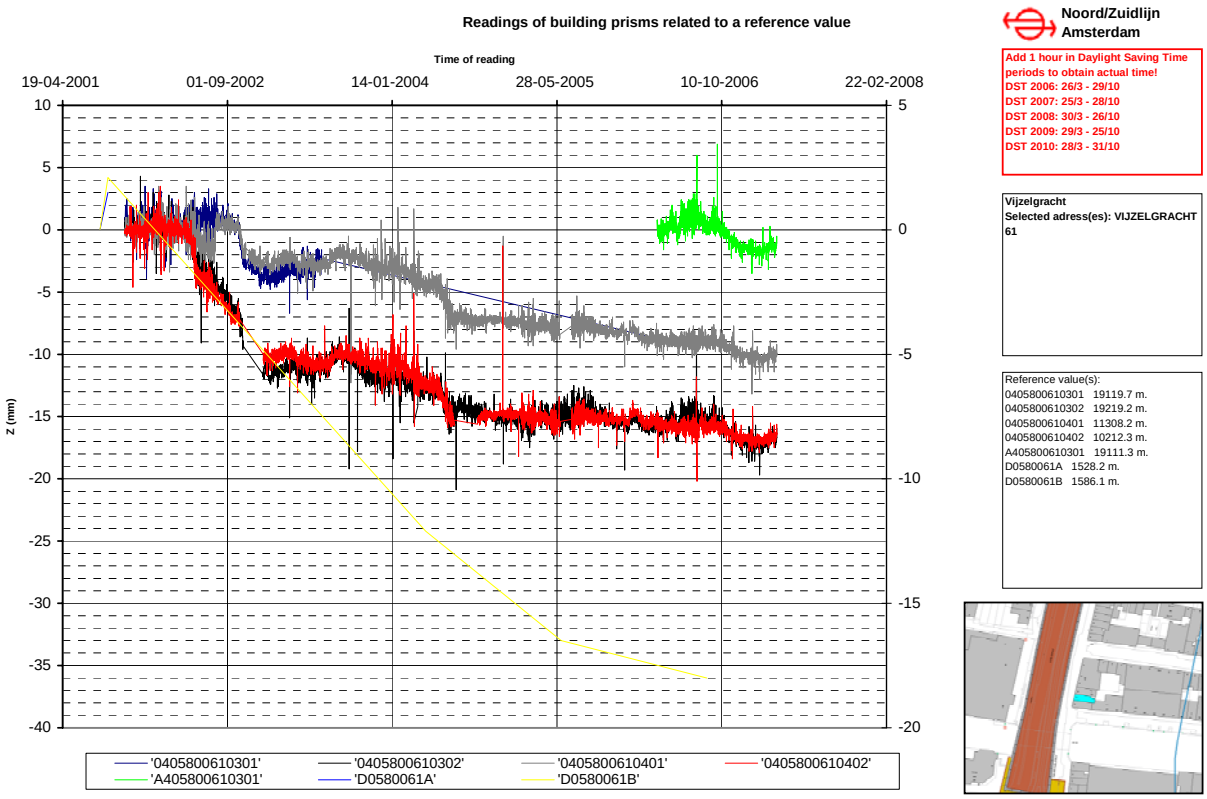

Figure 2.6: Example of vertical deformation observed by the monitoring system of the North/South metro line. The different colors reflect different prisms, located on one building

2.2 Alkmaar

The Alkmaar area, in the Province of Noord-Holland (Netherlands), is part of the validation experiment since it is an important on-shore gas-producing area of the Netherlands. The area comprises 16 gas fields of various sizes and elongated shape, oriented in a NW-SE direction along a stretch of about 30 km. The location of the gas fields is shown in Figure 2.7. Gas production started in the early 1970's and continues up to date for most of the fields. End of production is expected to be somewhere around 2010. The field known as 'Alkmaar' has already been closed for production and is now used for gas injection for peak gas buffering. Gas production is from a depth of over 2000m from reservoir rocks of the Rotliegend Slochteren Formation, Zechstein 3 Carbonate Member (Platten) and the Main Buntsandstein Subgroup (Bunter).

Figure 2.7*:* Location of the gas fields in the Alkmaar area in North-Holland.

Figure 2.8 shows a geological cross-section over the Alkmaar area.

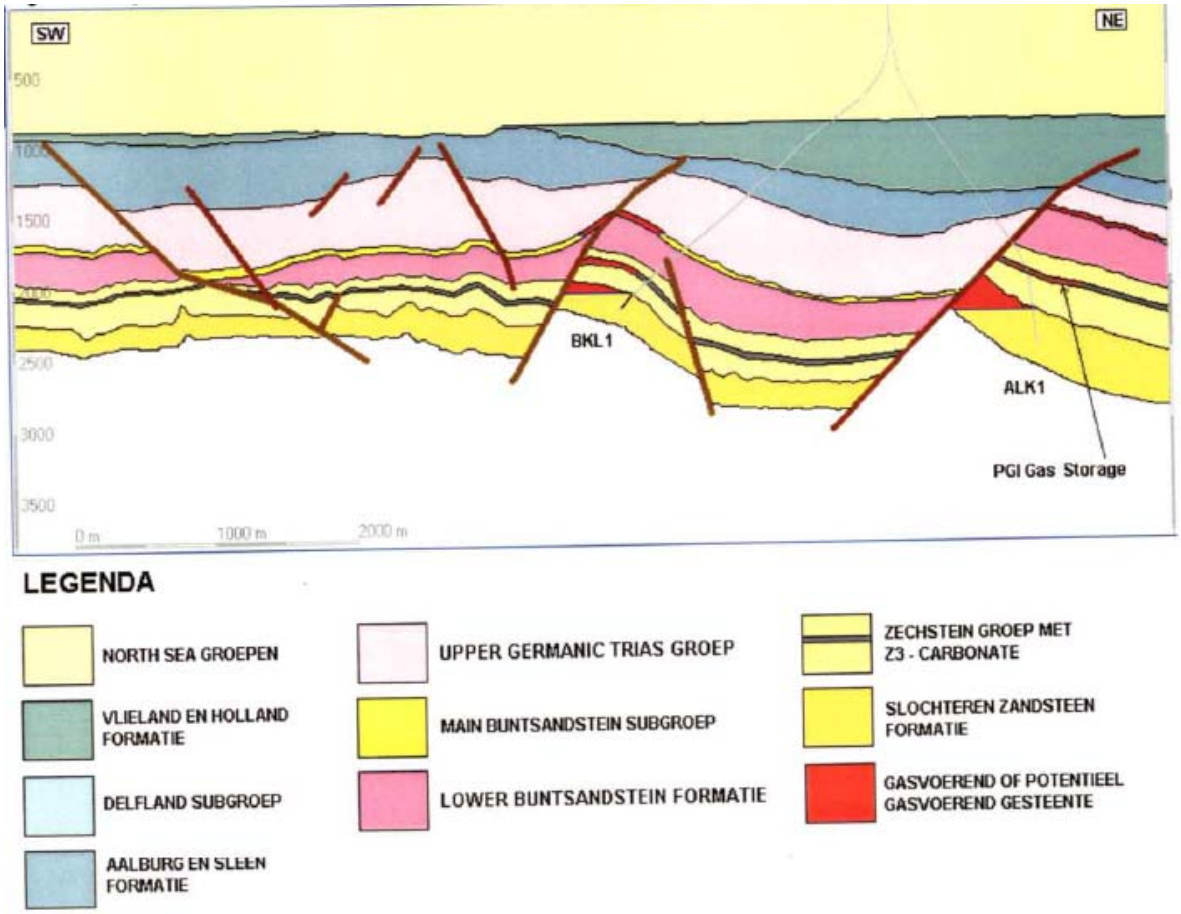

Figure 2.8: Geological cross-section over the Alkmaar area.

The expected maximum additional subsidence over the various gas fields for the total production period varies from less than 2 cm to an expected 8 cm maximum. Maximum average subsidence velocity with respect to natural movements over the period 1972 – 2003 is estimated to be of the order of 4 mm/yr.

Induced earthquakes occurred at least five times during the period of gas extraction:

1. 06-08-1994, magnitude (M_L) 3.0, 2.5 km depth
2. 21-09-1994, magnitude (M_L) 3.2, 2.5 km depth
3. 09-09-2001, magnitude (M_L) 3.2, 2 km depth
4. 10-09-2001, magnitude (M_L) 2.8, 2 km depth
5. 10-10-2001, magnitude (M_L) 2.7, 2.5 km depth

As a result of the 9-9-2001 quake, damage has been reported in an area of 100 km^2, see also Figure 2.9 and Figure 2.10.

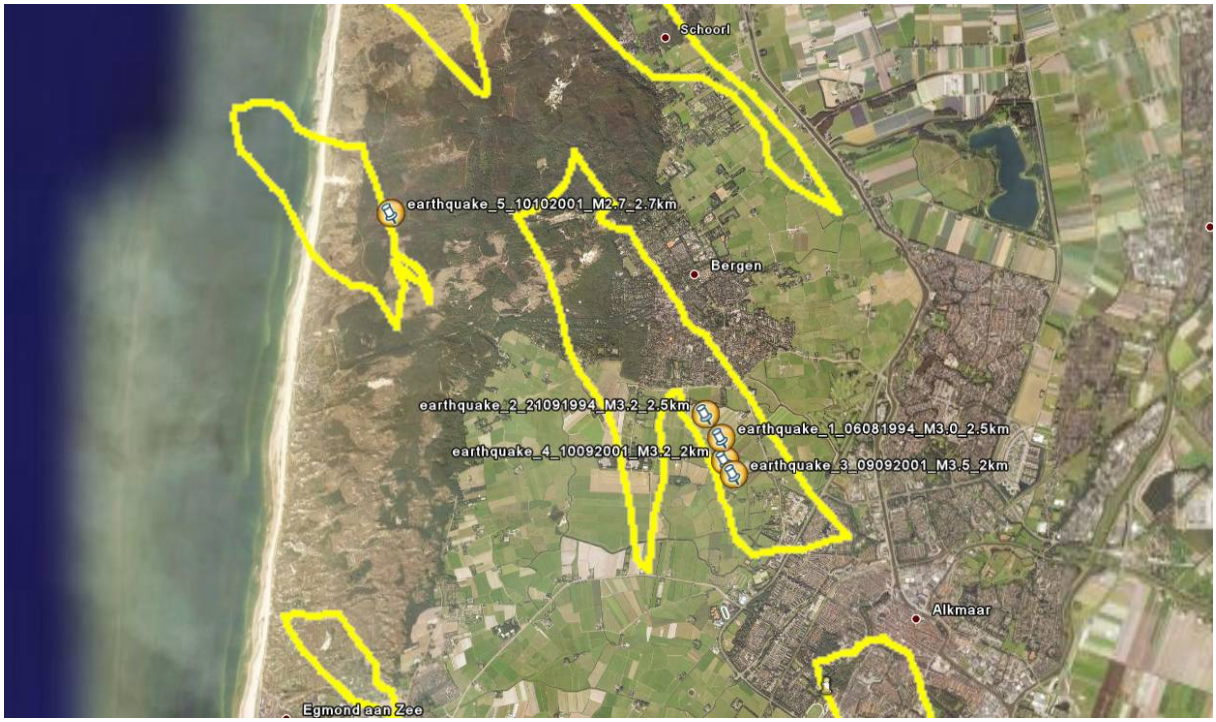

Figure 2.9: Earthquake epicenter positions in relation to gas reservoir boundaries.

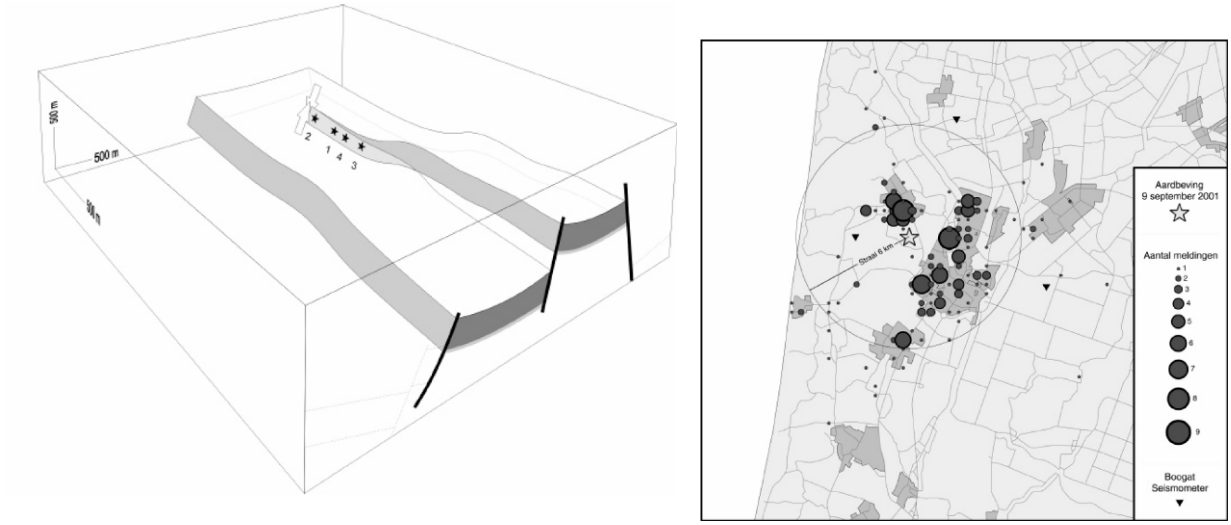

Figure 2.10: Left: Mechanism of the earthquakes of the Bergermeer gas field. Epicentres are indicated by the stars. The higher located gas bearing layers have moved downwards, causing the earthquakes. Right: overview of reported damage claims per location following the 9/10-09-2001 quakes (stars 3 and 4 on left image). Larger circles indicate more damage reports (max. 9) [Haak et al, 2001].

3 GENERAL ANALYSIS PSI DATA

The area around Alkmaar is analyzed using 83 ERS images (Apr. 1992 – Sep. 2000) and 39 Envisat images (Mar. 2003 – Mar. 2007). For Amsterdam only the Envisat data set is used. The number of detected Persistent Scatterers in each data set obtained by the four operational service providers (OSP) of TerraFirma is shown in Table 3.1 (Team A-D)[2]. The result of TU Delft is denoted by Team E and is used additionally throughout the validation process to improve the range of comparison. The table shows that the number of detected PS can be up to 7 times larger for one team compared to the other (18000 vs. 120000 in case of ERS-Alkmaar). It needs to be stressed that the sheer amount of PS is not necessarily indicative of the quality of the parameter estimation. The spatial distribution of the PS and optimized entanglement of the different phase contributors are equally, perhaps sometimes more important. In the following it is demonstrated that for some end-user applications the distribution of PS may be more important than the density.

Table 3.1: Number of Persistent Scatterers.

Number of PS	Team A	Team B	Team C	Team D	Team E
ASAR-Amsterdam	124318	91038	87911	74993	125739
ERS-Alkmaar	17859	121269	34103	25617	28421
ASAR-Alkmaar	63485	90209	54520	28363	88131

The analyses resulted in displacement time series, displacement velocities, relative heights and temporal coherence estimates of the PS. A general overview of the spatial distribution of the estimated displacement velocities is presented in Section 3.1. Section 3.2 shows the effect of de-trending of the data, whereas the estimated temporal coherences are presented in Section 3.3. For the estimated relative heights the reader is referred to Annex A.

3.1 Original linear displacement velocities

The spatial distribution of the original linear displacement velocities [mm/y] provided by the teams is shown in Figure 3.1 to Figure 3.3. For the Alkmaar region, the figures also contain the velocity estimates derived from levelling. The procedure used to obtain these velocities is described in Chapter 5. To enable comparison and further validation, the PSI derived velocities for the Alkmaar region are projected to the vertical direction using precise orbits. The Amsterdam results are in the original line-of-sight. The data are all referenced to a common reference point.

[2] It should be noted that during the project team C and D were requested to make a second delivery for the Amsterdam Envisat data set, with a larger number of detected PS, in order to increase the number of PS-NSline pairs for a more significant comparison. The first delivery of team C consisted of 33606 PS, the first delivery of team D of 21906 PS. In Table 3.1 the number of PS of the second delivery are given.

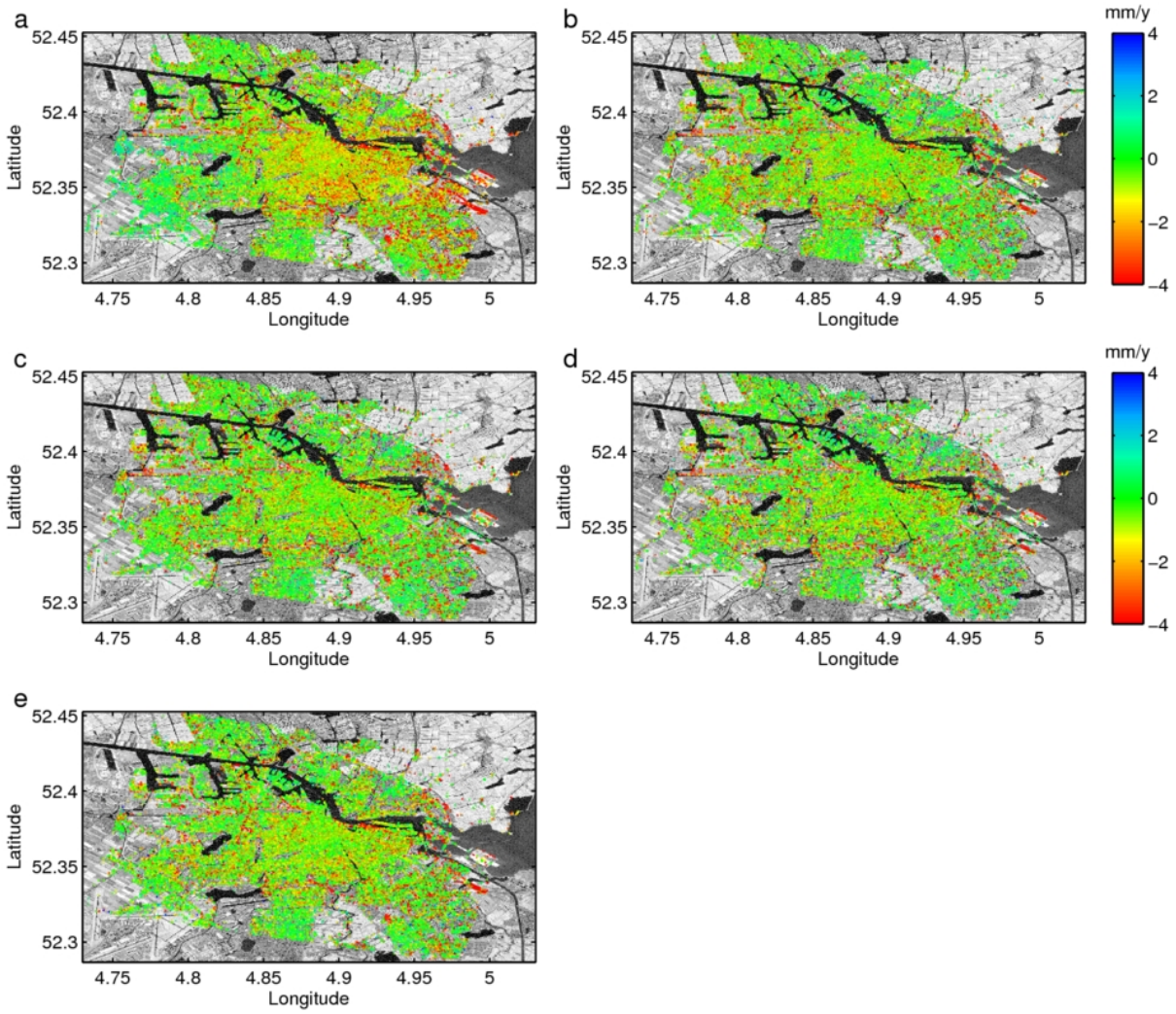

Figure 3.1: Amsterdam, Envisat. Linear displacement velocities (LOS) [mm/y] in Amsterdam obtained from Envisat time series (Mar. 2003 – Mar. 2007). a) - d) OSP results, e) TU Delft result. No corrections for spatial trends have been applied.

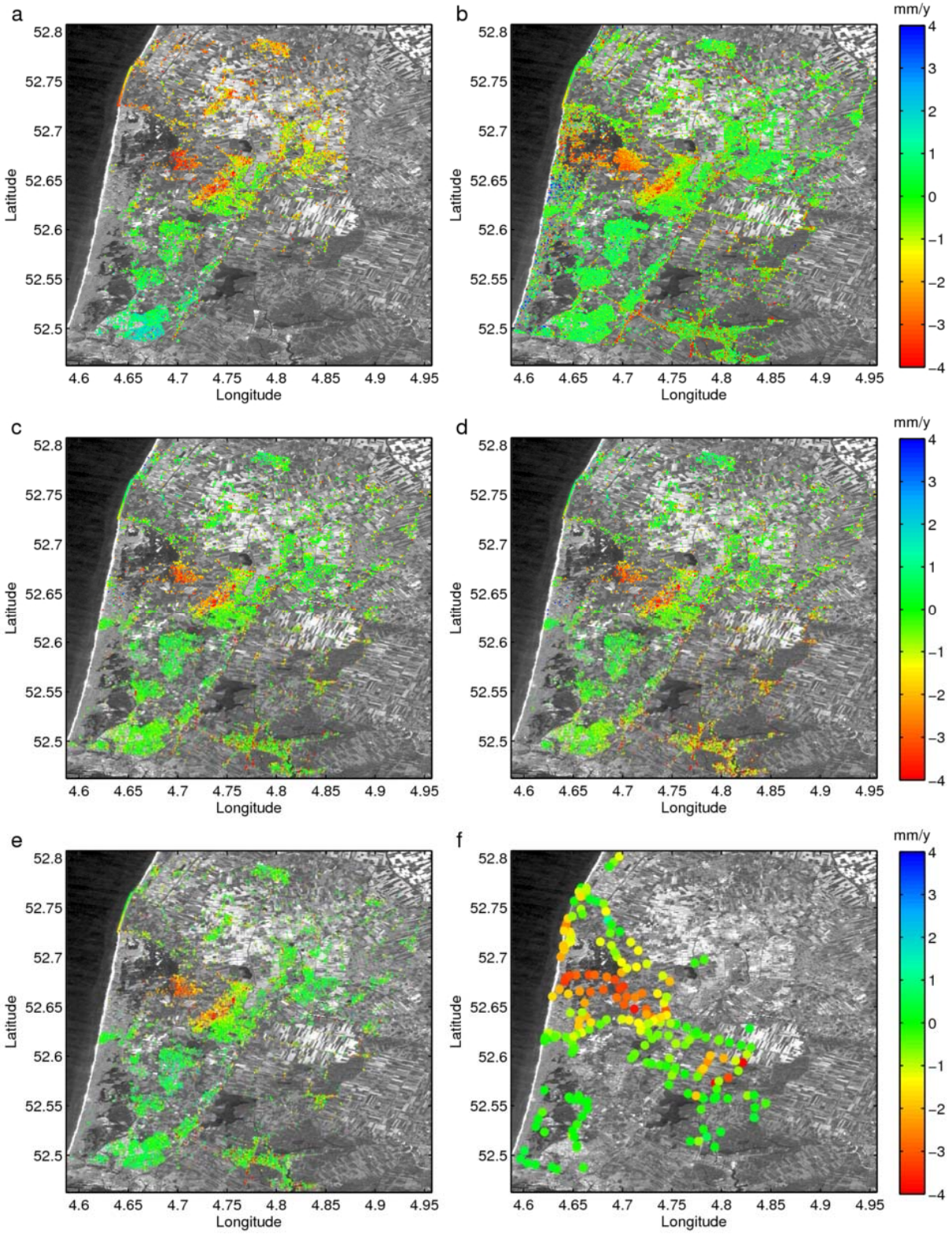

Figure 3.2: Alkmaar, ERS. Linear displacement velocities (vertical) [mm/y] around Alkmaar obtained from ERS time series (Apr. 1992 – Sep. 2000). a) - d) OSP results, e) TU Delft result, f) linear displacement velocities [mm/y] from levelling. No corrections for spatial trends have been applied.

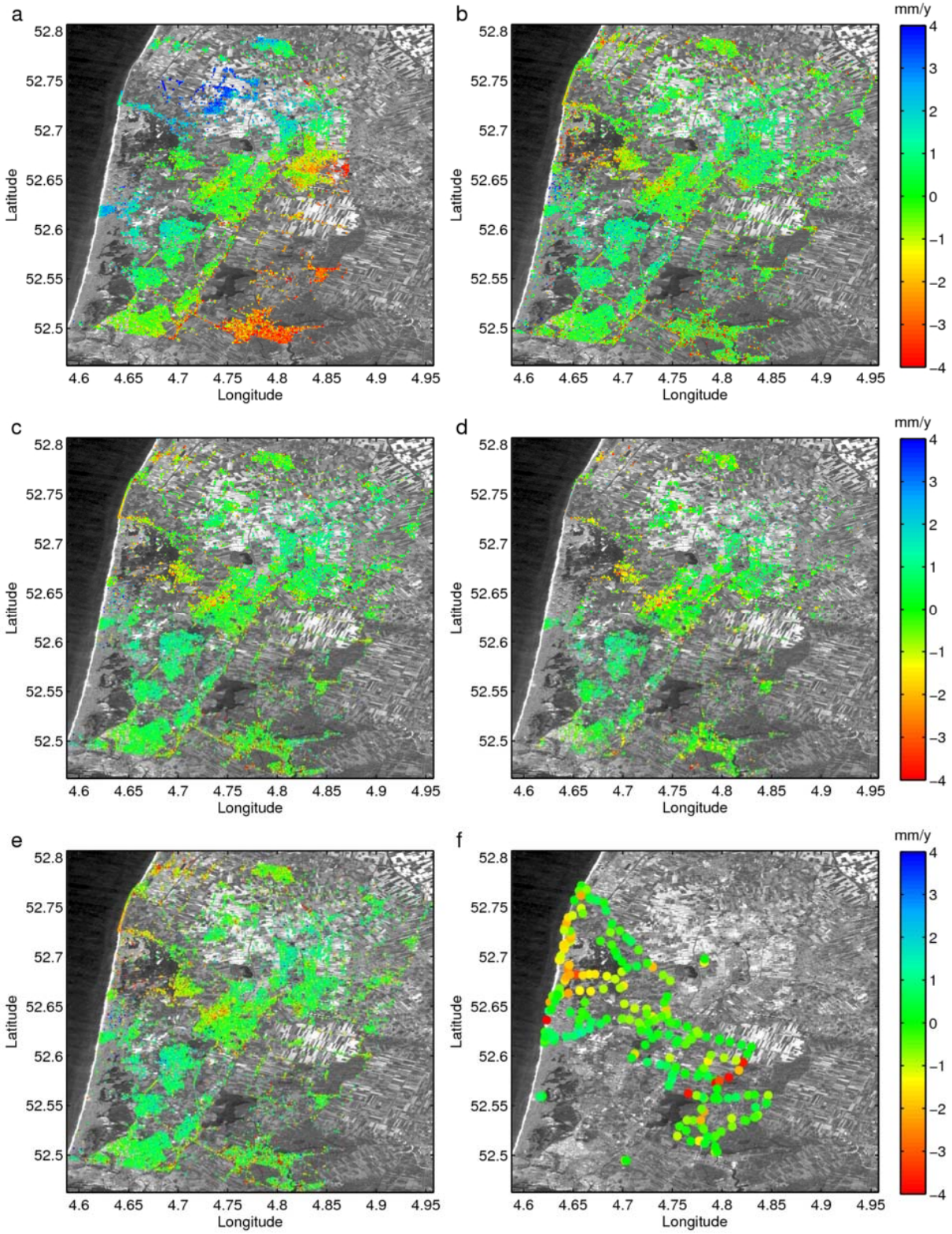

Figure 3.3: Alkmaar, Envisat. Linear displacement velocities (vertical) [mm/y] around Alkmaar obtained from Envisat time series (Mar. 2003 – Mar. 2007). a) - d) OSP results, e) TU Delft result, f) linear displacement velocities [mm/y] from levelling. No corrections for spatial trends have been applied.

3.2 De-trended linear displacement velocities

Although the results of the different teams in Figure 3.1 to Figure 3.3 show similar features, there are trends in the data that prevent a direct interpretation. Especially the results of team A contain a large trend. With current satellite state vector precisions, orbital trends are largely in the null-space of radar interferometric measurements and therefore not solvable without assumptions on the spatial characteristics (i.e. planar trend) or external measurements (e.g. GPS). To remove the trends, a linear plane is estimated through all data sets and removed. Per team all PS are used in the estimation process. The distribution of the linear deformation velocities are visualized in Figure 3.4 to Figure 3.6. After de-trending, the distribution of linear velocities of all teams is very similar.

Analysis of common PS between the teams within the process validation (Adam and Parizzi, 2008) identified an underestimation of the linear deformation velocities by a factor of 2 in the Alkmaar-Envisat data set of team D. Although this effect is not clearly visible in the histograms of all PS, the distribution of velocities is the narrowest, see Figure 3.6.

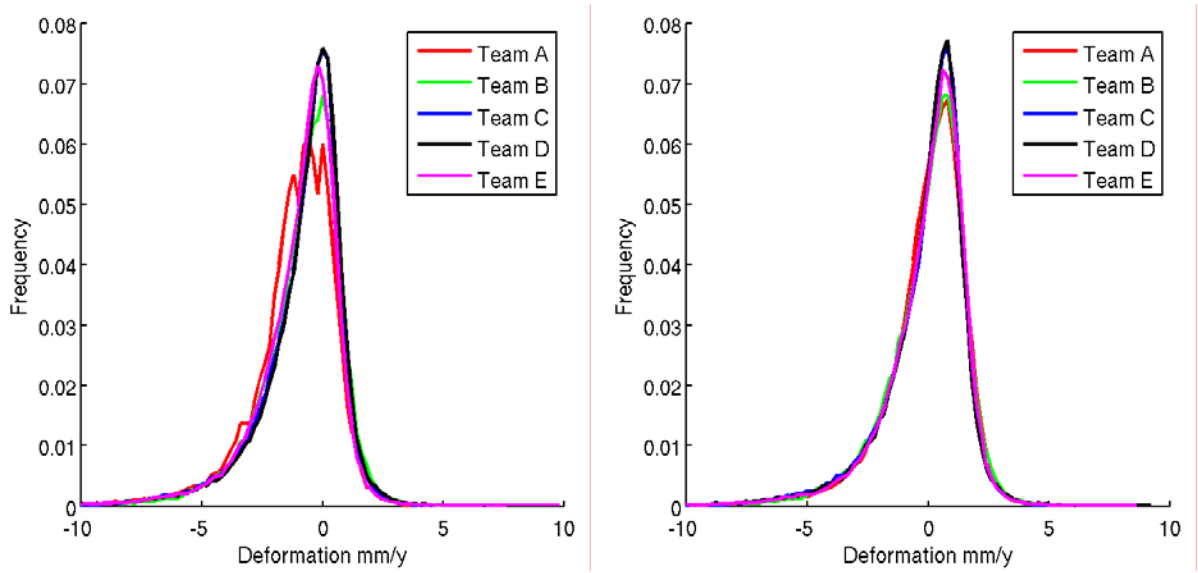

Figure 3.4: Amsterdam, Envisat. Normalized histograms of the estimated linear displacement velocities (LOS) in Amsterdam for the Envisat time series. Left) original data, right) de-trended data.

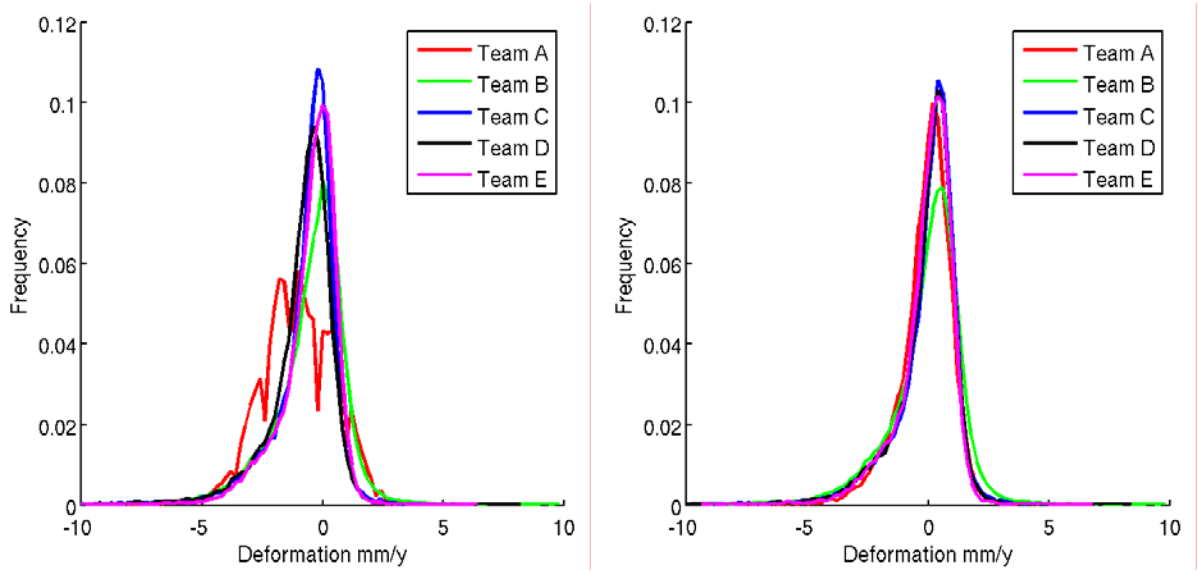

Figure 3.5: Alkmaar, ERS. Normalized histograms of the estimated linear displacement velocities (vertical) around Alkmaar for the ERS time series. Left) original data, right) de-trended data. The jagged nature of the red line is due to the discrete bin size and the relatively low amount of PS.

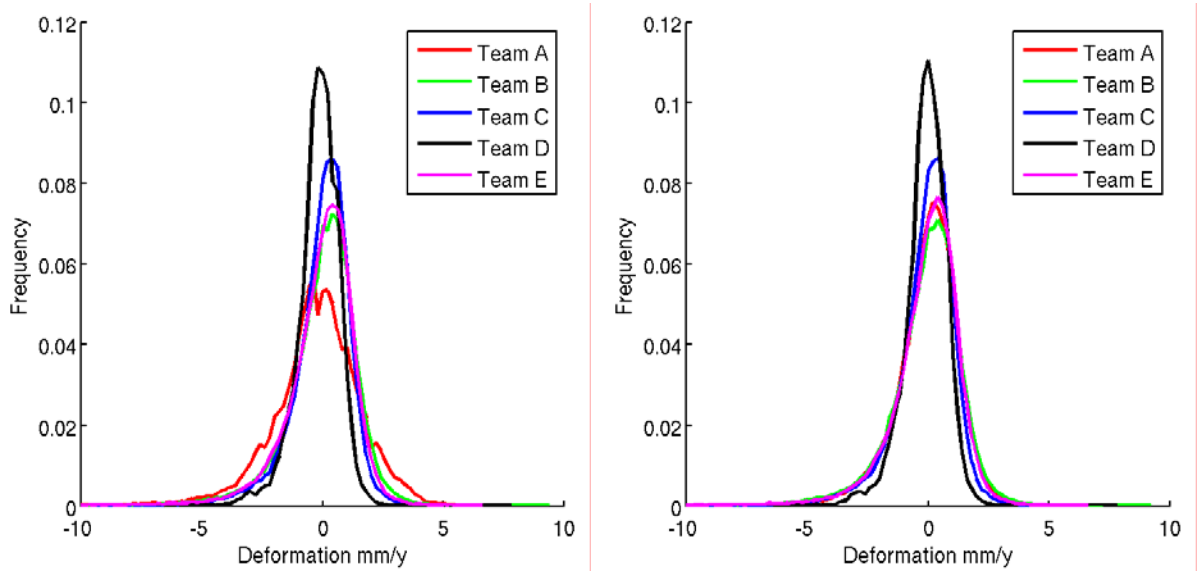

Figure 3.6: Alkmaar, Envisat. Normalized histograms of the estimated linear displacement velocities (vertical) around Alkmaar for the Envisat time series. Left) original data, right) de-trended data.

The spatial distribution of the de-trended linear displacement velocities is shown in Figure 3.7 to Figure 3.9. The results are more comparable and better interpretable. However, for team A there are still isolated patches which show an offset compared to the other teams. This effect is especially large in the Alkmaar-Envisat case. Integration errors during phase unwrapping are the most likely cause of this effect. A detailed analysis of this effect is reported in the process validation document (Adam and Parizzi, 2008).

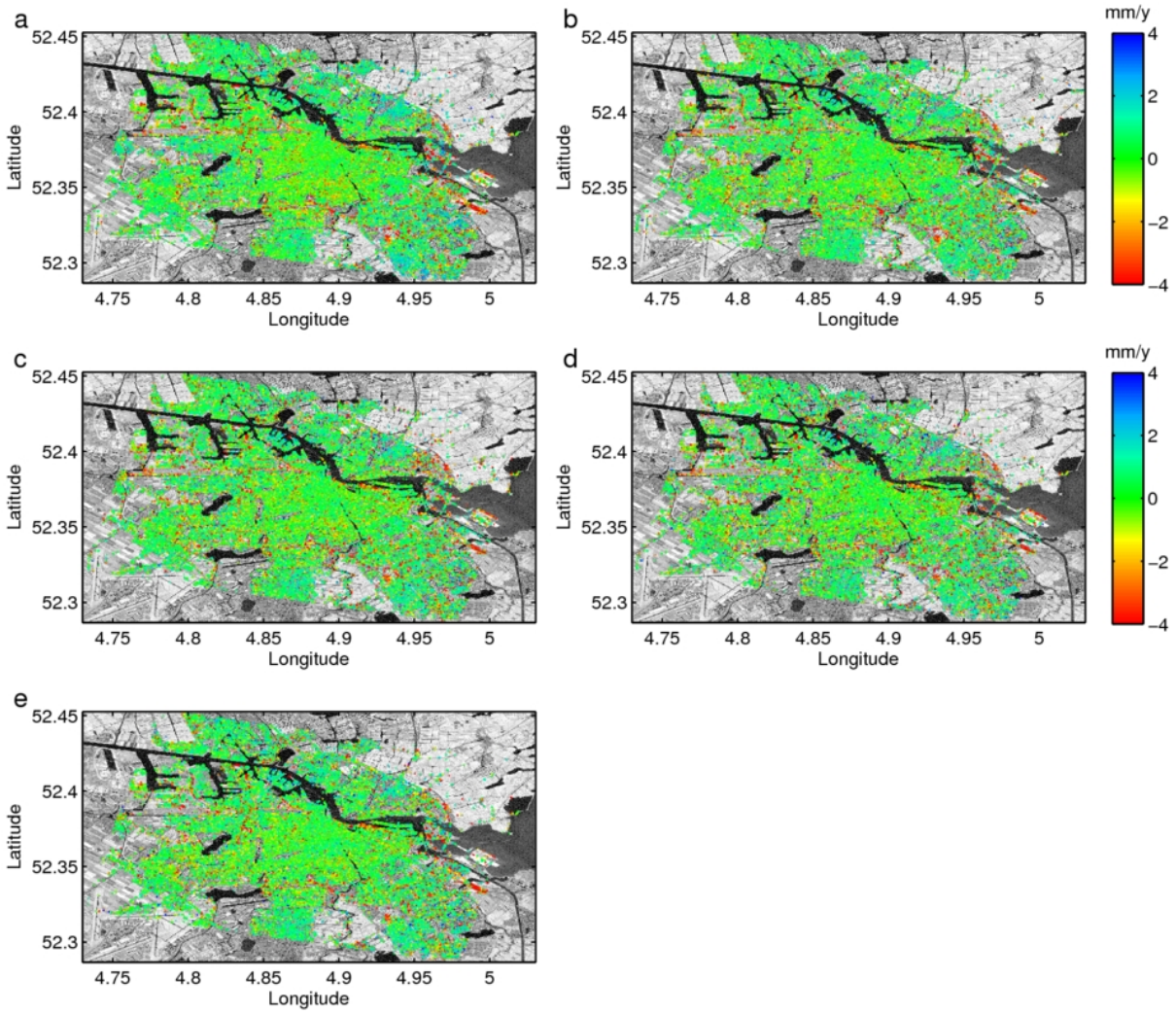

Figure 3.7: Amsterdam, Envisat. De-trended linear displacement velocities (LOS) [mm/y] in Amsterdam obtained from Envisat time series (Mar. 2003 – Mar. 2007). a) - d) OSP results, e) TU Delft result.

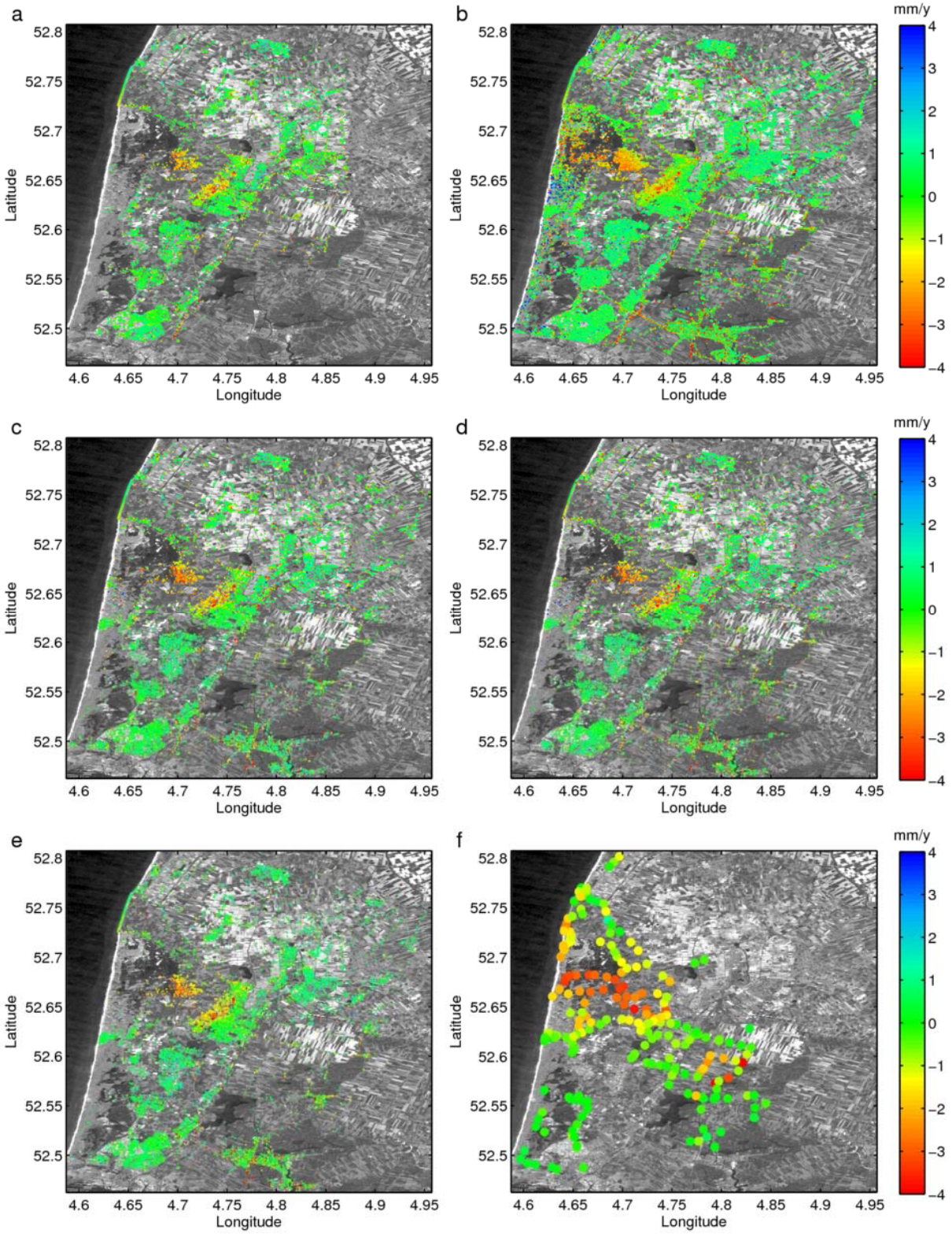

Figure 3.8: Alkmaar, ERS. De-trended linear displacement velocities (vertical) [mm/y] around Alkmaar obtained from ERS time series (Apr. 1992 – Sep. 2000). a) - d) OSP results, e) TU Delft result, f) linear displacement velocities [mm/y] from levelling.

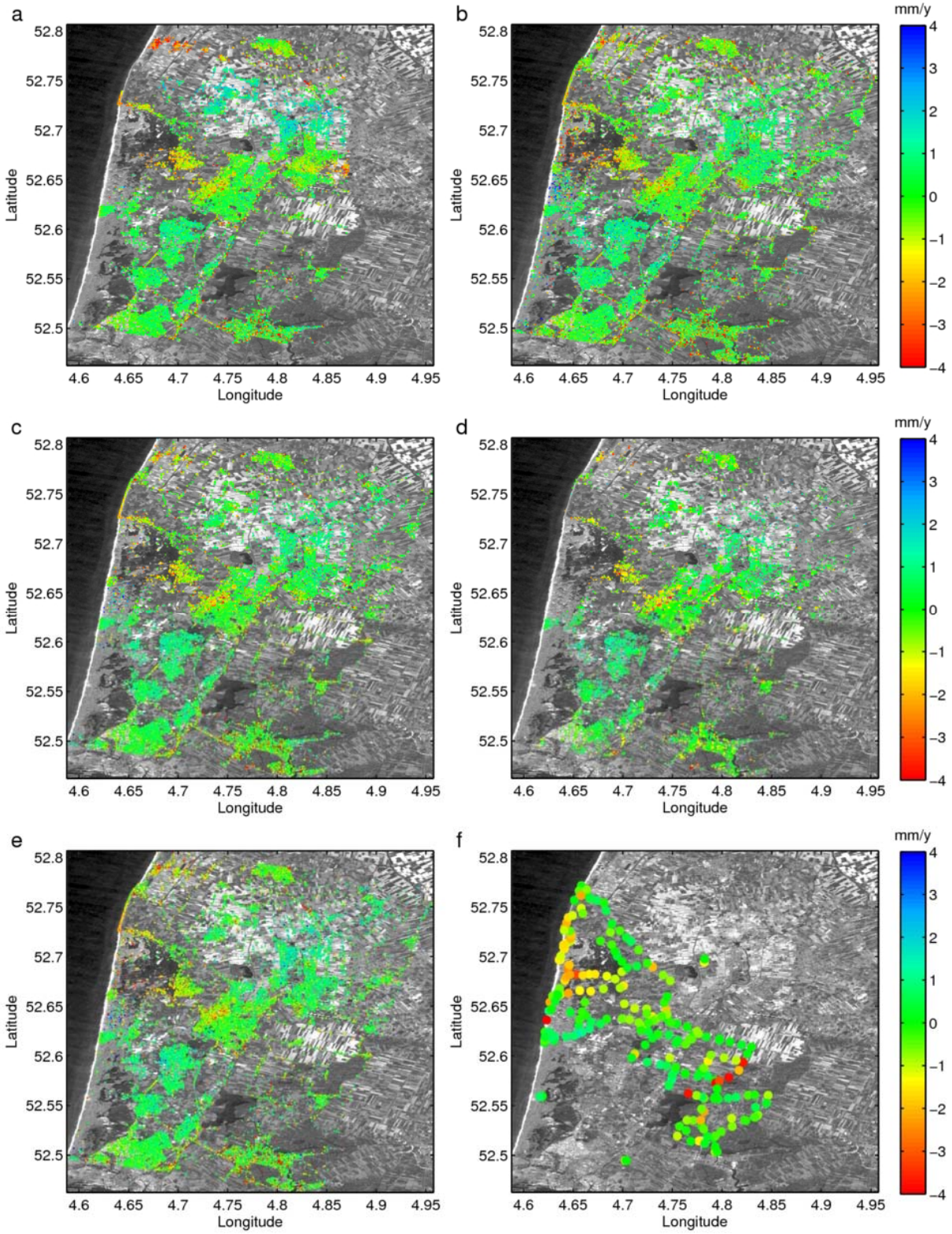

Figure 3.9: Alkmaar, Envisat. De-trended linear displacement velocities (vertical) [mm/y] around Alkmaar obtained from Envisat time series (Mar. 2003 – Mar. 2007). a) - d) OSP results, e) TU Delft result, f) linear displacement velocities [mm/y] from levelling.

3.3 Temporal coherence

A frequently used quality indicator is the temporal coherence. This parameter describes how well the interferometric phase observations fit to the temporal displacement model used (often a linear model). As an example, the spatial distribution of the estimated temporal coherence of the Alkmaar-ERS data sets is displayed in Figure 3.10. The Alkmaar-ASAR and Amsterdam-ASAR estimated temporal coherence plots can be found in Annex 1.

The histograms of estimated temporal coherence supplied by the teams for all areas are shown in the left column of Figure 3.11. Although the definition of the temporal coherence is clear, the estimates in the figures show large differences. These differences are either caused by conceptual differences in the processing chains, the location in the processing chain were the coherence is calculated (e.g., before or after filtering) or the amount of filtering. This makes interpretation of the estimated temporal coherence values hard, if not impossible. However, analysis in the process validation (Adam and Parizzi, 2008) shows that the ordering of the PS with respect to coherence value is equivalent. That is, PS with a high coherence have a high coherence for all teams.

The right column of Figure 3.11 shows the temporal coherence re-calculated from the time series. To obtain these values, the time series are translated from mm to phase, followed by a least-squares linear fit. No weight or covariance matrix is applied in this estimation. The temporal coherence γ is re-calculated by (Ferretti et al., 2001, see also Adam and Parizzi, 2007)

$$\gamma = \left| \frac{1}{N} \sum_{i=1}^{N} e^{j(\varphi^{\text{ifg}} - \varphi^{\text{model}})} \right|,$$

where φ^{ifg} is the interferometric phase of the time series, φ^{model} is the modeled phase and N is the number of interferograms. Note that the re-calculated values can differ slightly from the values supplied by the teams because a) an alternative estimation procedure compared to the least-squares principle is applied and/or b) a weight or covariance matrix describing the (relative) precision of the individual displacements in the time series is applied (as is for instance the case for the results of TU Delft (team E)). These differences in the estimation procedure cause small differences between the supplied and re-calculated coherence values, which is especially visible in the low end of the histograms in Figure 3.11 where a threshold is applied by the teams (e.g., compare the result of team B in figure 3.11a and d).

Apart from these small differences, especially the re-calculated coherence values of team A differ significantly from the supplied values, indicating additional a-posteriori smoothing of the time series. This is confirmed by the semi-variograms shown in Figure 3.12. The semi-variogram of team A shows a much smaller nugget than the rest of the teams, indicating a smooth signal at short time spans.

The dependency of the temporal coherence estimator on a displacement model is a disadvantage and does not allow for distinguishing between measurement noise and model imperfections. To overcome this dependency, a new quality indicator is proposed, the spatio-temporal consistency (STC), as described in Chapter 6.

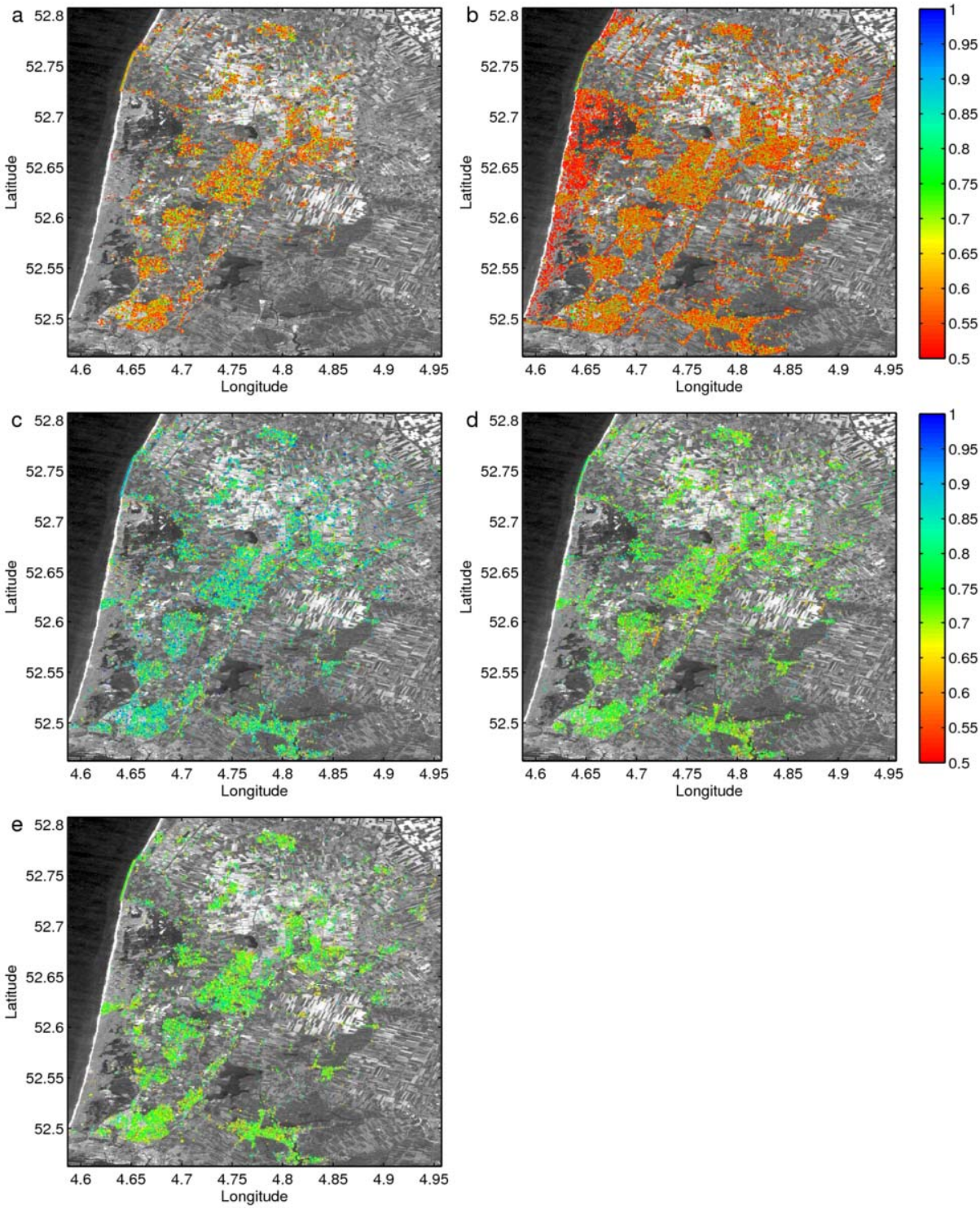

Figure 3.10: Alkmaar, ERS. Temporal coherence estimates around Alkmaar for ERS time series (Apr. 1992 – Sep. 2000). a) - d) OSP results, e) TU Delft result.

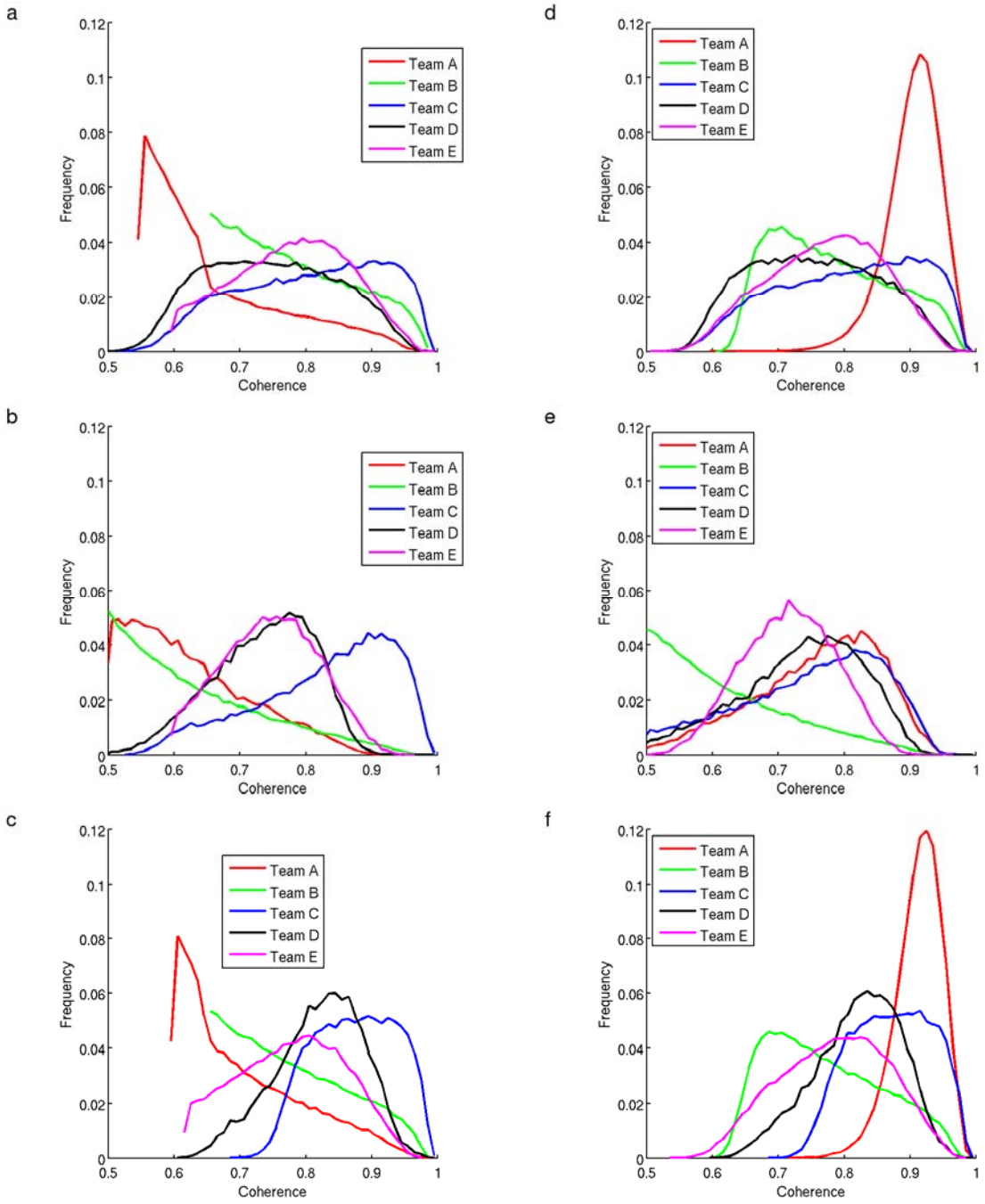

Figure 3.11: Normalized histograms of the estimated temporal coherence. Top) Amsterdam, Envisat time series, middle) Alkmaar, ERS time series, bottom) Alkmaar, Envisat time series. Left) Based on temporal coherence estimate values as supplied directly by the teams, right) based on temporal coherence estimate values re-calculated from time series (fit to linear model). Note that differences in the estimation procedure of the linear model can cause small differences between the supplied and re-calculated values, which is especially visible in the low end of the histograms when a threshold is applied (e.g., compare the result of team B in figure 3.11a and d). Apart from these small differences, especially the supplied and re-calculated coherence values of team A differ significantly, indicating additional a-posteriori smoothing of the time series. Curves that decrease for smaller coherence values indicate that the selection of PS is based on more criteria than estimated temporal coherence alone. These extra criteria applied by some teams were not explicitly stated, hampering fair comparative analysis.

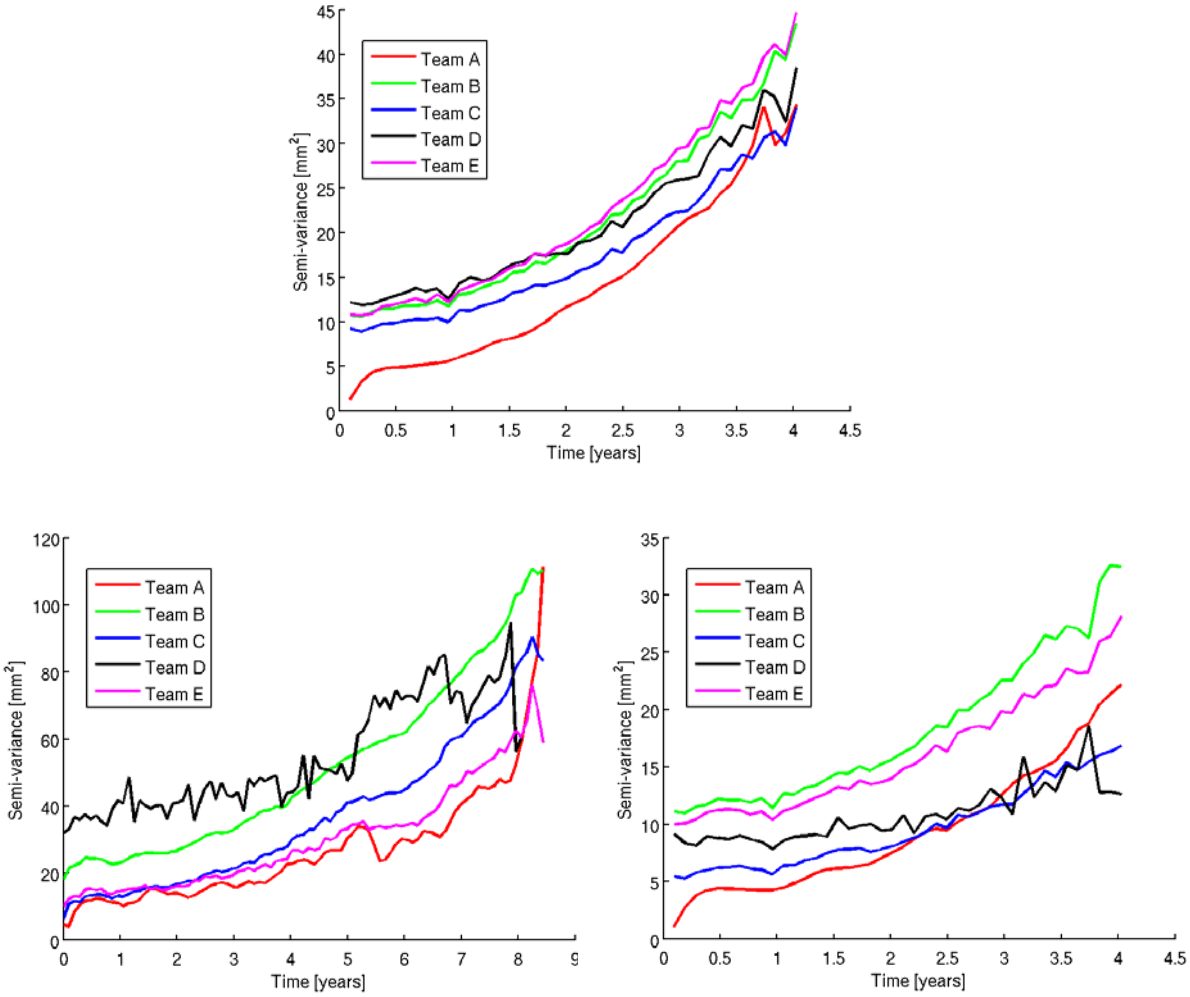

Figure 3.12: Semi-variograms of the PS time series, indicative of the roughness of the estimated time series. Roughness varies a factor 2-3 between teams. Above) Amsterdam, Envisat, left) Alkmaar, ERS, right) Alkmaar, Envisat. The semi-variogram of team A shows a much smaller nugget than the rest of the teams, indicating a smooth signal at short time spans. A limited roughness could be either due to better estimation of noise components, such as the atmospheric phase screen, or due to a posteriori smoothing of the time series. In the latter case, the criteria for this smoothing should be made explicit for comparison.

4 VALIDATION AMSTERDAM

In this chapter the Envisat data set of Amsterdam is validated. The reference data used is described in Section 5.1. The validation of the PSI measurements is reported in Section 5.2.

4.1 Reference data

A subset of the prisms on the buildings along the N/S-line is selected based on the vicinity of PS data. Displacements in the N/S-metro line region in Amsterdam are caused by geotechnical instability[3] and localized construction work. These displacements are expected to be mainly spatially uncorrelated. The on-line monitoring system comprises about 74 Total Stations, computer controlled tachymeters, that monitor some 5350 prisms on about 1800 buildings in continuous operation. The average sampling rate is once every 4 hours.

To be able to compare PS data with the N/S-line tachymeter data, some pre-processing has to be done.

1) Correction of geographical coordinates.
2) Selection of suitable prisms based on available closeby scatterers along the N/S-line.
3) Referencing of tachymeter data and PS data to a common velocity-frame per tachymeter reference group (see Section 2.1).
4) Transformation of the tachymeter data (x, y, z) to radar line-of-sight (LOS) displacements.
5) Averaging of measurements per day and per (part of a) building, if applicable.
6) Estimation of linear deformation rates of the tachymeter data, to allow direct comparison with PSI deformation rates.

These steps are described in Sections 4.1.1 to 4.1.5. After these preparation steps, both time series and velocities are compared. A detailed description and results are given in Sections 4.2.1 and 4.2.2.

4.1.1 Correction of geographical coordinates

During inspection of the position of the scatterers in geographical coordinates, it was observed that there were discrepancies between the teams when looking at common scatterers, i.e. scatterers with the same radar coordinates (range, azimuth). An example of the observed offset is given in Figure 4.1 (left). These discrepancies affect the selection of scatterers, for example when a search radius is used around a building. It therefore affects the comparison of the teams. In order to minimize these geo-coding errors, a correction has been applied to the geographical coordinates of the Amsterdam Envisat dataset. The correction factor per team is determined by calculating the average difference in distance of common scatterers with respect to team A. Next, the correction has been applied to all scatterers, not only the common scatterers. The choice of team A as a reference is arbitrary. The right picture of Figure 4.1 shows the corrected positions of the common scatterers.

[3] Note that we refer to this movement as autonomous deformation, as it is prior to, and therefore not related to the construction works, and frequently localized to a specific building.

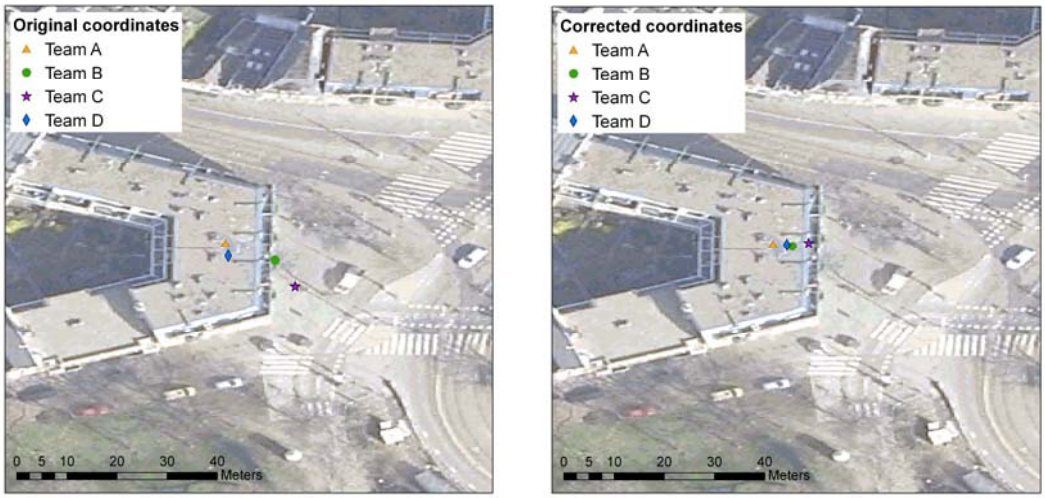

Figure 4.1: Amsterdam, Envisat. Example of the position of common scatterers, i.e. scatterers with the same radar coordinates, before (left) and after correction (right).

4.1.2 Selection of prisms

The prisms on the buildings along the N/S-line are selected based on the presence of PS data within a certain radius. Taking into account the absolute positioning accuracy of the scatterers, the overall size of the target buildings, and a workable dataset size, a radius of 15 meter is chosen. After the selection based on radius, a visual inspection is performed of the selected prisms and buildings, to assess whether the scatterers are really reflected by the specific building. Furthermore, if the building is large and different sides of the building show distinct displacement patterns the building is handled in parts. The final number of buildings and selected scatterers of each team is given in Table 4.1. An overview of selected N/S-line prisms per reference group is given in Figure 4.2.

Table 4.1: Final number of selected buildings and scatterers per team

# selected buildings	PS Team A	PS Team B	PS Team C	PS Team D
354	665	392	467	522

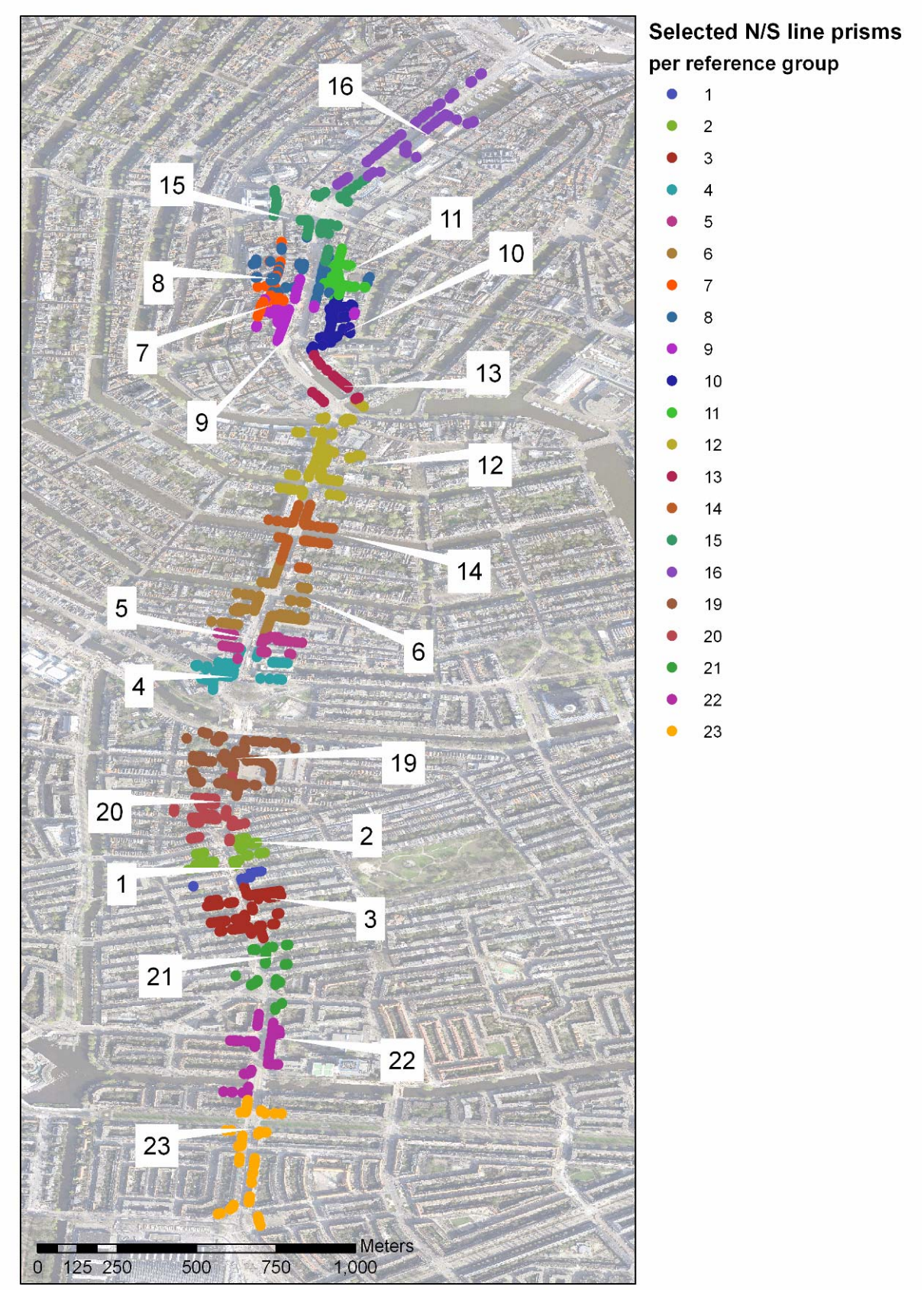

Figure 4.2: Overview of selected N/S-line prisms per reference group

4.1.3 Referencing to a common velocity-frame per reference group

The geodetic monitoring system of the N/S-line is a fully automated system consisting of 74 robotic tachymeters (Total Stations) aimed at 5350 prisms on 1500 constructions along the 3.8 km transect measuring individual prisms in (x,y,z). The Total Stations are split into groups of up to five instruments each to form a local geodetic network. In total there are 23 different groups (Figure 4.2). The measurements are made within the local system and are linked to other groups through a referencing based on least-squares adjustment. Despite the referencing, still artifacts are visible in the transitions between the local networks. To circumvent the influence of these artifacts in the validation process, the PS data is referenced to the individual reference groups.

The data set that was made available by N/S-line project office consists of the 4-hourly measurements at the selected buildings on the dates of the InSAR measurements. For each prism the reference group is given. For each reference group, one building is selected as the reference building. This selection is based on several criteria.

- The building should have scatterers of all four teams within its search radius (15m.).
- Preferably the building was also considered as reference building by N/S-line project office.
- The building has a low LOS velocity.
- The building has a time series with data points on as many SAR acquisition dates as possible.

Next, for each selected reference building and for each team a velocity correction is estimated by taking the difference between the building LOS velocity and the mean velocity of the scatterers per team,

$$V_{corr} = V_{NSbuilding} - \frac{1}{N}\sum_{i=1}^{N} V_{team}^{i} \,. \qquad (1)$$

The velocity corrections range from -1.67 mm/yr to +1.83 mm/yr. This way, the corrected velocities can be considered to be in the same reference frame as the tachymetric data, enabling a direct comparison.

4.1.4 Transformation to LOS and averaging of measurements per day

Generally, along the N/S-line there are measurements every 4 hours. This means that there is a maximum of 6 measurements per day. PSI offers one measurement for that day so that in order to compare tachymetry and PSI the former is averaged over one day. Prism x- y- and z-displacements are averaged per day by taking the median of the measurements. The median is taken in order to have less effect of outliers. Histograms of the daily standard deviation in x, y, and z are given in Figure 4.3. Next, the (x,y,z) displacements are transformed into LOS displacements to get the LOS time series for each individual prism.

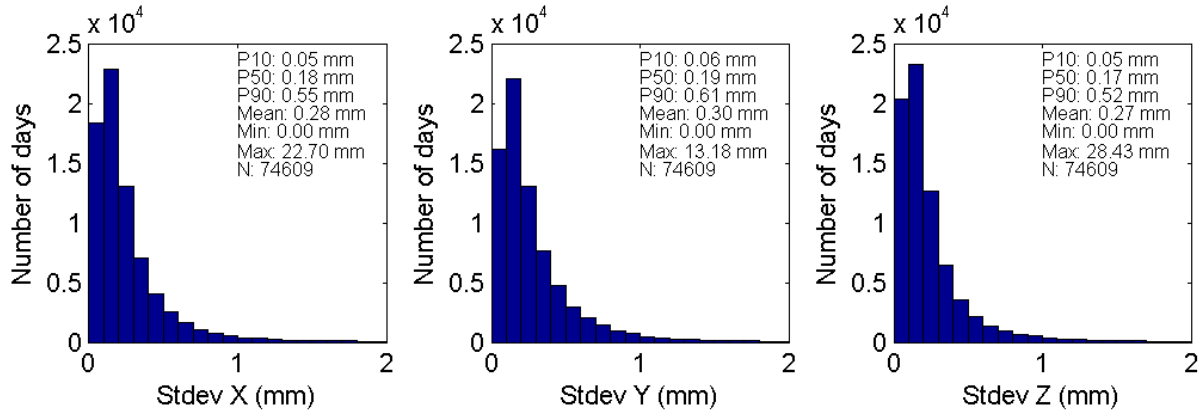

Figure 4.3: Daily standard deviation of prism displacements in X, Y, and Z (mm), as observed from tachymetry.

4.1.5 Averaging per building and estimation of linear deformation rates

On each building two or more prisms are attached depending on the size of the building. After transformation to LOS and averaging per day, the building LOS time series is calculated by taking the median of the time series of the available prisms. Finally, the linear deformation rate is determined by estimating the trend through the median LOS time series. Figure 4.4 shows the histogram of the standard deviation in linear deformation rate due to the averaging per building. Figure 4.5 shows an example of transformation to LOS and averaging per building. If the building is large, the prism time series are averaged in groups per part of the building.

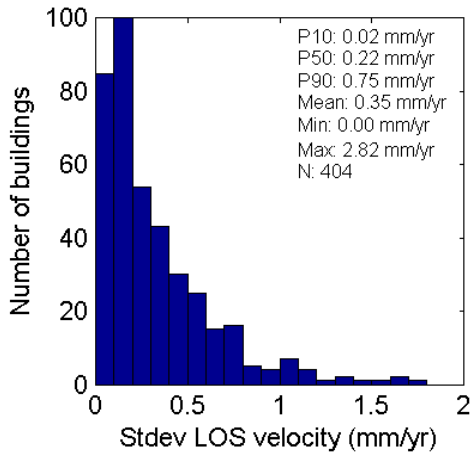

Figure 4.4: Building standard deviation in LOS velocity (mm/yr), as observed from tachymetry.

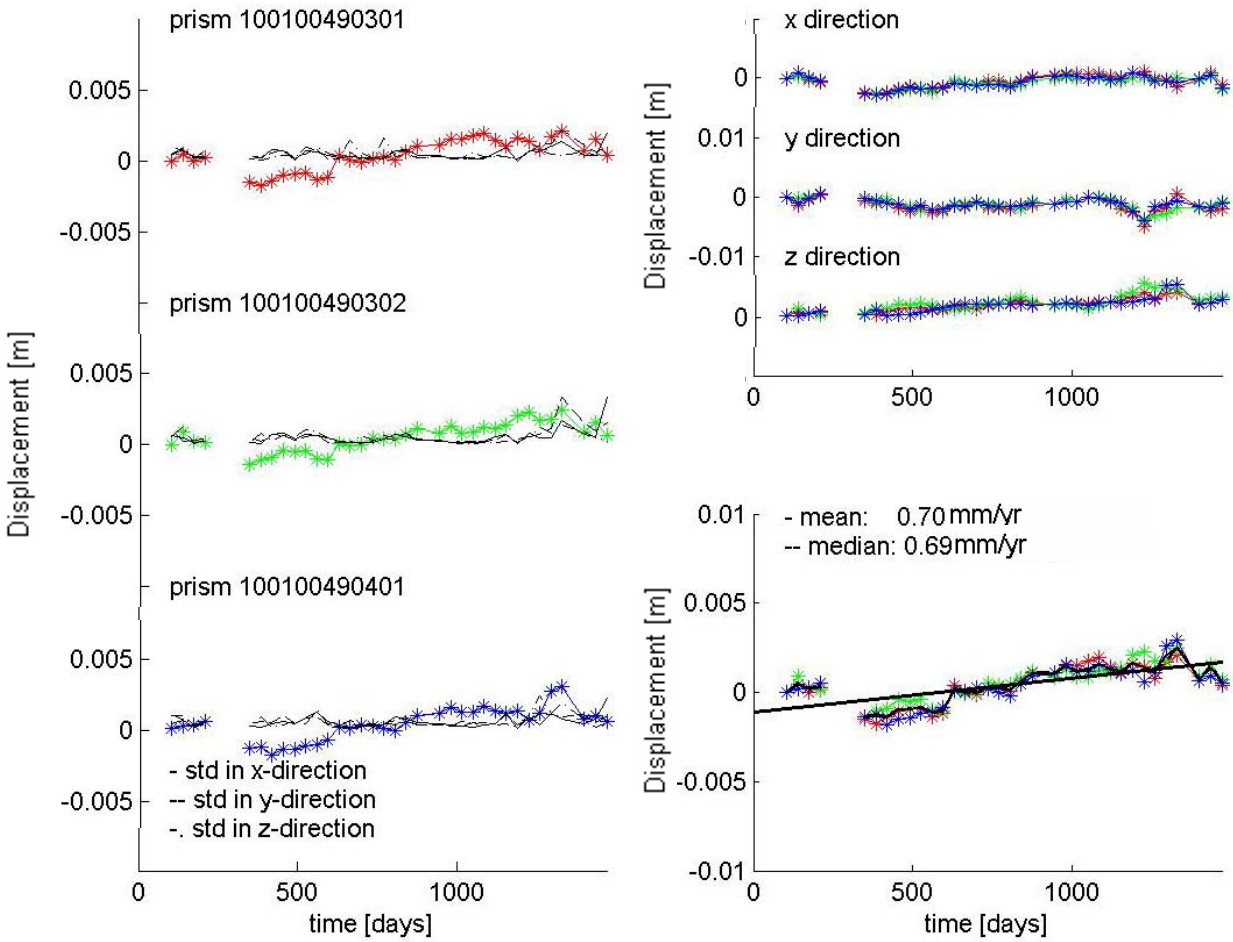

Figure 4.5: Example of the averaging per day and building. On the left the LOS time series of three individual prisms are displayed, together with the daily standard deviation. On the upper right the x, y, and z time series are displayed. On the lower right the LOS time series of the individual prisms and the median and mean LOS time series of the building are displayed, together with their trend line. In this case, mean and median give almost the same LOS time series, therefore they are not clearly distinguishable.

4.2 Validation in the measurement space

In this chapter the comparison of PS data with groundtruth data in Amsterdam is described. As described in the previous section, a selection of prisms is made primarily based on the availability of scatterers within a search radius of 15 meter. The comparison of tachymetry (ground truth) data and PS data comprises a validation of linear displacement velocities and a time series validation.

4.2.1 Validation linear displacement velocities

As can be seen in the example in Figure 4.6, the number of scatterers near a building varies per team (and also per building), ranging from 0 up to ~10 scatterers. Also, the variability and trend of the scatterers can vary widely. For a sound comparison between N/S velocity and PS velocity, for each building an average PS velocity is calculated per team by taking the median of the velocities of the individual scatterers. An overview of the standard deviation of this averaging is given in Figure 4.7. This figure shows that the mean standard deviation varies between 0.53 mm/yr and 0.60 mm/yr. This is almost twice as much variation as in the N/S-line velocities (mean of 0.35 mm/yr), shown in Figure 4.4. However, the standard deviation shown in Figure 4.4 has a slightly different meaning, since the actual variability in N/S-line velocity does not only depend on the averaging of the prism LOS velocities, but also on previous processing steps such as the averaging per day.

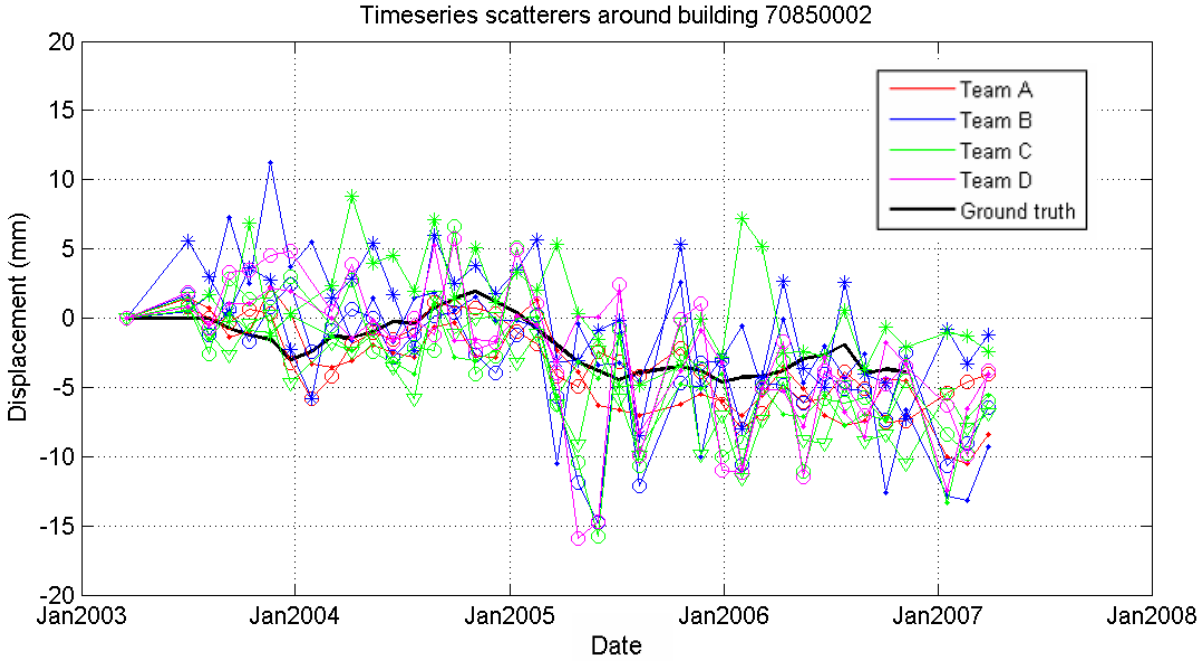

Figure 4.6: Example of PS time series and median N/S-line LOS time series. Some teams identified more PS than others.

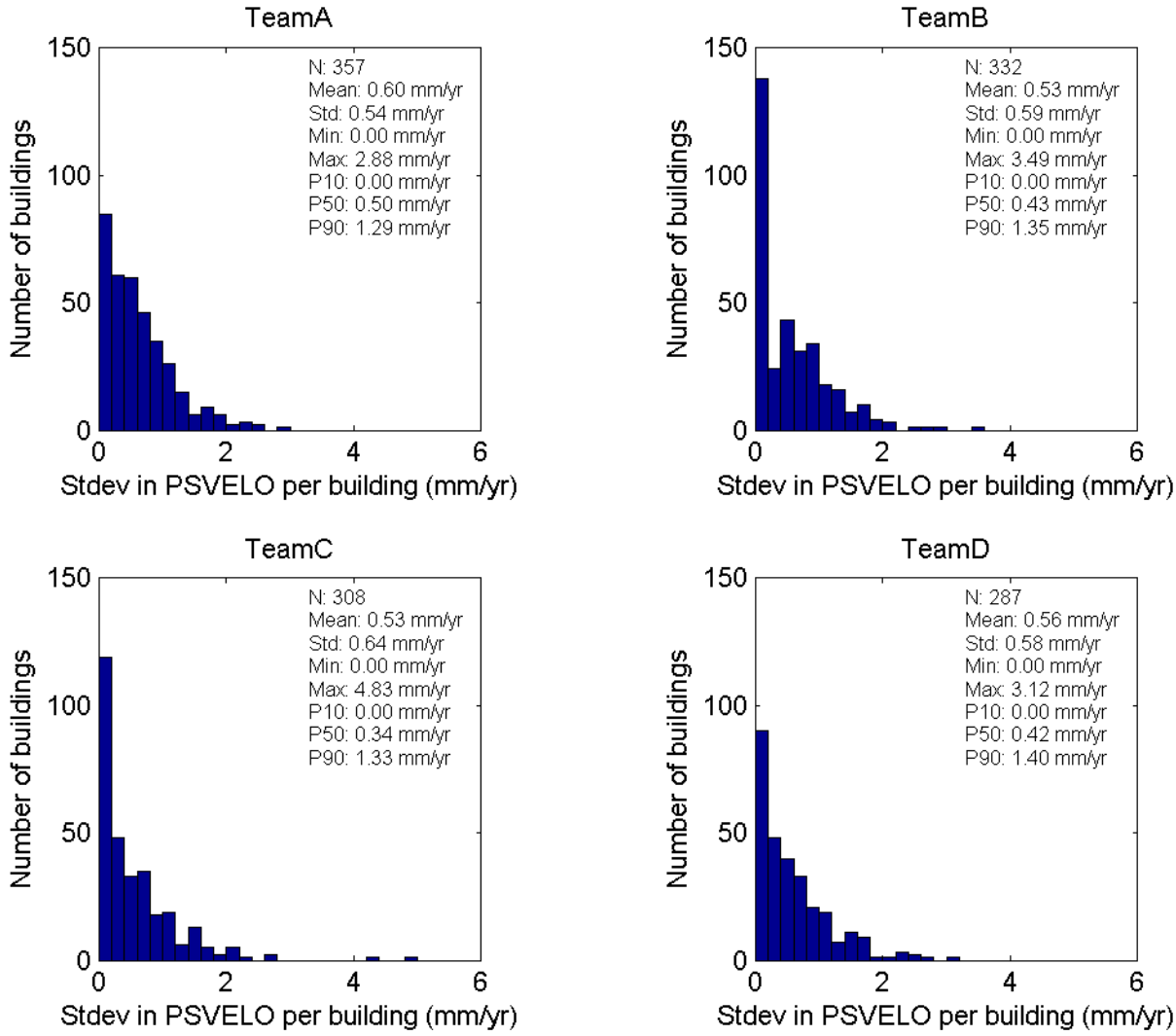

Figure 4.7: Standard deviation in LOS PS velocity per building.

Figure 4.8 shows scatter plots of (referenced) PS velocities against tachymetry velocities for each team. In each plot the trend line through the data points is plotted, together with the correlation coefficient R. The difference in velocity (PS velocity minus tachymetry velocity) is plotted in histograms in Figure 4.9.

To check the velocity corrections applied during referencing, the scatter plots of (referenced) PS velocity against N/S-line velocity for each team can be examined per reference group. If the cross plot of a particular reference group shows a distinct offset from the x=y line, the correction velocity is under- or overestimated. In this case, the choice of reference building is reconsidered, and processing steps are repeated with a new reference building.

A different approach to velocity validation is to calculate 'double differences' in stead of 'single differences' as displayed in Figure 4.9. The double difference in velocity is defined by the difference in PS velocity between two buildings, compared with the difference in N/S-line velocity of the same pair of buildings:

$$V_{DD} = \left(V_{PS}^{i} - V_{PS}^{j}\right) - \left(V_{N/S}^{i} - V_{N/S}^{j}\right) = \Delta V_{PS} - \Delta V_{N/S} \qquad (2)$$

In which i and j denote different buildings. The advantage of using double differences in stead of single differences is the fact that the two datasets do not need to be referenced to the same reference frame, as long as the datasets are not split into different reference groups. In the case of Amsterdam, the N/S-line data is split into reference groups, whereas the (original) PS data is not. Comparison of velocity differences between building pairs can only be done for building pairs belonging to the same reference group. This is taken into account when calculating the double difference in velocity, displayed in Figure 4.10.

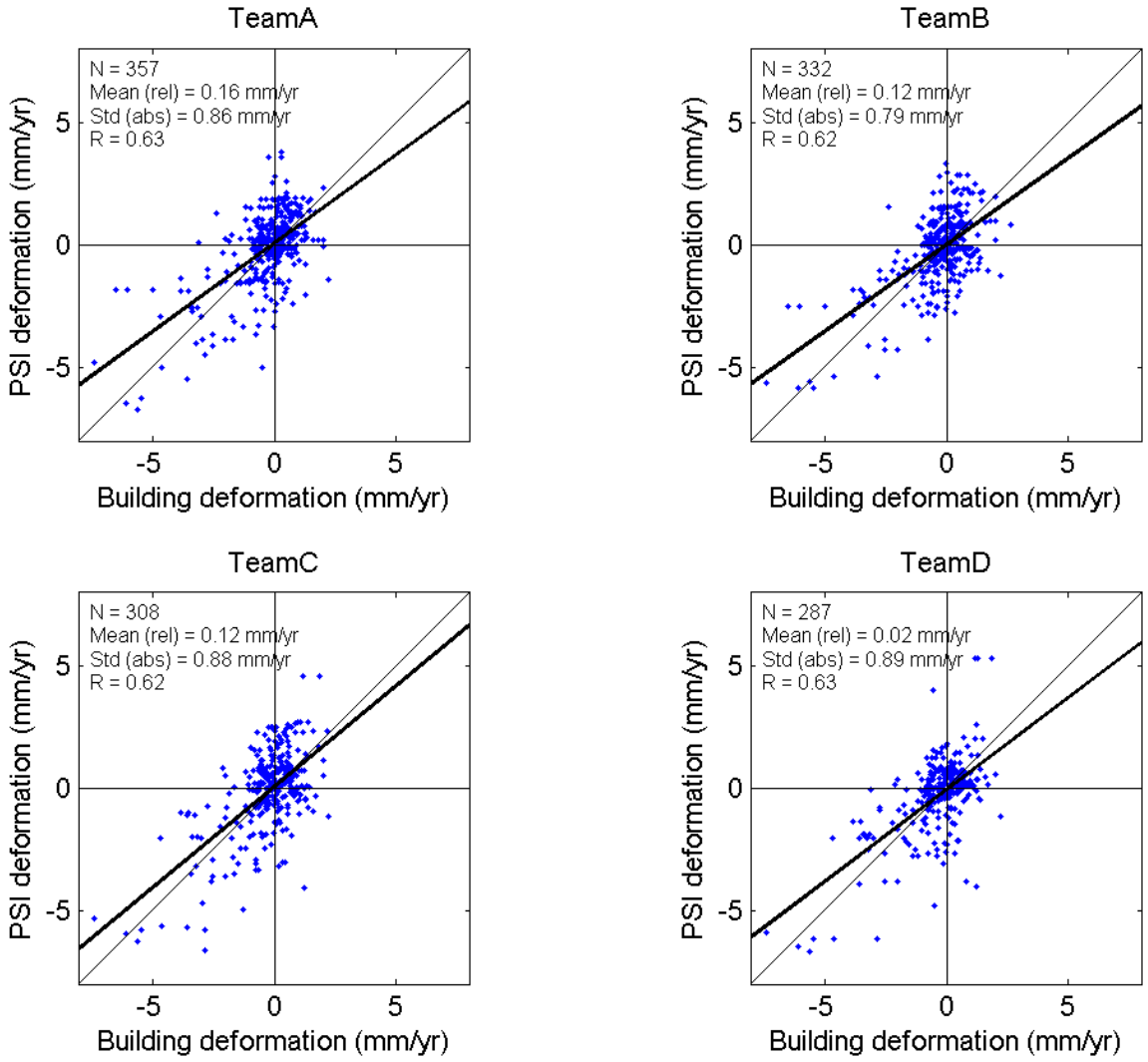

Figure 4.8: Scatter plots of PSI deformation and tachymetry-based building deformation per team. For visualisation the thin black lines x=y, x=0 and y=0 are plotted. The thick black off-diagonal line is the trend line through the data points. Generally, the PS velocity estimates appear to be slightly smaller than the tachymetry-based estimates.

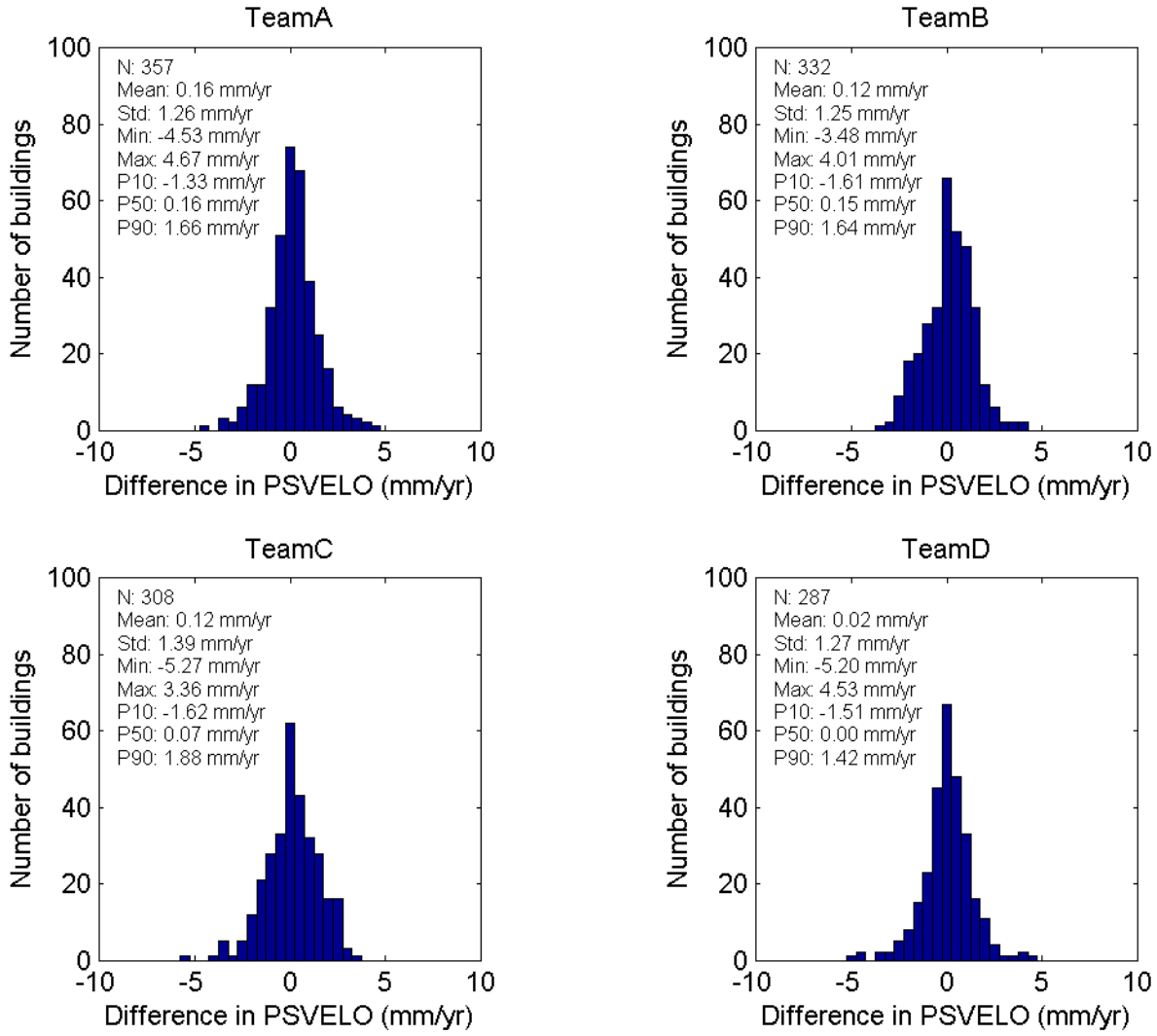

Figure 4.9: Histograms of the (single) difference in velocity (PS velocity minus tachymetry-based velocity). The mean values differ slightly from zero, indicating that there is a small remaining bias between the tachymetry and PSI datum.

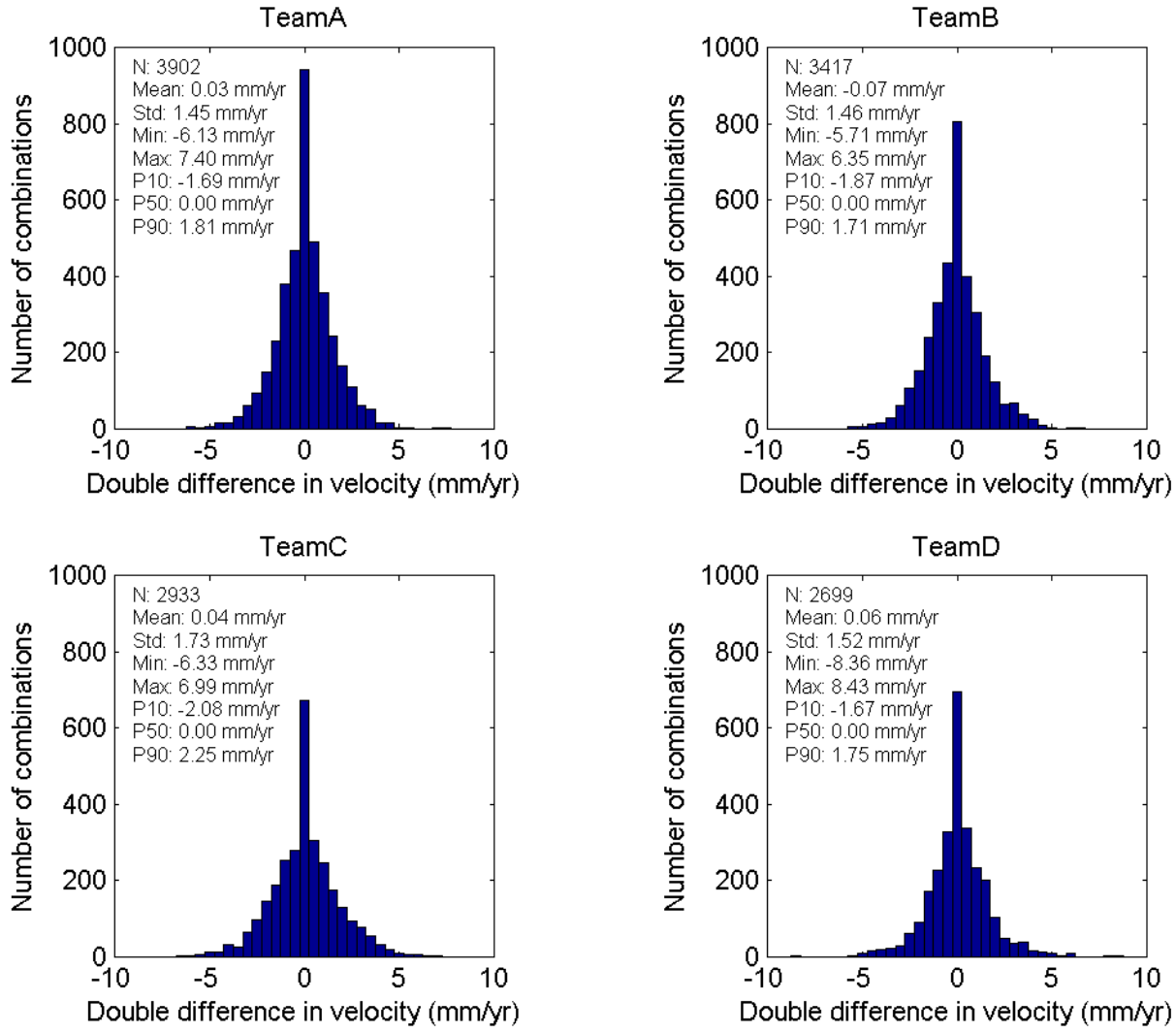

Figure 4.10: Histograms of the double difference in LOS velocity.

Finally, the velocities are plotted in the map of Amsterdam for visual inspection. In Figure 4.11 the linear deformation velocities of the selected N/S-line prisms are displayed. In Figure 4.12 to Figure 4.15 the linear deformation velocities of the selected scatterers of the individual teams are displayed. From these plots it is clear to see that some areas with significant deformation observed from the tachymetry data are also observed from the radar data. Mainly these are locations where construction works for the metro stations are performed.

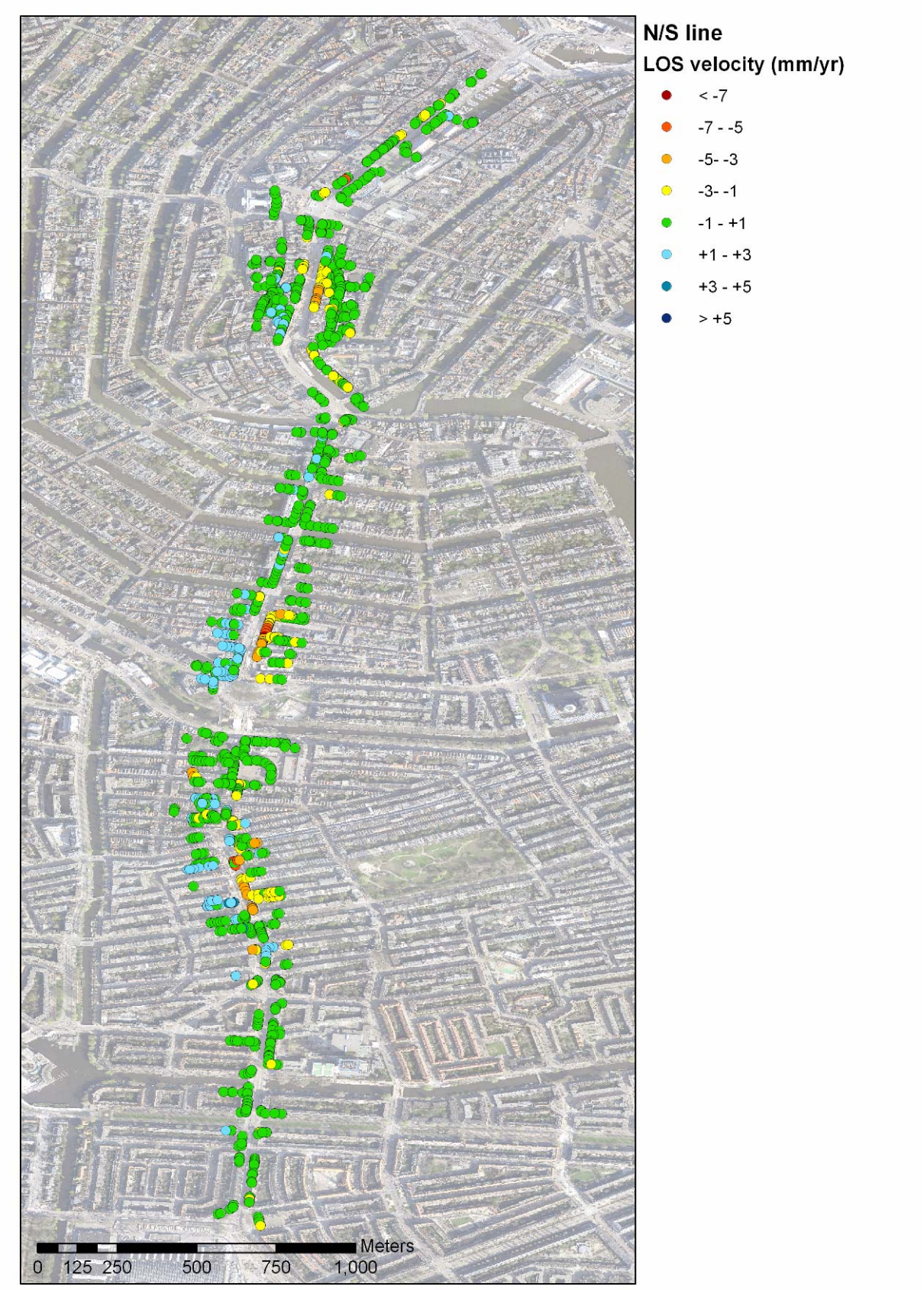

Figure 4.11: N/S-line prism LOS velocities along the N/S-line trajectory

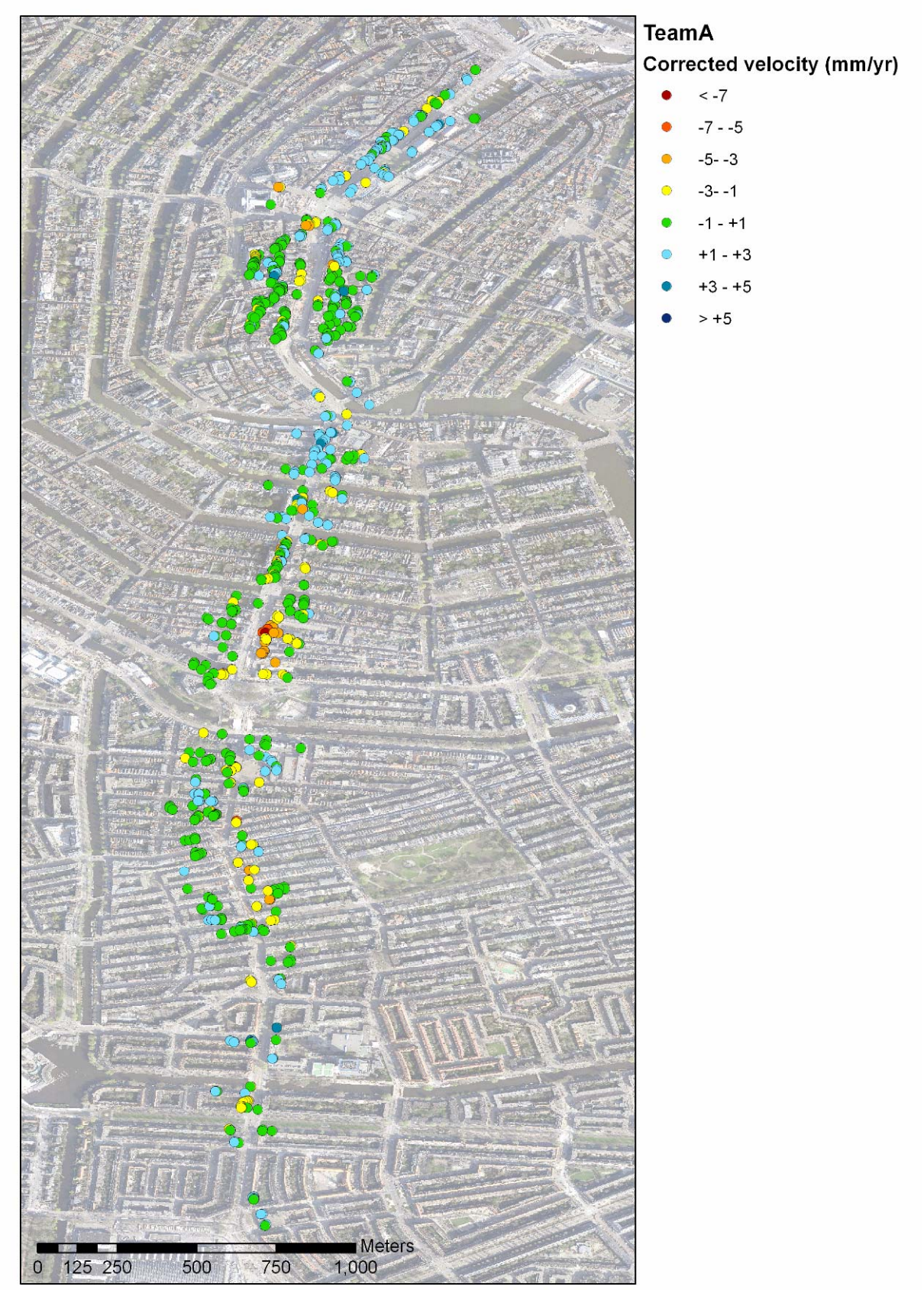

Figure 4.12: Linear deformation velocities of the selected scatterers of team A along the N/S-line trajectory

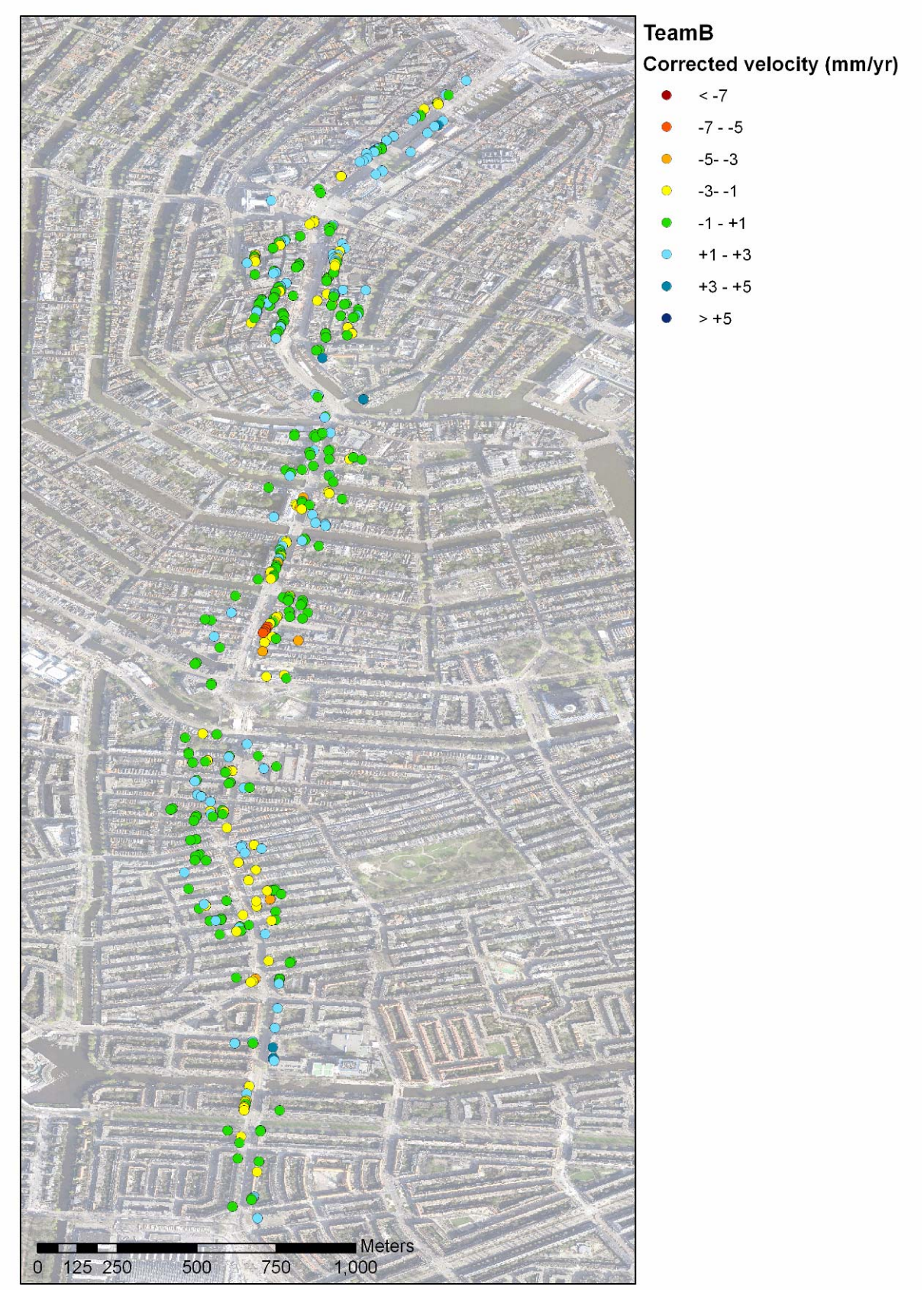

Figure 4.13: Linear deformation velocities of the selected scatterers of team B along the N/S-line trajectory

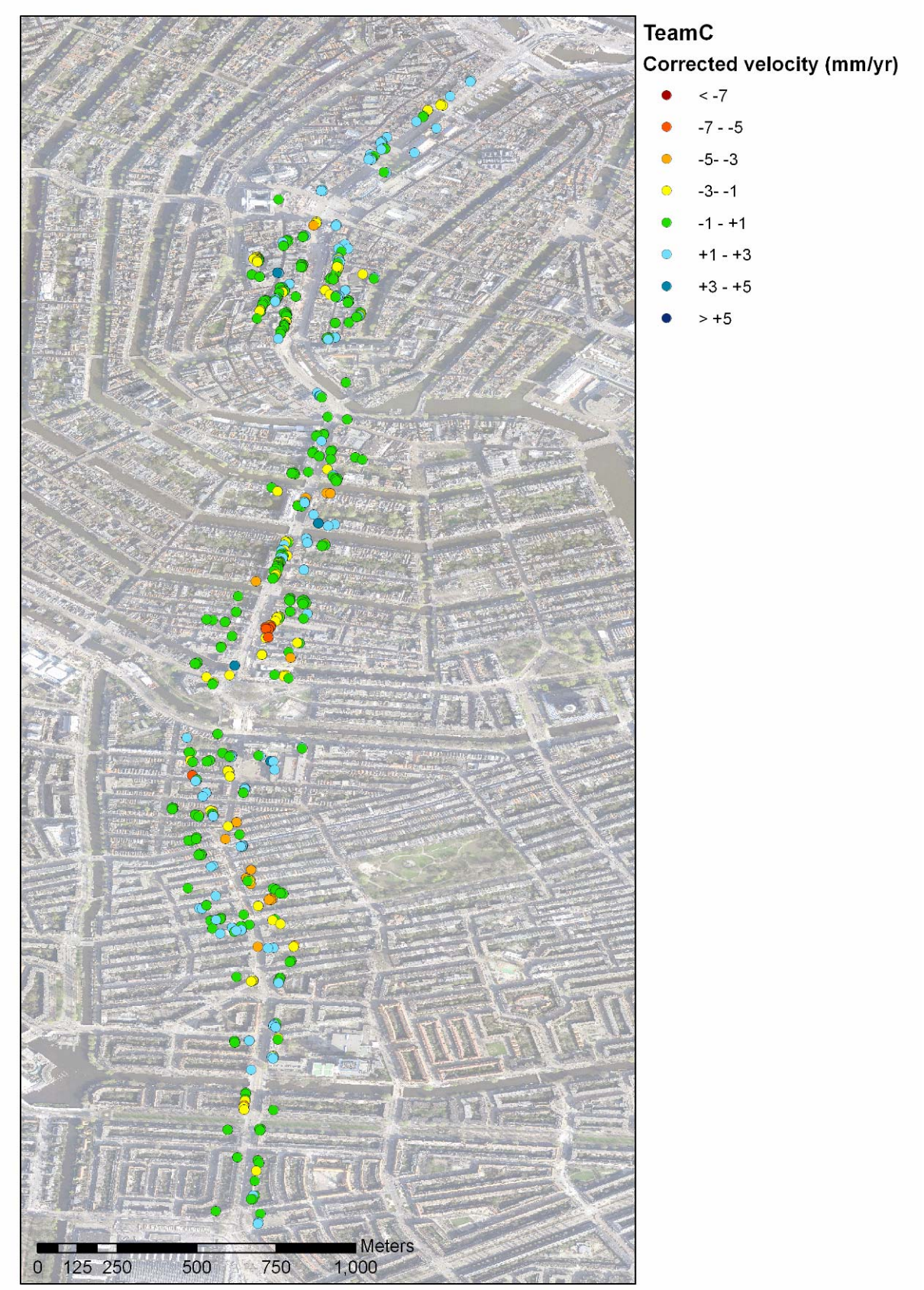

Figure 4.14: Linear deformation velocities of the selected scatterers of team C along the N/S-line trajectory

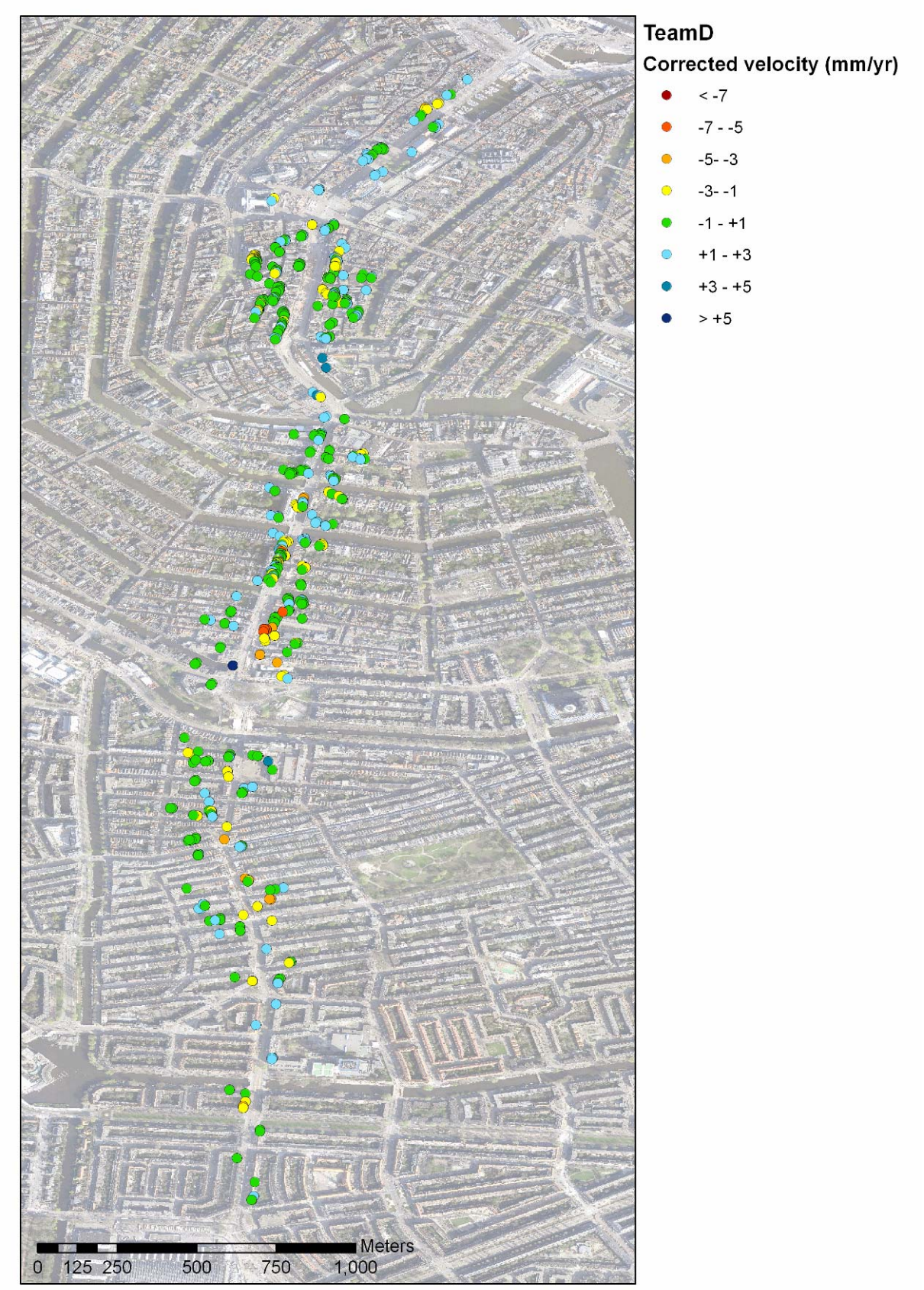

Figure 4.15: Linear deformation velocities of the selected scatterers of team D along the N/S-line trajectory

4.2.2 Validation displacement times series

For the time series validation a statistical approach is used in which the average difference of the PSI time series and the tachymetry LOS time series is quantified by means of a Root Mean Squared Error (RMSE).

RMSE is a common statistic for quantifying errors (Shearer, 1990). RMSE for a time series is defined by

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X_L(i) - X_{PS}(i)\right)^2} \quad . \tag{3}$$

Here: N = number of paired observations
$X_L(i)$ = observation of tachymetry at time i
$X_{PS}(i)$ = observation of PSI at time i

RMSE represents the typical size of discrepancy between the two datasets at a certain prism location, with values equaling or near zero indicating perfect or near perfect fit. The smaller the RMSE, the better is the fit between the two-datasets.

For each team the time series of all the scatterers within the 15 meter radius are compared with the tachymetry time series, taking into account reference groups and parts of large buildings. An example of PS time series and a tachymetry time series of a single building is displayed in Figure 4.6. In Figure 4.16 the results of the RMSE analysis are shown. It can be seen that the RMSE values range between 4.19 and 5.48 mm. This is also summarized in Table 4.2.

Table 4.2: Summary of RMSE analysis: the mean of the RMS error between PS and NSline timeseries for all point pairs on buildings.

	Team A	**Team B**	**Team C**	**Team D**
Mean RMSE (mm)	4.19	5.48	5.09	5.25
N (number of point pairs)	1424	869	1011	1103

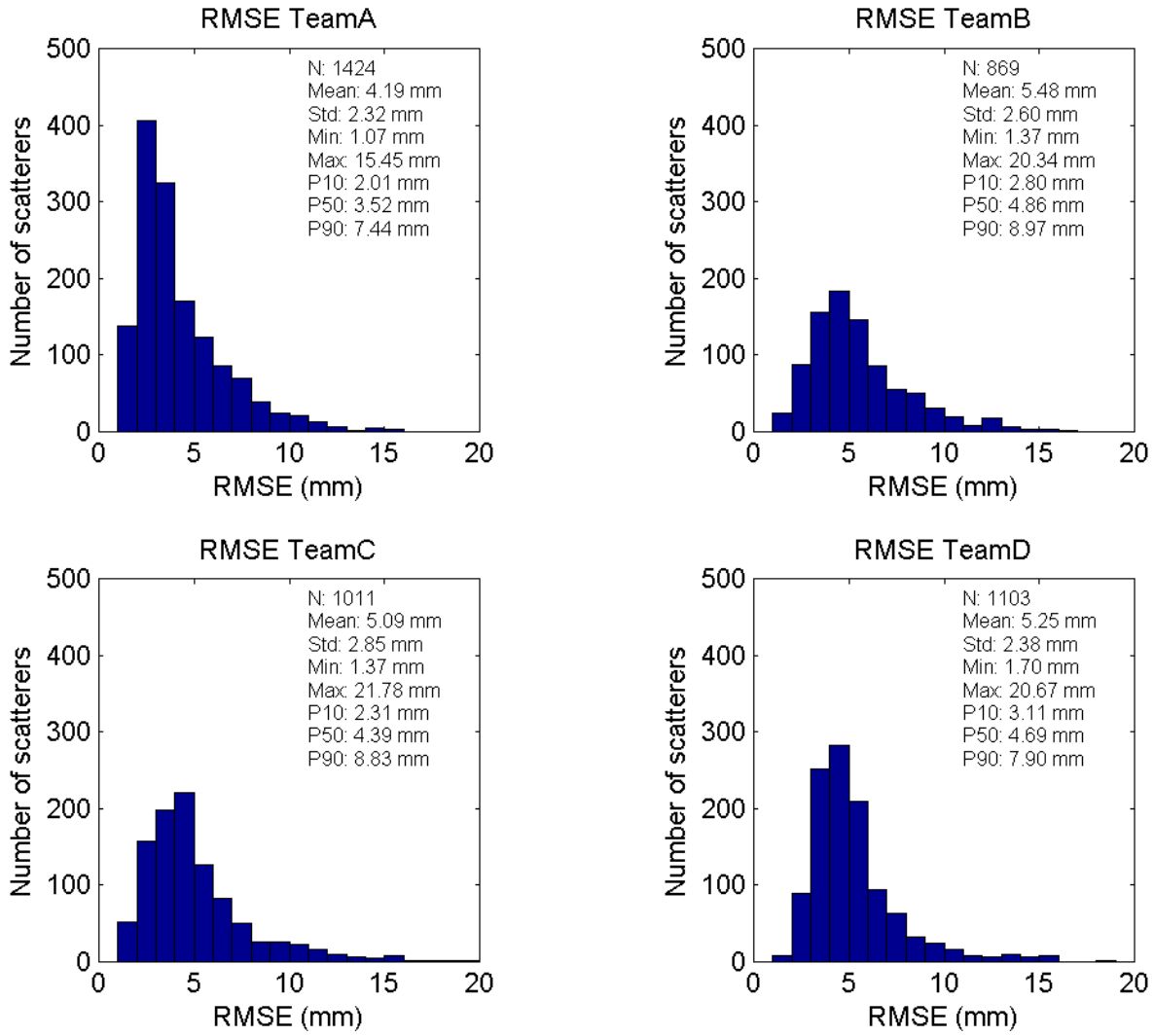

Figure 4.16: RMSE of the time series per team

5 VALIDATION ALKMAAR

In this chapter the ERS and Envisat data sets of the Alkmaar region are validated. The reference data used is described in Section 5.1. The validation of the PSI measurements is reported in Section 5.2. Section 5.3 contains the validation in the parameter space.

5.1 Reference data

Data from levelling benchmarks have been obtained from the Geo-Information and ICT Advisory Department (AGI) of the Directorate-General of Public Works and Water Management (Rijkswaterstaat, RWS). AGI manages the national ordnance surveys and the database containing all national precision levelling campaigns.

The movement of the benchmarks is considered to be a summation of three components.

- Autonomous geological movement.
- Movement caused by the extraction of material.
- Movement caused by settlement.

Autonomous geological movement in the Alkmaar region is mainly due to tectonics, isostatic motion and compaction of deep seated clay layers. Added up, these processes result in a regional subsidence of about -0.5 to -1 mm/yr with respect to the average over the Netherlands. Figure 5.1 shows the vertical movement of deep founded ordnance benchmarks which define the Dutch ordnance level. These benchmarks together form the 1st-order system of benchmarks and are founded on Pleistocene sand deposits. In the Alkmaar area these deposits are about 30 meter below ground level.

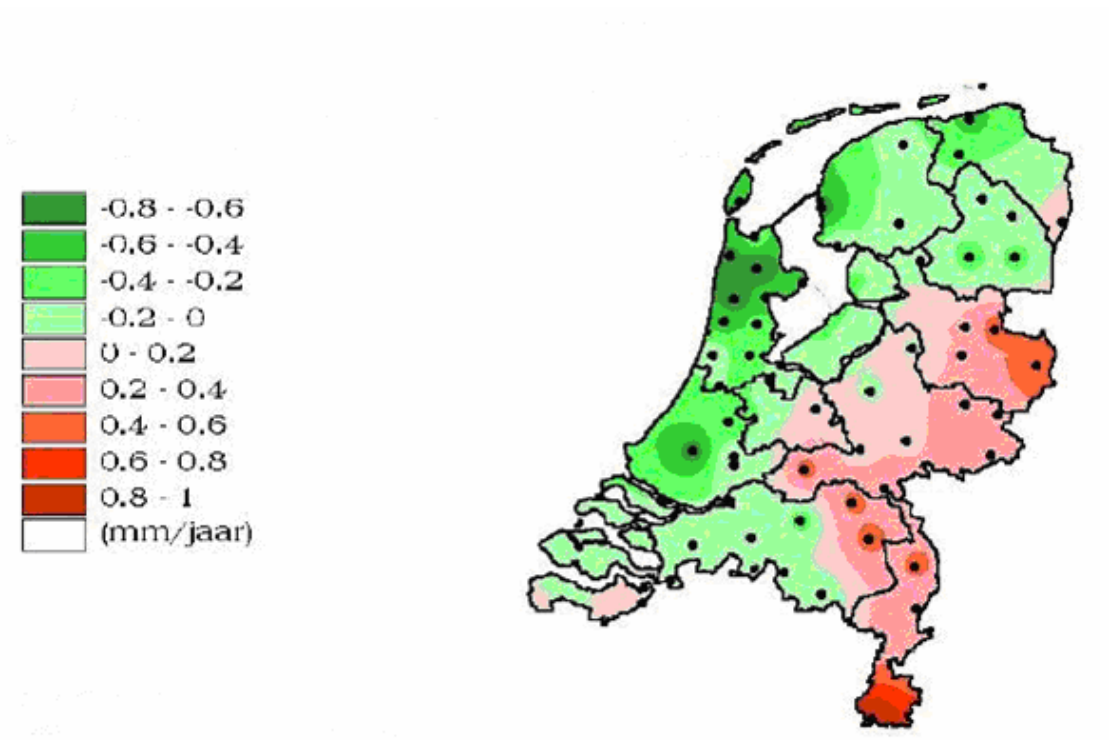

Figure 5.1: Ground movement of the Netherlands in mm/y derived from deep founded 1st-order benchmarks (RWS-DID, 1997)

Since levelling campaigns are closed on the 1st-order system of ordnance benchmarks effects are to be seen over large distances. However, on the scale of the Alkmaar area no large effects are to be expected since the 1st-order benchmarks in this area show differences of maximum 0.2 mm/yr. Differences in velocity are primarily contributed to shallow phenomena (like settlement of constructions in which the 2nd-order benchmarks have been placed) or subsidence due to gas production.

Shallow phenomena relate to the nature of the shallow geological deposits in this area. The area has a thick deposit of Holocene marine and fluvio-marine material and is prone to settlement. Figure 5.2 shows the distribution of settlement-prone areas in the Netherlands, related to the presence of clay and peat deposits. As can be seen, apart from the dune area, most of the Alkmaar-area is prone to settlement. Depending on the foundation of the construction this effect will add up to the total movement.

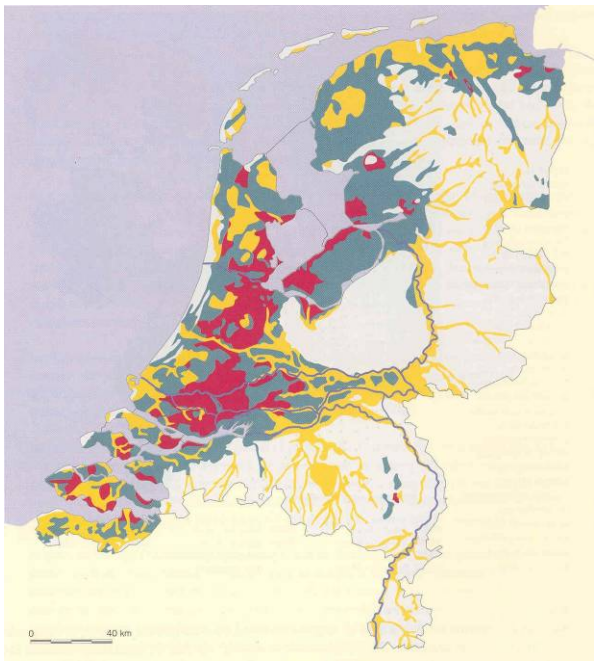

Figure 5.2: Distribution of settlement sensitive areas in the Netherlands (brown = predominantly peat, dark blue = predominantly clay, orange = sandy clay).

The quality of the estimated heights of the levelling benchmarks can only be estimated, as precise information is unavailable. The original levelling data can be characterized by a precision of 0.7 mm per square kilometer. However, due to unknown network design and adjustment procedures, the variance-covariance matrix of the heights cannot be reconstructed. Therefore, in the further validation procedure the assumption is made that the heights are uncorrelated and have a standard deviation of 1 mm/year.

Levelling data for the period under consideration is available for 945 benchmarks in the area around Alkmaar. The spatio-temporal distribution of the benchmark heights is shown in Figure 5.3a (red dots). The green and blue dots represent the acquisition times of ERS and Envisat, respectively. The projections of the plot in the spatial and temporal domain are shown in Figure 5.3b and c. It shows that there are hardly any measurements in the eastern

part after 1997 (near the Middelie gas field). This is probably because the extraction stopped in 1992 (Winningsplan Midelie area, 2005) and no major displacement is expected in this region afterwards[4]. Furthermore, the number of levelling benchmarks coinciding with the time spans of radar acquisitions is limited, especially in the Envisat case. Because of the difference in available levelling heights between the ERS and Envisat set, a different strategy for the analysis is chosen for both sets.

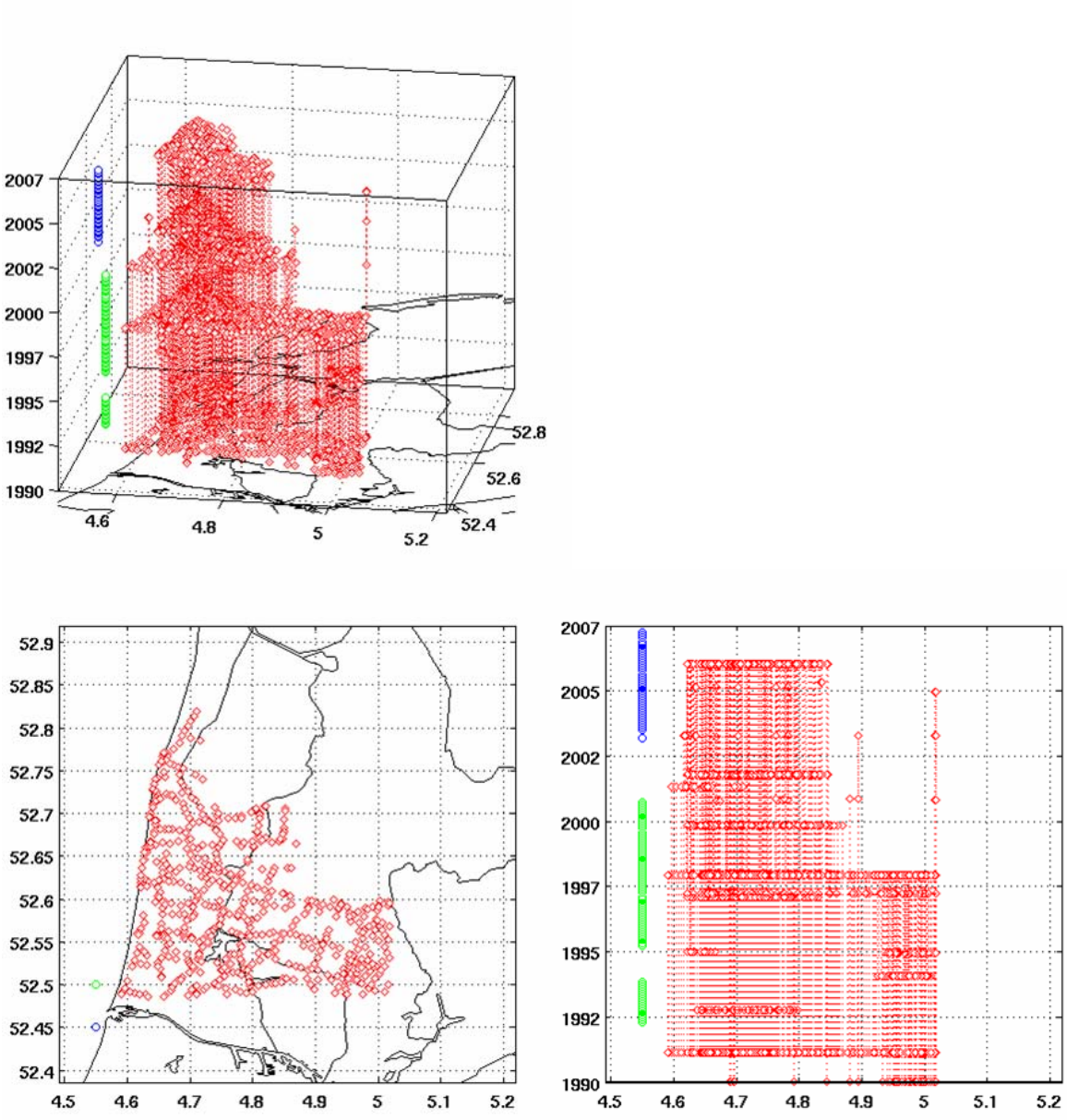

Figure 5.3: Above) Spatio-temporal distribution of levelling heights in the area around Alkmaar from 1990 to 2007 (red dots). The green and blue dots represent the acquisitions by the ERS and Envisat satellite respectively. Left) Distribution of heights in the spatial domain. Right) Distribution of heights in the temporal domain.

ERS

In case of the ERS, levelling campaigns up to two years before and after the radar time series are used, see Figure 5.4. The time span of the radar acquisitions is marked with the gray

[4] Extraction re-started in 2007

vertical lines. To enable the validation of the PS velocities, a linear velocity is estimated through the levelling heights. A testing scheme is applied to remove outliers, adopting a standard deviation of 1 mm and a critical value of 1.96 mm. Only the benchmarks with, after testing, at least 3 measurements are used for the validation. Figure 5.4 shows an example of levelling data for a benchmark. The blue dots are the accepted heights, whereas the red measurement is rejected. The solid blue line is the final estimated linear model. The dashed lines show the 1 mm standard deviation interval. The example shows that the linear model is a good representation of the subsidence due to gas extraction.

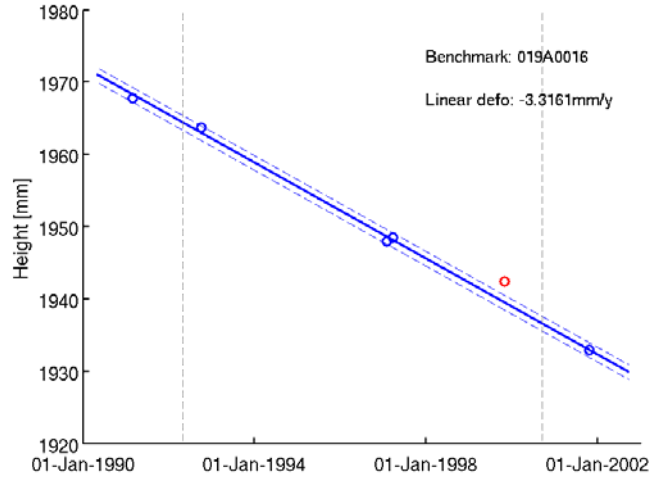

Figure 5.4: Example of estimated linear displacement model for a certain levelling benchmark covering the ERS time span. Levelling campaigns up to 2 years before and after the radar time series (marked by vertical gray lines) are used. A testing scheme is applied to remove outliers (red dot).

In total 219 benchmarks fulfill the criteria and are used in the validation. The estimated linear velocities [mm/year] for these benchmarks are shown in Figure 5.5a. Subsidence phenomena due to the gas extraction are clearly visible. The number of remaining measurements per benchmark after testing ranges from 3 to 7, which is shown in Figure 5.5b.

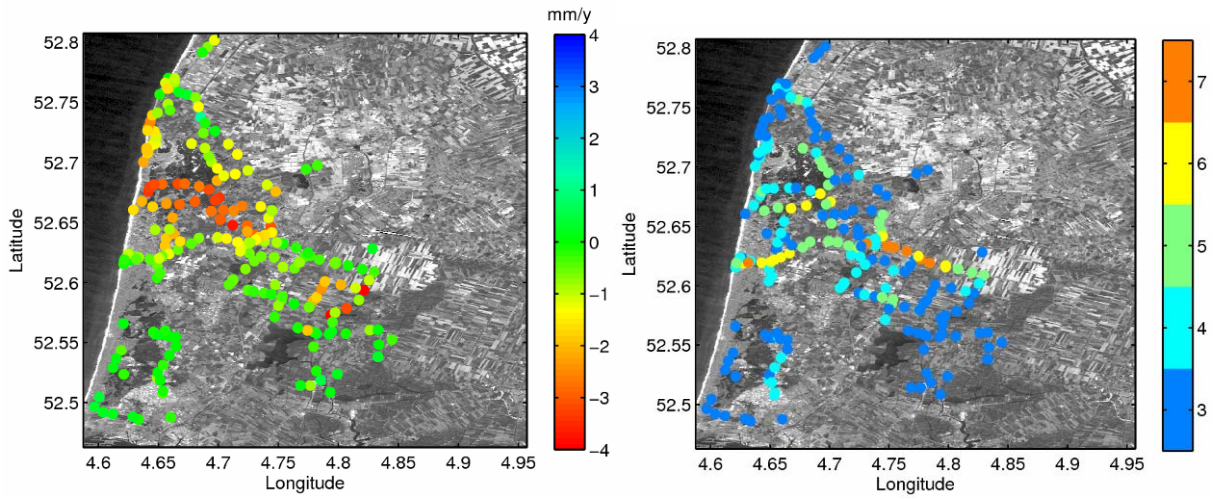

Figure 5.5: a) Estimated linear displacement velocities from levelling data covering the ERS time span. In total 219 benchmarks fulfill the testing criteria and are used in the validation process. b) Number of measurements per benchmark (after testing).

Envisat

The number of levelling campaigns covering the Envisat time span is unfortunately very limited, even with an extra buffer of two years. As an example, the available data for the same benchmark as in Figure 5.4 is shown in Figure 5.6. As for most benchmarks, only two measurements are available (see Figure 5.3, right plot). Therefore no testing could be applied and the calculated velocities are very vulnerable for measurement errors. Note the relatively large reduction in displacement velocity between this time span and the ERS time span for this example (-0.82 vs. -3.32 mm/y). Hence, caution should be taken by interpretation of the validation results.

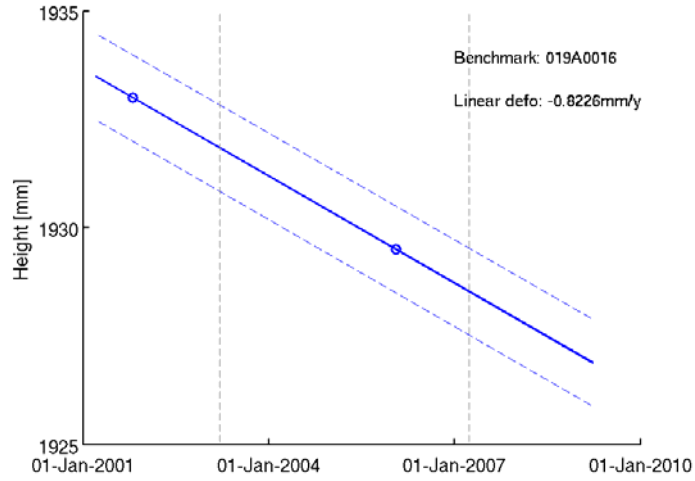

Figure 5.6: Example of estimated linear displacement model for a certain levelling benchmark covering the ASAR time span. Levelling campaigns up to 2 years before and after the radar time series (marked by vertical gray lines) are used. No testing could be applied due to the limited number of available measurements.

In total 230 benchmarks are available for the validation of the Envisat data. The spatial distribution of the linear velocities and the number of available levelling campaigns is shown in Figure 5.7.

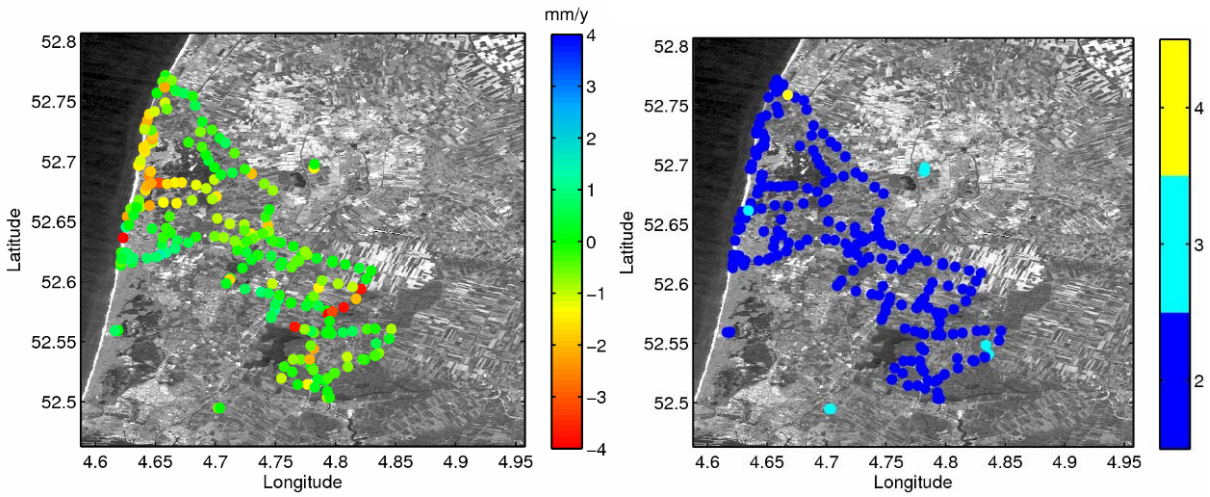

Figure 5.7: a) Estimated linear displacement velocities from levelling data covering the Envisat time span. In total 230 benchmarks are used in the validation process. b) Number of measurements per benchmark.

5.2 Validation in the measurement space

The PSI results are directly validated against levelling measurements, denoted by a validation in the measurement space. The PS data is first transformed to the vertical direction to coincide with the levelling data. Note that we can safely assume that for the area around Alkmaar, gravimetric changes are negligible, which enables us to consider the orthometric levelling information as geometric deformation and thus comparable with PSI. Subsequently, for each levelling benchmark (see Figure 5.5 and Figure 5.7) the PS within a radius of 50 meter are selected per team. The representative value of PSI for each benchmark is obtained by selecting the nearest neighbor within the 50 meter region. The results taking the mean and median values are also evaluated, but do not result in significant changes. Only levelling benchmarks which have at least one PS within the radius are used for the evaluation. This results in case of ERS1/2 in 36 to 151 benchmarks per team. This indicates the large dispersion in PS density per team.

Because levelling and PSI result in measurement in a different datum (orthometric heights vs. geometric heights), the datums have to be connected. This is done by taking the mean difference between the estimated linear deformation velocity from levelling and PSI, and adjusting the PSI estimates. Note that also the displacement time series are corrected for this datum shift.

The validation results for the Envisat time series have to be interpreted with care. Major reason is the lack of levelling epochs in the Envisat time span. The estimated linear velocities are often based on only two measurements, which prevents a proper testing of the levelling measurements. Moreover, the Envisat data set of team A is assumed to contain significant phase integration errors. However, this effect seems limited in the area of interest and therefore does not seem to influence the validation results much. Within the Envisat data set of Team D the displacements are underestimated with a factor 2. Yet, this effect does not propagate clearly in the product validation results, probably because of the low signal-to-noise ratio.

5.2.1 Validation linear displacement velocities

The linear velocities derived from the PSI data at the levelling benchmarks obtained by the nearest neighbor approach are visualized in Figures 5.8 and 5.10 for ERS and Envisat data respectively. The differences between the PSI and the levelling estimates is shown in Figures 5.9 and 5.11. In the lower right plot the levelling estimates are shown as reference.

The correlation between the PSI and levelling estimates is more clear from the scatter plots in Figures 5.12 and 5.13. The plots show the results for the original and the de-trended data, with or without correction for the bias between the levelling and the PSI datum. The de-trended and bias corrected results (black crosses) are used in the validation process. The plots show that there are only few benchmarks with significant signal (deformation), and that the signal observed is just above the general noise level. Even though these plots suggest the absence of significant signal, from Figures 5.8 and 5.10 it is clear that the signal to noise level is sufficient to unambiguously detect the deformation signal of interest in the active areas.

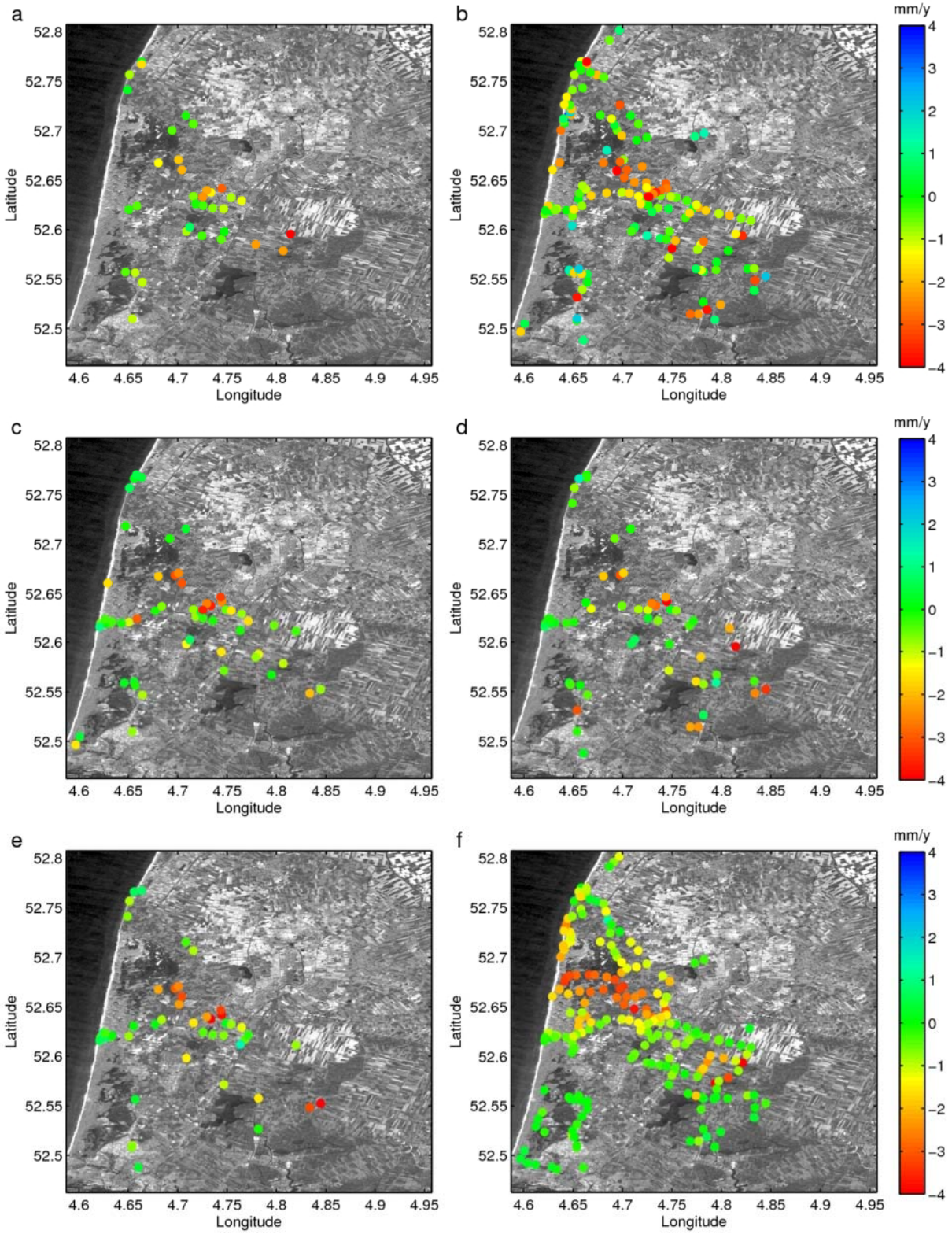

Figure 5.8: Alkmaar, ERS. Linear displacement velocity [mm/y] obtained from PSI at the levelling benchmark location where a nearest neighbor PS is available within a 50 m radius. Alkmaar-ERS time series (Apr. 1992 – Sep. 2000). a) - d) OSP results, e) TU Delft result, f) original linear displacement velocities [mm/y] from levelling.

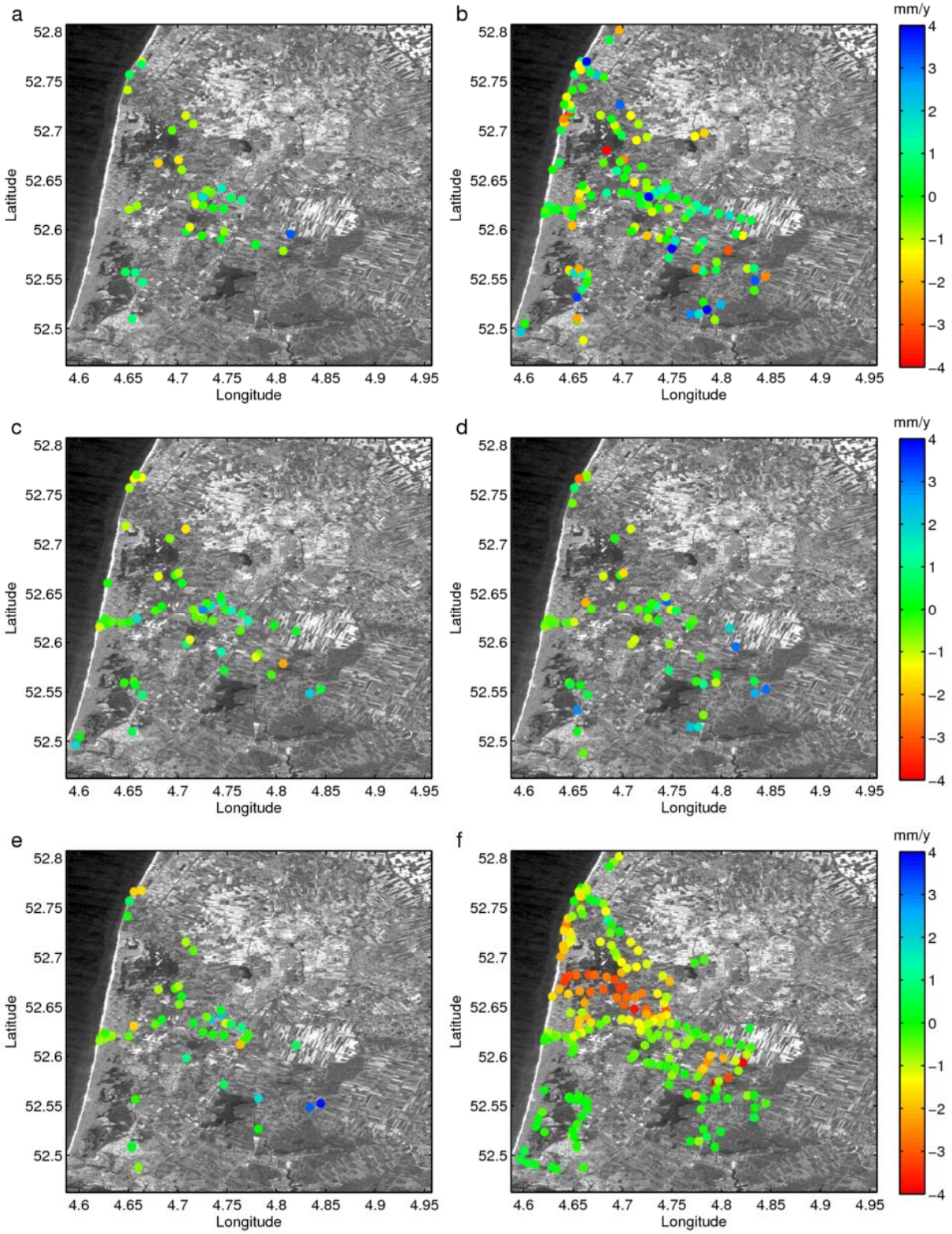

Figure 5.9: Alkmaar, ERS. Difference between linear displacement velocity [mm/y] obtained from PSI and levelling. Alkmaar-ERS time series (Apr. 1992 – Sep. 2000). a) - d) OSP results, e) TU Delft result, f) original linear displacement velocities [mm/y] from levelling.

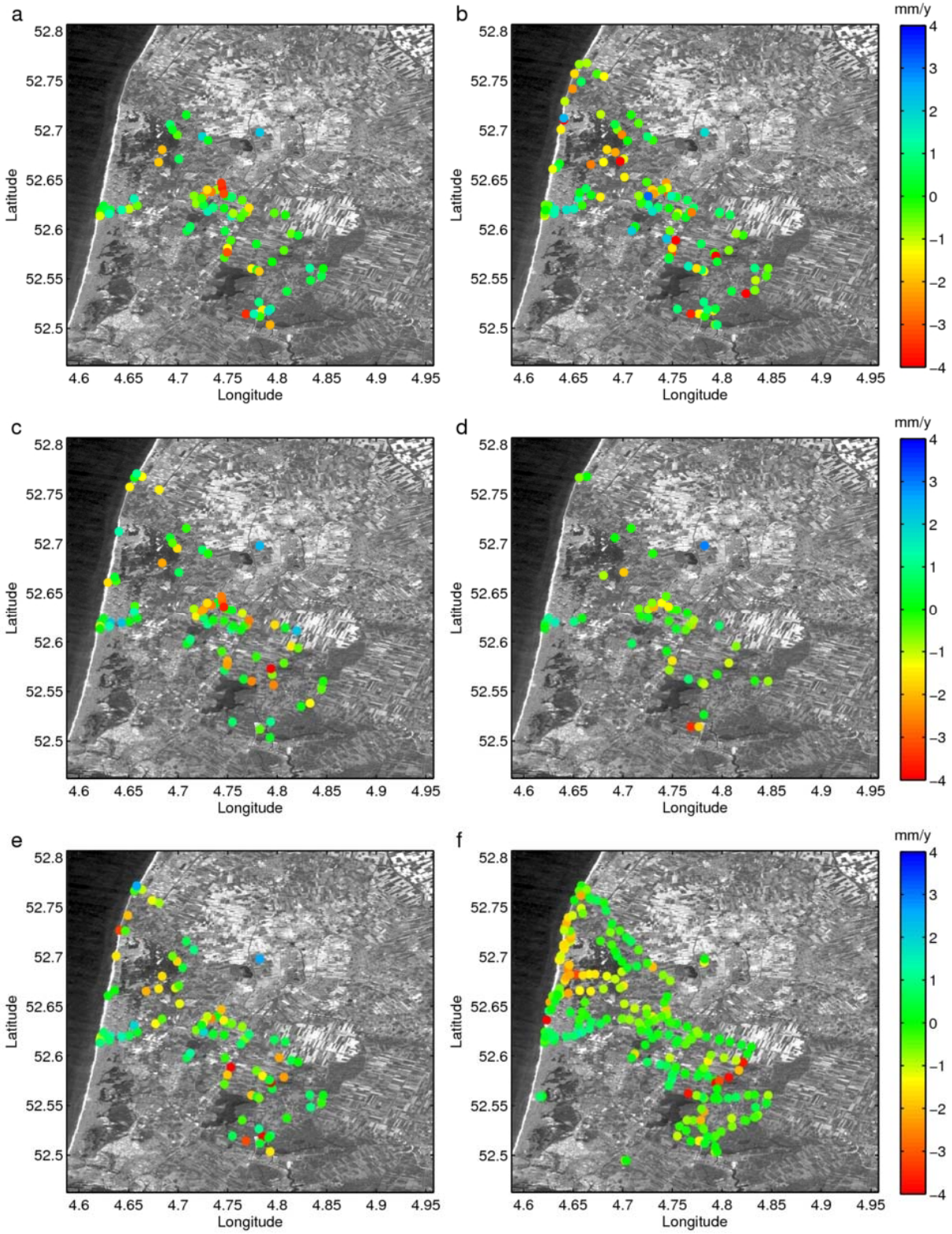

Figure 5.10: Alkmaar, Envisat. Linear displacement velocity [mm/y] obtained from PSI at the levelling benchmark location where a nearest neighbor PS is available within a 50 m radius. Alkmaar-Envisat time series (Mar. 2003 – Mar. 2007). a) - d) OSP results, e) TU Delft result, f) original linear displacement velocities [mm/y] from levelling.

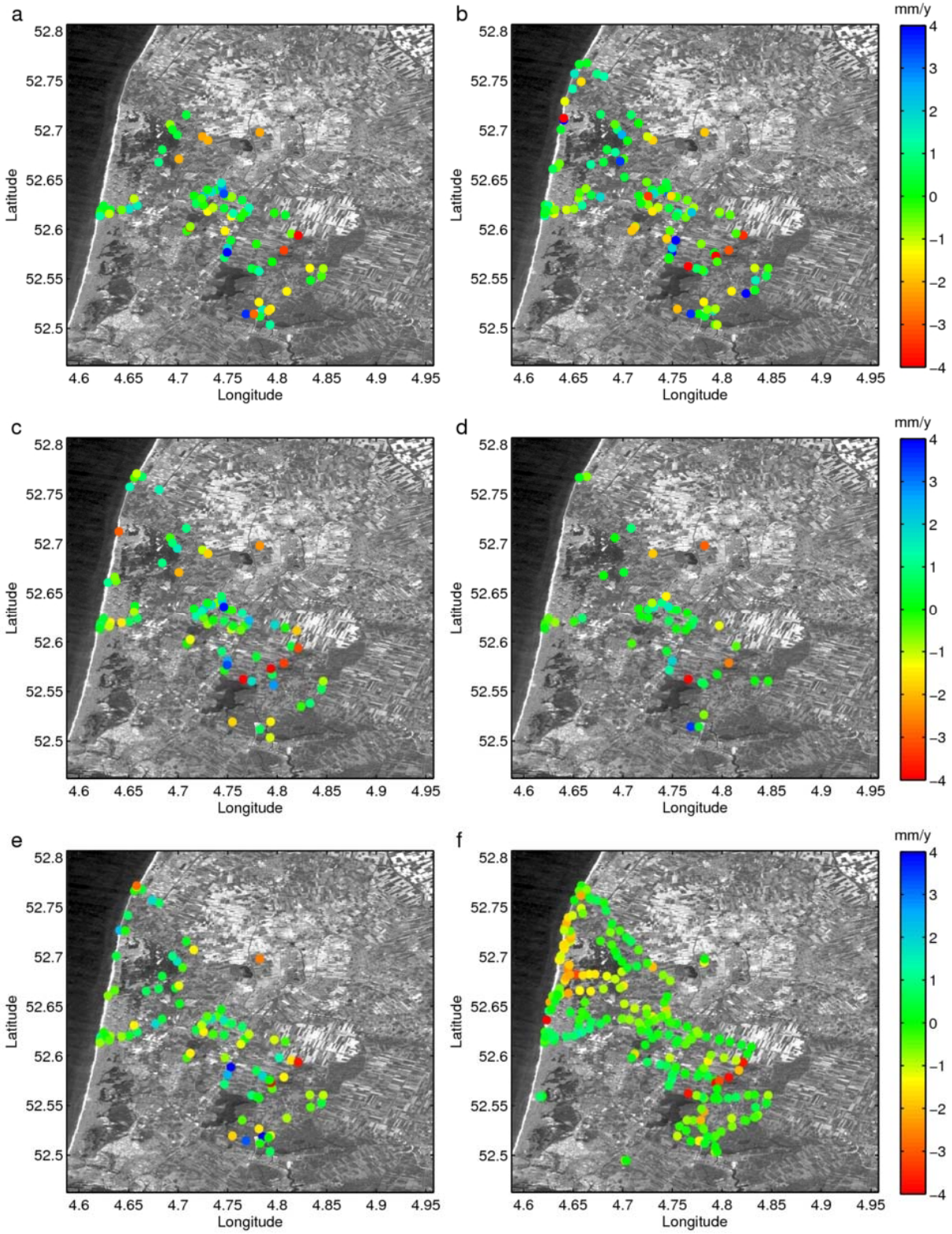

Figure 5.11: Alkmaar, Envisat. Difference between linear displacement velocity [mm/y] obtained from PSI and levelling. Alkmaar-Envisat time series (Mar. 2003 – Mar. 2007). a) - d) OSP results, e) TU Delft result, f) original linear displacement velocities [mm/y] from levelling.

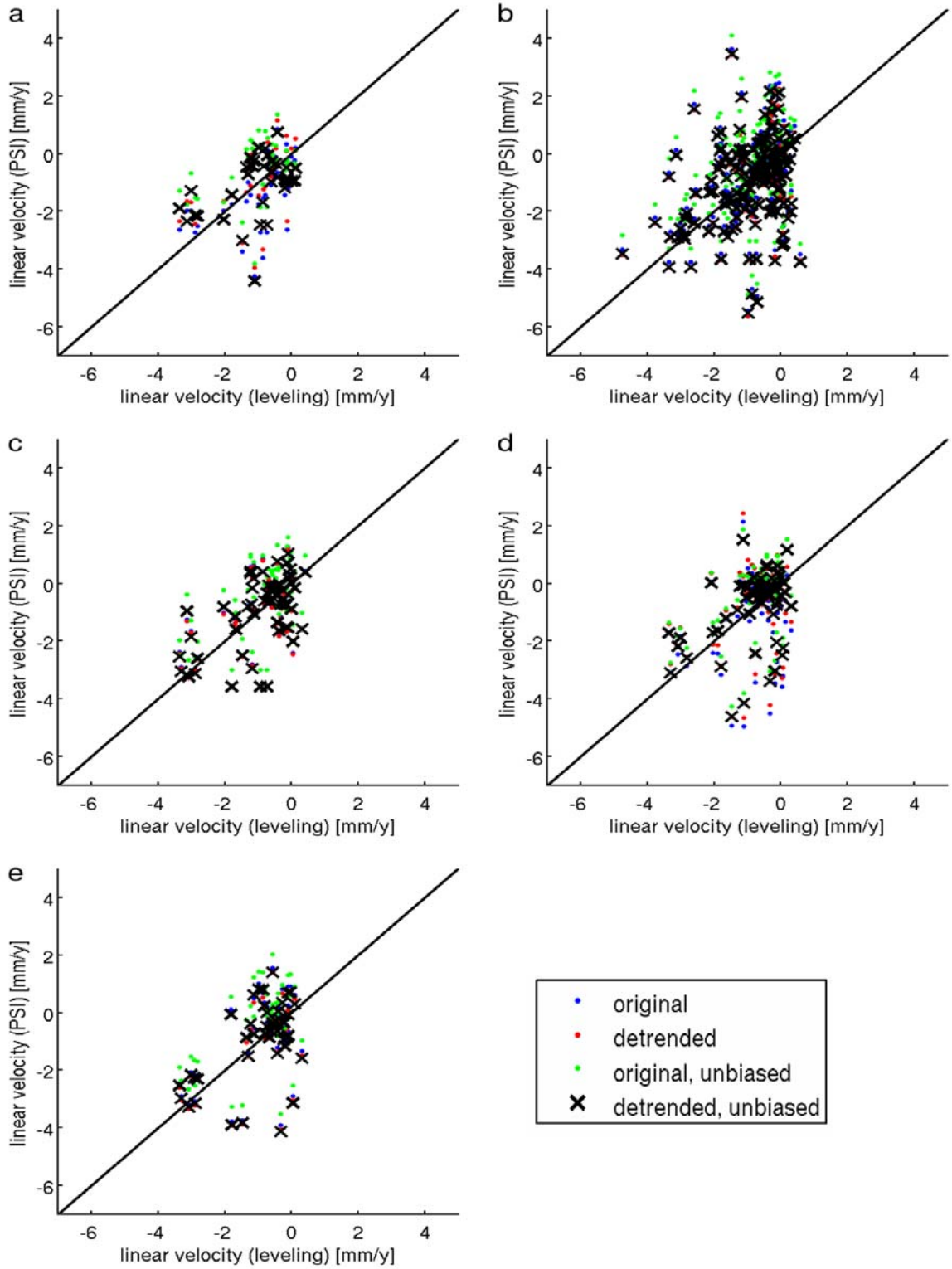

Figure 5.12: Alkmaar, ERS. Scatter plots of linear displacement velocity [mm/y] obtained from PSI and levelling. Alkmaar-ERS time series (Apr. 1992 – Sep. 2000). a) - d) OSP results, e) TU Delft result. The black crosses show the final PSI data set used for validation (de-trended, unbiased). The red, green and blue dots show the effect of de-trending and correction for the bias (datum connection). Note that the axes have a limited range, and that the relative absence of a clear elongated shape indicates that there is not much signal in the data (relatively stable areas). Therefore, these plots should be interpreted as representing the noise floor of PSI and levelling data.

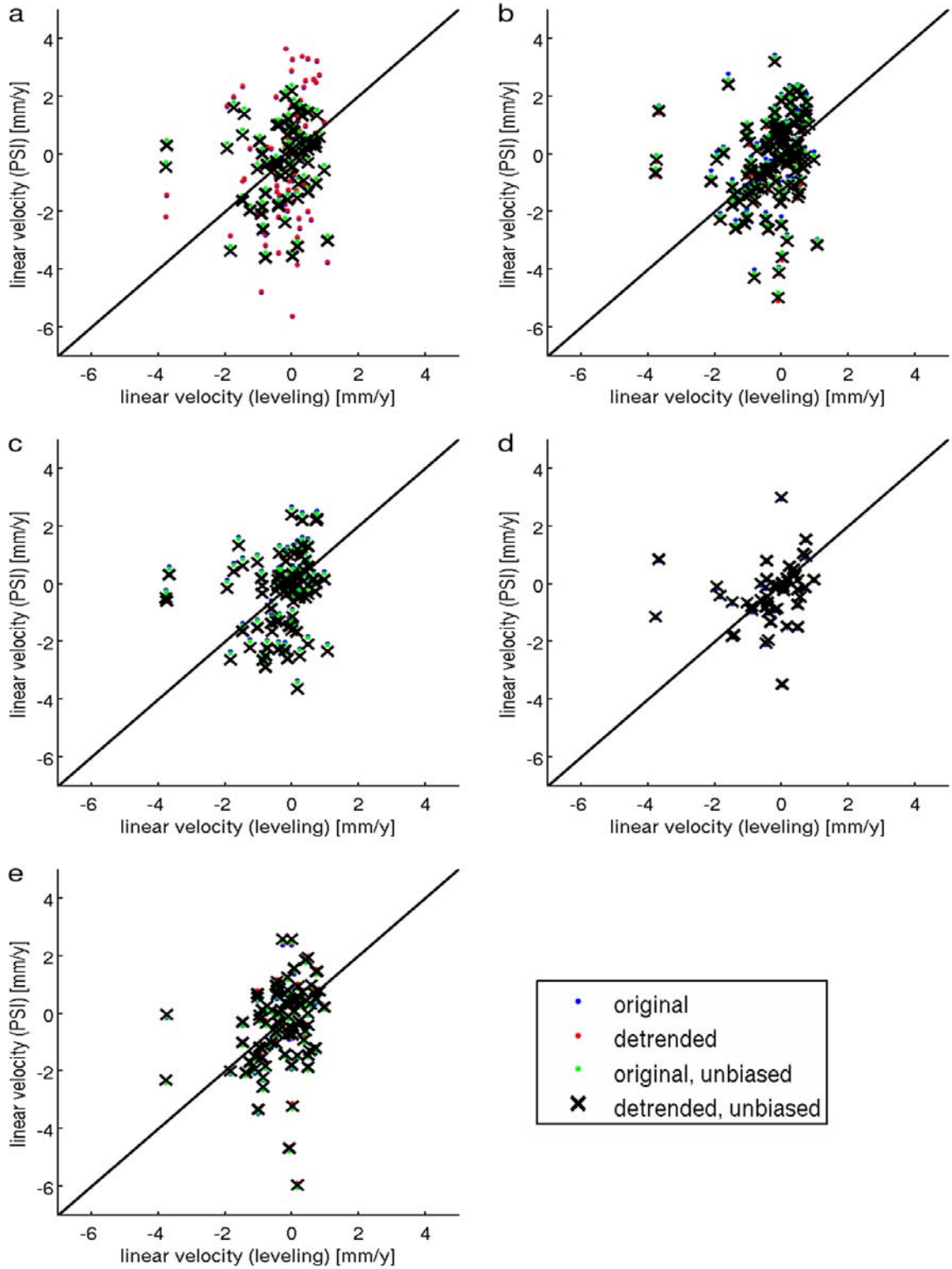

Figure 5.13: Alkmaar, Envisat. Scatter plots of linear displacement velocity [mm/y] obtained from PSI and levelling. Alkmaar- Envisat time series (Mar. 2003 – Mar. 2007). a) - d) OSP results, e) TU Delft result. The black crosses show the final PSI data set used for validation (de-trended, unbiased). The red, green and blue dots show the effect of de-trending and correction for the bias (datum connection). Note that the axes have a limited range, and that the relative absence of a clear elongated shape indicates that there is not much signal in the data. Therefore, these plots should be interpreted as representing the noise floor of PSI and levelling data.

The RMSE (Eq. 3, Section 4.2.2) of the differences between the PSI and levelling estimates are denoted in Table 5.1, together with the number of benchmarks per team that could be used for the validation. These results are obtained from the de-trended and unbiased data sets. Note that because of the removed bias (zero mean), the RMSE is equal to the standard deviation. The RMSE ranges from 1.0-1.5 mm/y for ERS and 1.3-1.8 mm/y for Envisat.

Table 5.1: RMSE of PSI versus levelling displacement velocities [mm/y].

	Team A	Team B	Team C	Team D	Team E
Alkmaar-ERS					
Number of benchmarks	36	151	58	58	47
RMSE PSI-levelling [mm/y]	1.07	1.54	1.04	1.23	1.18
Alkmaar-ASAR					
Number of benchmarks	76	118	80	49	90
RMSE PSI-levelling [mm/y]	1.51	1.79	1.63	1.26	1.54

To investigate whether there is a correlation between the RMSE of the difference between PSI and levelling and the quality measures of PSI, the RMSE is calculated per temporal coherence and spatio-temporal consistency class (see Chapter 6). Because the coherence values as supplied by the teams show a strong variation in distribution, an absolute classification is not possible. Therefore a relative classification is used, taking the best 20%, 40%, 60% and 80% of the PS per team. The results are given in Table 5.2. The results show no general trend. It changes per team whether there is a positive or negative correlation between the RMSE and coherence. The same applies for the classification according to the spatio-temporal consistency. These results can be found in Annex A.4. Here a relative as well as an absolute classification is used.

Table 5.2: RMSE of PSI versus levelling displacement velocities [mm/y] based on temporal coherence class.

	Team A	Team B	Team C	Team D	Team E
Alkmaar-ERS					
Top 20% coherence					
Number of benchmarks	3	24	3	9	7
RMSE PSI-levelling [mm/y]	0.96	1.51	1.05	0.63	1.09
Top 40% coherence					
Number of benchmarks	15	48	13	22	17
RMSE PSI-levelling [mm/y]	0.99	1.39	1.24	0.94	0.98
Top 60% coherence					
Number of benchmarks	21	71	27	34	19
RMSE PSI-levelling [mm/y]	1.00	1.41	1.06	1.20	1.28
Top 80% coherence					
Number of benchmarks	28	117	42	49	32
RMSE PSI-levelling [mm/y]	1.11	1.57	0.97	1.24	1.14
100% coherence					
Number of benchmarks	36	151	58	58	47
RMSE PSI-levelling [mm/y]	1.07	1.54	1.04	1.23	1.18
Alkmaar-ASAR					
Top 20% coherence					
Number of benchmarks	10	11	5	9	9
RMSE PSI-levelling [mm/y]	2.08	1.64	1.90	1.07	0.68
Top 40% coherence					
Number of benchmarks	24	26	20	19	23
RMSE PSI-levelling [mm/y]	1.54	1.73	2.21	1.30	0.92
Top 60% coherence					
Number of benchmarks	37	62	35	27	40
RMSE PSI-levelling [mm/y]	1.66	1.80	1.79	1.25	1.48
Top 80% coherence					
Number of benchmarks	57	97	63	36	67
RMSE PSI-levelling [mm/y]	1.54	1.80	1.73	1.37	1.66
100% coherence					
Number of benchmarks	76	118	80	49	90
RMSE PSI-levelling [mm/y]	1.51	1.79	1.63	1.26	1.54

5.2.2 Validation displacement time series

Apart from the estimated linear displacement rates, also the PSI time series, that is, displacements per epoch, are validated. Due to the sparse temporal sampling of levelling, the PSI time series are interpolated towards the levelling epochs[5]. An averaging window of 6 acquisitions (3 before and 3 after the levelling epoch) is applied to reduce the PSI noise. An example of the levelling measurements (black crosses) together with the nearest neighbor PSI time series of the different teams (colored lines) is shown in Figure 5.14. The PSI estimates at the levelling epochs are indicated with the colored crosses. In case the levelling epochs are outside the radar time span, a pseudo-levelling observation (gray crosses) is created based on the levelling trends (black line).

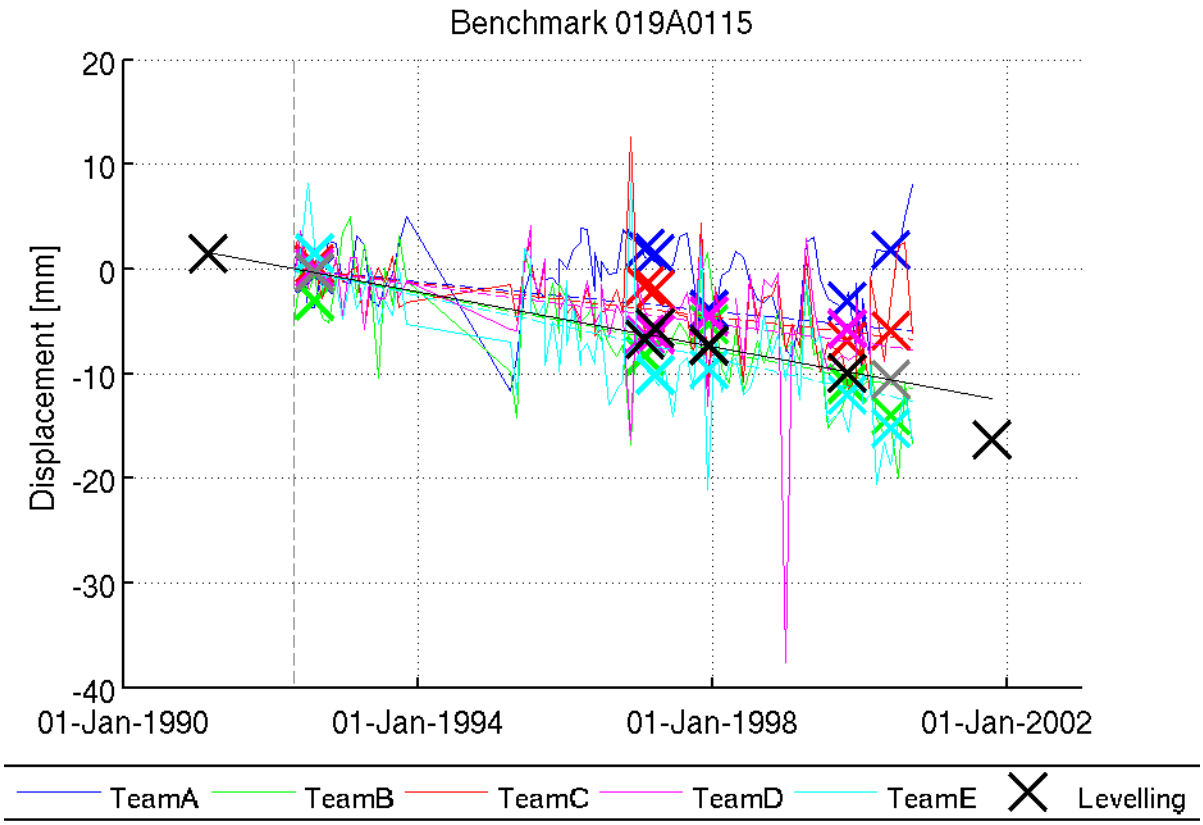

Figure 5.14: Example of levelling measurements (black crosses) at a certain benchmark, together with the nearest neighbor PSI time series of the different teams (colored lines). The PSI estimates at the levelling epochs are indicated with the colored crosses. In case the levelling epochs are outside the radar time span, a pseudo-levelling observation (gray crosses) is created based on the levelling trends (black line).

From Figure 5.14 it can be observed that the absolute offsets between the datums of the levelling and PSI time series is removed at the time of the first radar acquisition based on the estimated linear trends. This is however an arbitrary choice and the offset could for example also be removed in the middle of the time series. This choice will influence the comparison of

[5] Note that this procedure is different compared to the PSIC4 study, because there the temporal sampling of levelling epochs was higher, enabling interpolation towards the dates of radar acquisitions.

the time series. To circumvent this effect, the validation is based on the double differences between the levelling epochs (differences between PSI and levelling, between 2 epochs). The absolute offset is hereby eliminated.

The RMSE of the double differences are shown in Table 5.3, together with the number of double differences available for evaluation. The RMSE ranges from 4.2-5.9 mm for ERS and between 4.6 and 6.1 mm for Envisat. Hence, the precision of the ERS and Envisat time series is comparable. Note that in case of the Envisat time series typically only two levelling measurements are available and therefore only one double difference per benchmark.

Table 5.3: RMSE of PSI versus levelling time series (double difference) [mm].

	Team A	**Team B**	**Team C**	**Team D**	**Team E**
Alkmaar-ERS					
Number of double differences	122	485	209	192	166
RMSE PSI-levelling [mm]	4.98	5.90	4.16	5.59	5.45
Alkmaar-ASAR					
Number of double differences	78	121	81	50	92
RMSE PSI-levelling [mm]	4.67	5.53	6.11	4.58	5.49

As for the velocities, the RMSE is also calculated per quality class. The results can be found in Table 5.4. Again there is no clear correlation visible. The same holds for the classification according to the spatio-temporal consistency, which results are presented in Annex A.4.

Table 5.4: RMSE of PSI versus levelling displacement time series [mm] based on temporal coherence class.

	Team A	Team B	Team C	Team D	Team E
Alkmaar-ERS					
Top 20% coherence					
Number of double differences	6	79	11	31	28
RMSE PSI-levelling [mm]	3.22	5.73	4.16	4.86	4.83
Top 40% coherence					
Number of double differences	49	160	44	81	59
RMSE PSI-levelling [mm]	4.54	5.56	4.30	5.03	4.50
Top 60% coherence					
Number of double differences	69	236	105	110	67
RMSE PSI-levelling [mm]	4.63	5.58	3.68	5.47	5.59
Top 80% coherence					
Number of double differences	96	377	154	164	111
RMSE PSI-levelling [mm]	5.14	5.95	3.74	5.41	5.42
100% coherence					
Number of double differences	122	485	209	192	166
RMSE PSI-levelling [mm]	4.98	5.90	4.16	5.59	5.45
Alkmaar-ASAR					
Top 20% coherence					
Number of double differences	10	11	5	9	9
RMSE PSI-levelling [mm]	7.19	6.79	7.39	4.46	3.33
Top 40% coherence					
Number of double differences	25	27	20	20	25
RMSE PSI-levelling [mm]	5.10	6.47	10.05	3.75	2.93
Top 60% coherence					
Number of double differences	38	64	35	28	42
RMSE PSI-levelling [mm]	5.72	6.45	7.86	4.16	6.10
Top 80% coherence					
Number of double differences	58	99	64	37	69
RMSE PSI-levelling [mm]	5.02	5.67	6.55	4.83	5.81
100% coherence					
Number of double differences	78	121	81	50	92
RMSE PSI-levelling [mm]	4.67	5.53	6.11	4.58	5.49

5.3 Validation in the parameter space

The direct comparison of different types of measurements (levelling and PSI), as described in Section 5.2, has the disadvantage that results are only meaningful under the assumption that both techniques measure the same parameter at the same location and the same time. Usually this is not the case, as radar scatterers are typically not located on the same objects as levelling benchmarks. This hampers fair validation and data interpretation, and leads to an overestimation of the PSI error budget. Moreover, the most important aspect of the evaluation of the techniques — especially from an end-user point of view — is their ability to estimate the same source parameters. Source parameters depend on the type of parameterization chosen for the application. They could range from a geometric description of a subsidence bowl as a function of space and time, to a description of inflating/deflating volume sources at depth.

For this reason, here we perform a validation in the parameter space. The main advantage of this approach is that the influence of the physical difference in measurement techniques is reduced, and that the influence of reciprocal characteristics, such as point density versus estimation precision, are evaluated in concert. By estimating the same parameters from two different sets of data (levelling and radar) one can demonstrate whether that it is possible to reliably estimate the same parameters, even though the actual measurements are at totally different locations.

The applied parameterization uses the Mogi source model (see Anderson, 1936, Mogi, 1958) and is applied both to the PSI and levelling data. The Mogi sources have been placed in the reservoir locations, considering the location, dimensions and depth of the reservoir as a priori knowledge. For elongated or irregularly shaped reservoirs (see Figure 2.7 and Figure 2.9) 2-5 sources are placed. Both PSI and levelling data are considered to be uncorrelated and normal distributed, with $\sigma = 5$ and $\sigma = 1$ mm, respectively. The data are inverted in a Bayesian sense to derive source parameters (volume change) (Tarantola, 2005). It was assumed that the Mogi source parameters are correlated depending on distance, using correlation function $\exp(-r^2/\rho^2)$, where ρ = 1 km and r is the distance from the source center. Figure 5.15 shows the displacement field after modeling of the gas reservoirs for five PSI processing chain results and for the levelling data. The Mogi sources represent a volume change that can be directly related to the amount of gas extraction, thereby closing the gap between geodesy and geophysics. The results show ERS1/2 data, but similar results are obtained for Envisat. The most important conclusion from this approach is that despite the different nature, quality, and density of PSI and levelling measurements, similar displacement fields are estimated.

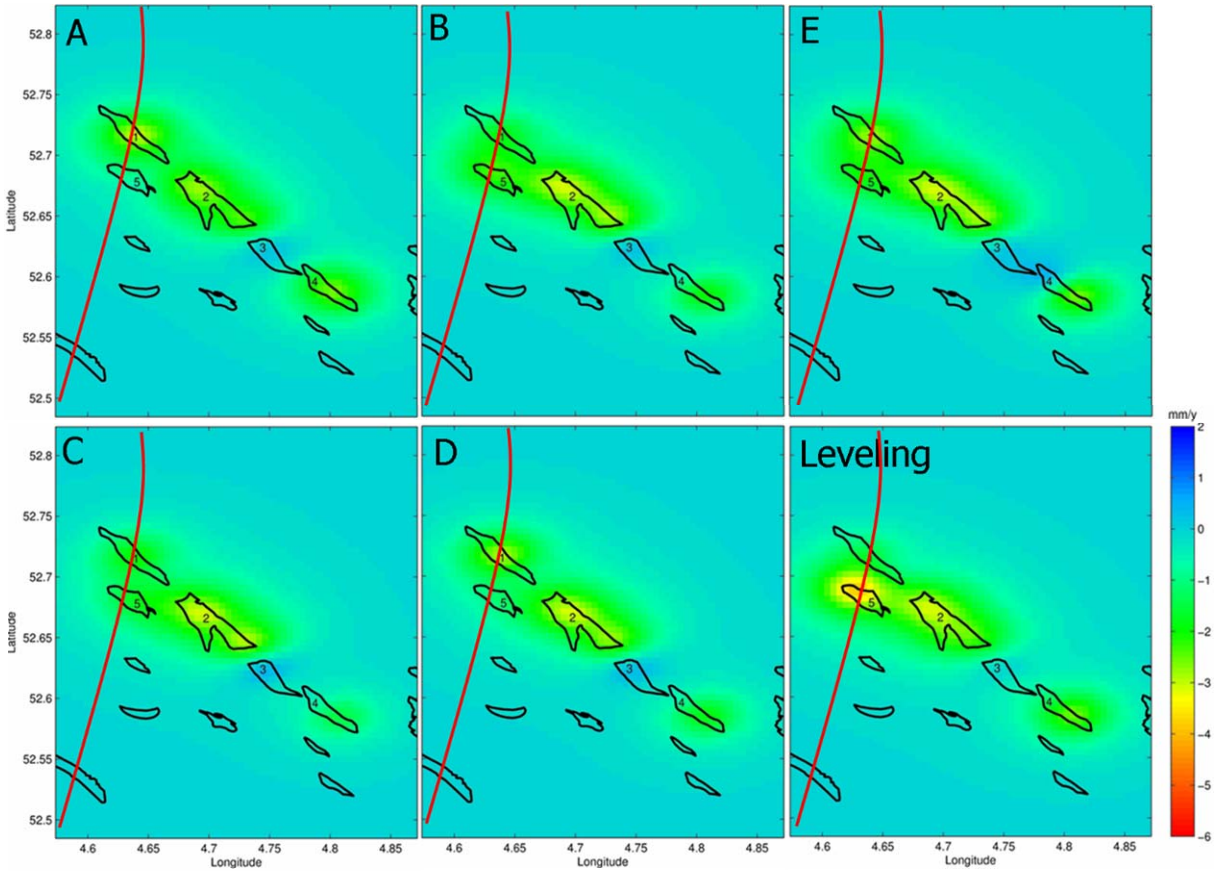

Figure 5.15: Alkmaar, ERS. Results of validation in the parameter space. Mogi sources have been placed in the reservoir locations 1-5, considering the location, dimensions and depth of the reservoir as a priori knowledge. PSI and levelling data are inverted to derive source parameters (volume change). The plots show the displacement field after modeling of the gas reservoirs by Mogi sources. The coast line is indicated in red. A-E are the results for five different processing chains. Note that the high rates for reservoirs observed by levelling are located offshore, and are likely to be extrapolation errors.

Values of the estimated volume change parameters are shown in Figure 5.16. Here the results for the five gas reservoirs, as numbered in Figure 5.15, are compared (arbitrary units). The dark blue bars are related based on the levelling data, whereas the other five colored bars represent the results for the different processing chains. The estimates for gas field 2, 3 (uplift) and 4 show a very good match, and the signal-to-noise ratio indicated by the error bars shows significant detection. The results for field 1 and 5 differ from the levelling, but show considerable internal consistency between the PSI processing chains. These gas fields are located partially offshore, which limits the spatial sampling of both techniques. The large error bars, especially for field 5, are mainly due to the unbalanced spatial sampling of measurements.

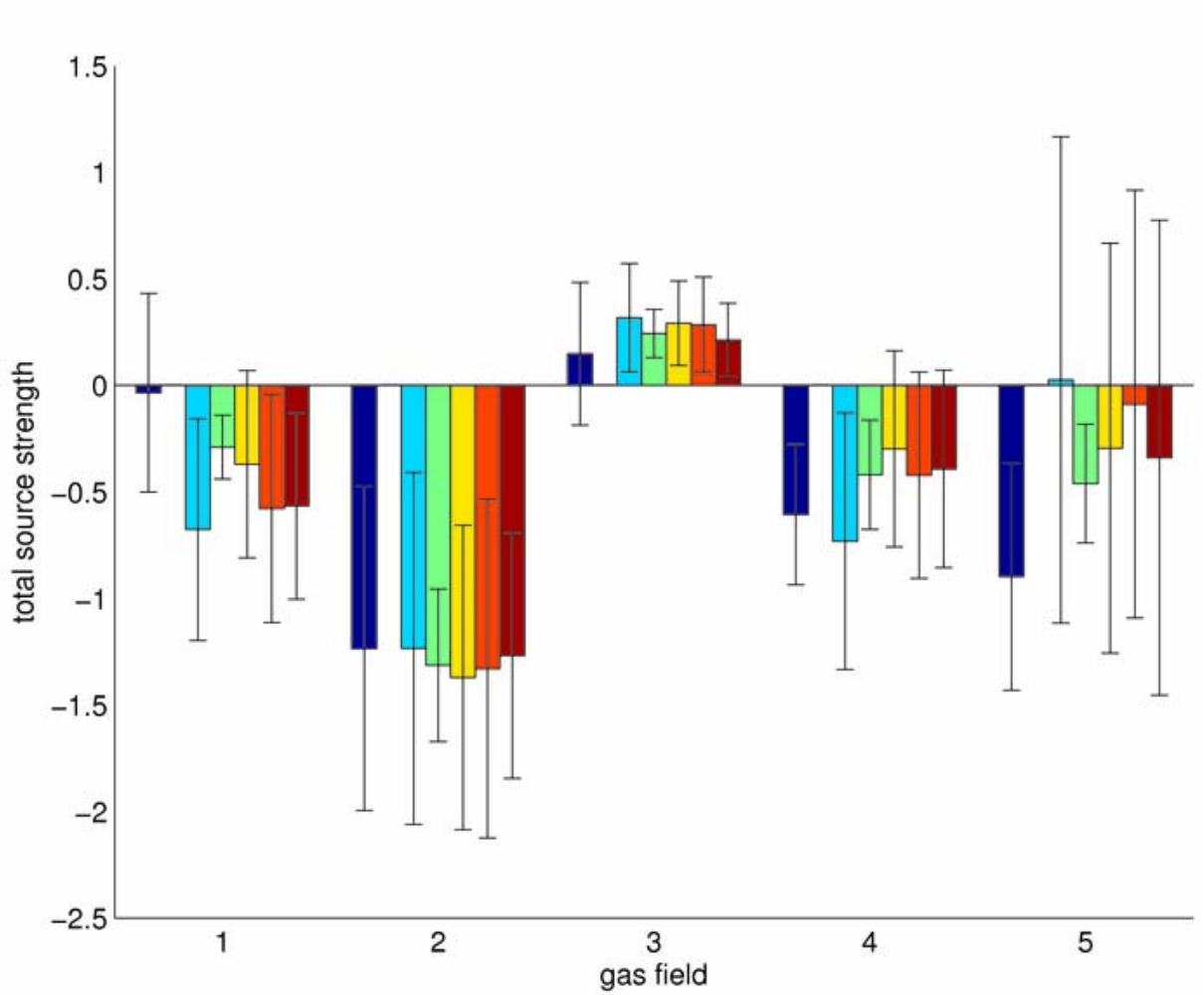

Figure 5.16: Comparison between total source strength (related to volume change) derived from levelling and PSI for different gas fields. Dark blue bars are based on levelling, and light blue, green, yellow, orange, red bars relate to teams A,B,C,D,E, respectively. The numbers of the gas fields on the horizontal axis correspond to the ones in Figure 5.15. The estimates for gas field 2, 3 and 4 show a very good match, whereas the results for field 1 and 5 differ. This is most likely due to the low density of PS in these areas (dunes) and the fact that both gas fields are largely situated below sea. The relatively large error bars for these estimates confirm these sub-optimal conditions for a reliable estimate.

These figures show that despite differences in PS density and quality among the various processing chains, they are all able to significantly estimate the signal of interest. This indicates that the sampling of PS, i.e., the actual locations of the PS, may be more important than the density of PS.

6 SPATIO-TEMPORAL CONSISTENCY

Coherence is a theoretically uniquely defined parameter, similar to, e.g., variance or standard deviation. This definition uses the expectation operator E{.}(Born and Wolf, 1980)

$$\gamma = \frac{E\{c_1 c_2^*\}}{\sqrt{E\{|c_1|^2\}E\{|c_2|^2\}}}, \tag{4}$$

where γ is the complex coherence and c1 and c2 are the complex radar observations for a pixel at a specific location. Unfortunately, as the expectation value of the signals c1 and c2 is unknown (there is only one observation at a given position and time) it is impossible to 'determine' the coherence; it can only be estimated. The coherence estimator $\hat{\gamma}$ for rezel (i,j) is generally defined as

$$\hat{\gamma} = \frac{\sum_R c_1(i,j)c_2^*(i,j)e^{-j\phi(i.j)}}{\sqrt{\sum_R |c_1(i,j)|^2 \sum_R |c_2(i,j)|^2}}, \tag{5}$$

where the summation over the range R is used as an estimate of the expectation value in eq. (1) The range R can be defined spatially, using samples in a window around rezel (i,j), or temporally, using subsequent observations of the same rezel in time. The latter is generally used in PSI applications. Both spatial and temporal coherence estimators rely heavily on the assumption of ergodicity, i.e., it is assumed that the stochastic properties of other observations in space and time are identical to the stochastic properties of the rezel under evaluation. It is evident that for (PS)InSAR, this is frequently not true, particularly for the phase component of the observations. This problem is mitigated in Eq. (5) by a (phase) correction factor $e^{-j\phi(i,j)}$, that adapts for the known phase behavior within range R. For the temporal coherence estimator, this implies that the phase of an arbitrary prior model (frequently a linear or steady-state model) is subtracted before the complex multiplication. Consequently, the value of the estimated temporal coherence is not only due to the quality (dispersion) of the interferometric observations, but also largely on the applied displacement model. In other words, it reflects not only measurement noise, but also model imperfections.

Example. Consider a group of noise-free PS, i.e., coherence is 1, that are all affected by the same, very non-linear, deformation history. Due to the non-linearity, the correction factor $e^{-j\phi(i,j)}$ is sub-optimal, which results in a lower temporal coherence estimate. Obviously, this is not dependent of the data quality as frequently assumed when using the term coherence, but only on model imperfections.

When considering this same group of PS, and evaluating their behavior relative to each other, no temporal model assumptions need to be made, and the dispersion between the observations would reflect measurement quality. This is the main idea behind the definition of Spatio-Temporal Consistency (STC) as alternative quality indicator, complementary to the temporal coherence estimator.

The spatio-temporal consistency of a PS is obtained by comparing its behavior in time and space with nearby PS. The nearby PS are selected between a minimum and a maximum distance. The minimum distance should avoid the selection of points which are related to the same physical scatterer, i.e., side lobes. The maximum range is set to avoid the influence of the low frequency deformation patterns on the quality estimate (which should only reflect measurement noise). In case of ERS/Envisat, a minimum range of 50 meters and a maximum range of 250 m is used. Once the nearby PS are identified for a certain PS, the double differences (in space and time) between the time series of the PS and the nearby PS are calculated. After calculation of the RMS error [mm] of these double differences, the minimum RMS error is assigned as the STC of the PS. Hence, the spatio-temporal consistency ρ is defined as

$$\rho = \min_{\forall p} \frac{\lambda}{4\pi} \sqrt{\frac{1}{N-1} \sum_{n=1}^{N-1} ((\phi_R^n - \phi_p^n) - (\phi_R^{n+1} - \phi_p^{n+1}))^2} , \qquad (6)$$

where p indicates the neighboring PS, R the PS under consideration, N the number of SAR acquisitions, λ the radar wavelength and ϕ the unwrapped phase in the time series. Taking the minimum RMS error, the influence of the spatial deformation pattern is assumed to be minimized, thereby best representing the measurement noise.

This procedure is repeated for each PS. When no nearby PS are present within the minimum-maximum range, no STC is assigned. Autonomous moving PS will get a high STC value. Therefore, this quality measure is especially suitable for spatially correlated deformation phenomena, as in the Alkmaar area.

Application of the algorithm, which can be performed on existing processing results and is therefore independent of the processing chains, results in the STC values which are presented in Figure 6.1 to Figure 6.3. Histograms of the estimates are shown in Figure 6.4. The results of team A show low values, which indicates a high spatio-temporal consistency. The most likely cause of this distinctive difference compared to the other teams is a stronger (a posteriori) filter on the time series. It can be noted that this quality indicator is discriminative, as results between different teams differ significantly. Compare these results with the coherence estimates earlier.

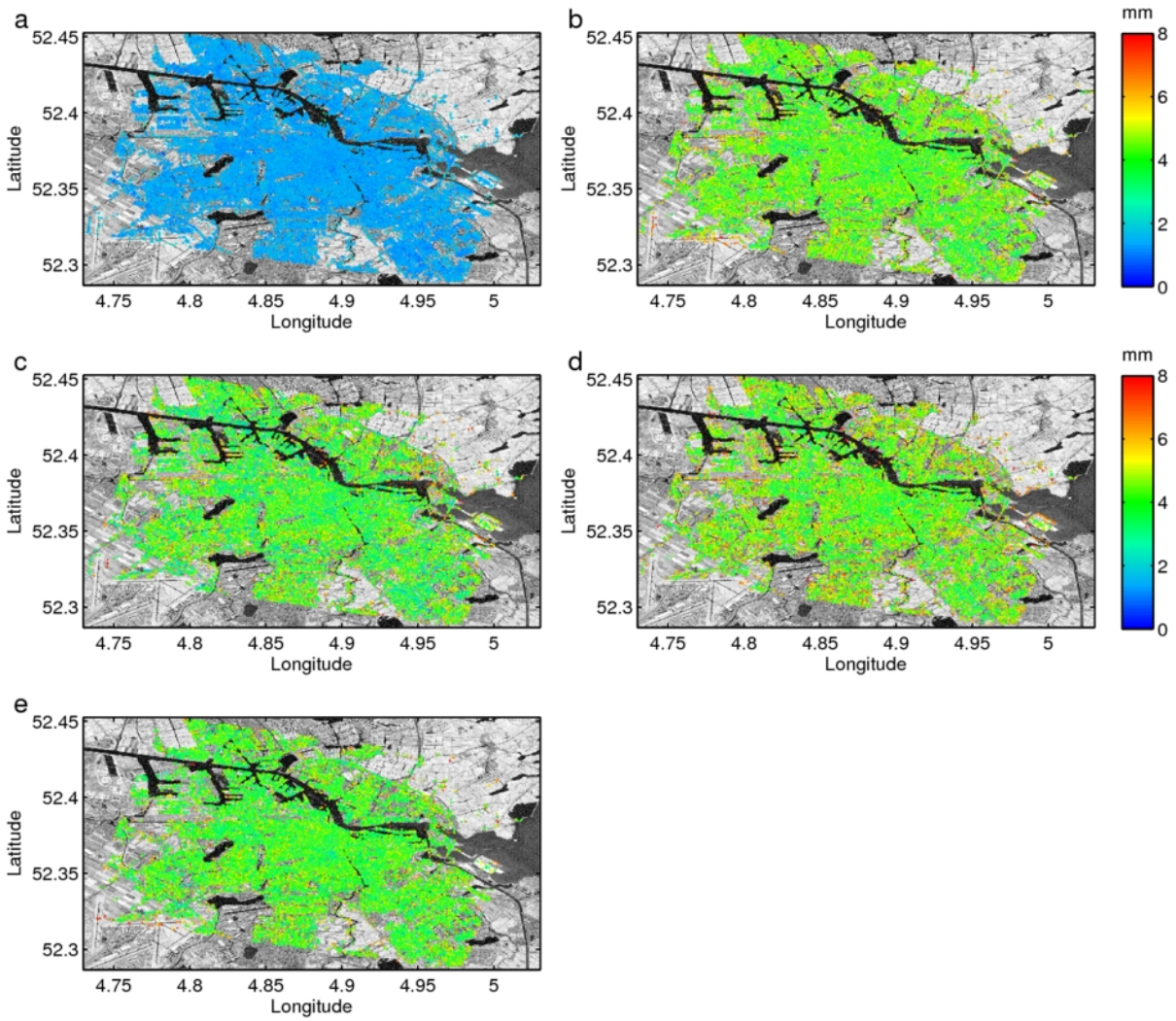

Figure 6.1: Amsterdam, Envisat. Spatio-temporal consistency (STC) [mm] in Amsterdam obtained from Envisat time series (Mar. 2003 – Mar. 2007). a) - d) OSP results, e) TU Delft result.

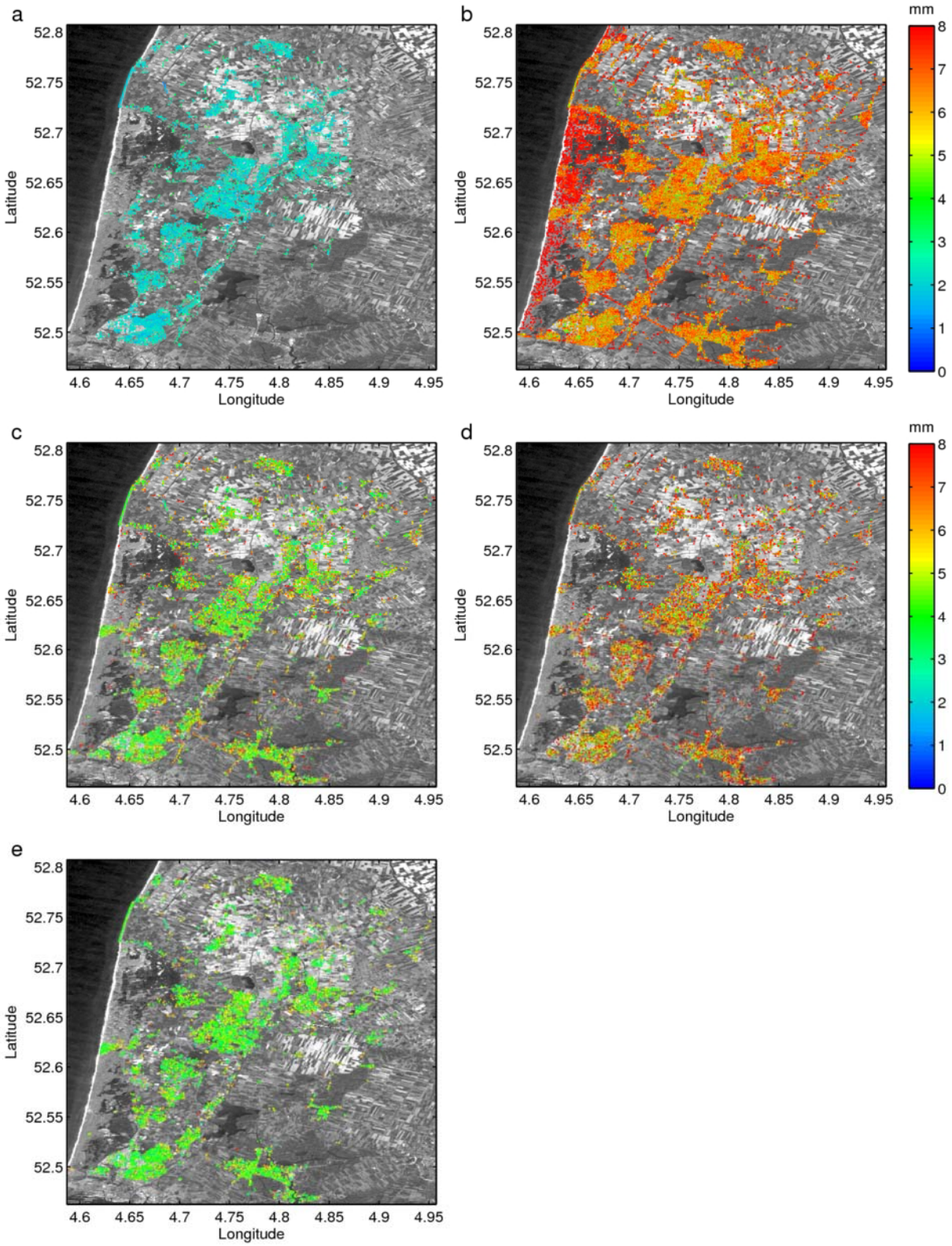

Figure 6.2: Alkmaar, ERS. Spatio-temporal consistency (STC) [mm] around Alkmaar obtained from ERS time series (Apr. 1992 – Sep. 2000). a) - d) OSP results, e) TU Delft result.

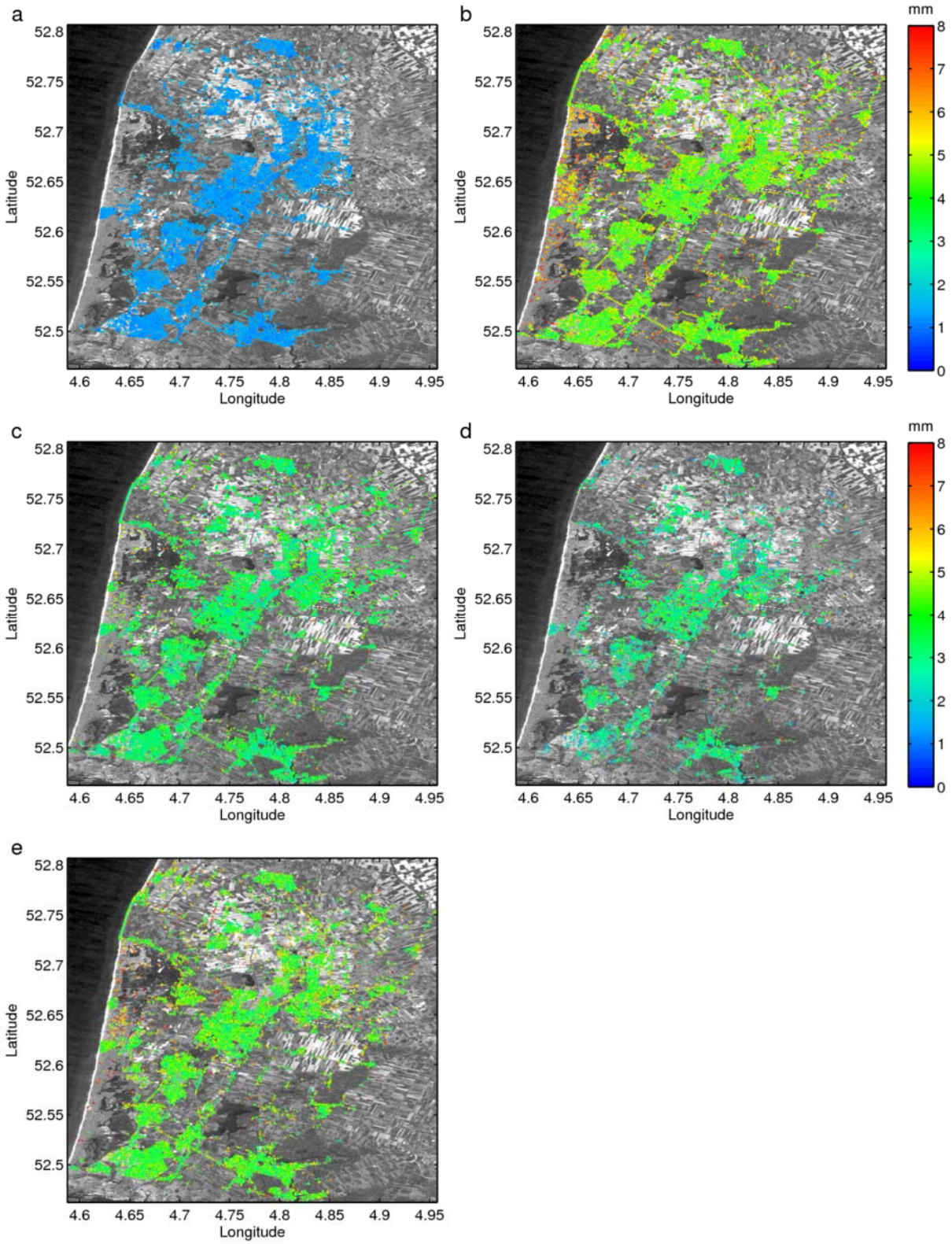

Figure 6.3: Alkmaar, Envisat. Spatio-temporal consistency (STC) [mm] around Alkmaar obtained from Envisat time series (Mar. 2003 – Mar. 2007). a) - d) OSP results, e) TU Delft result.

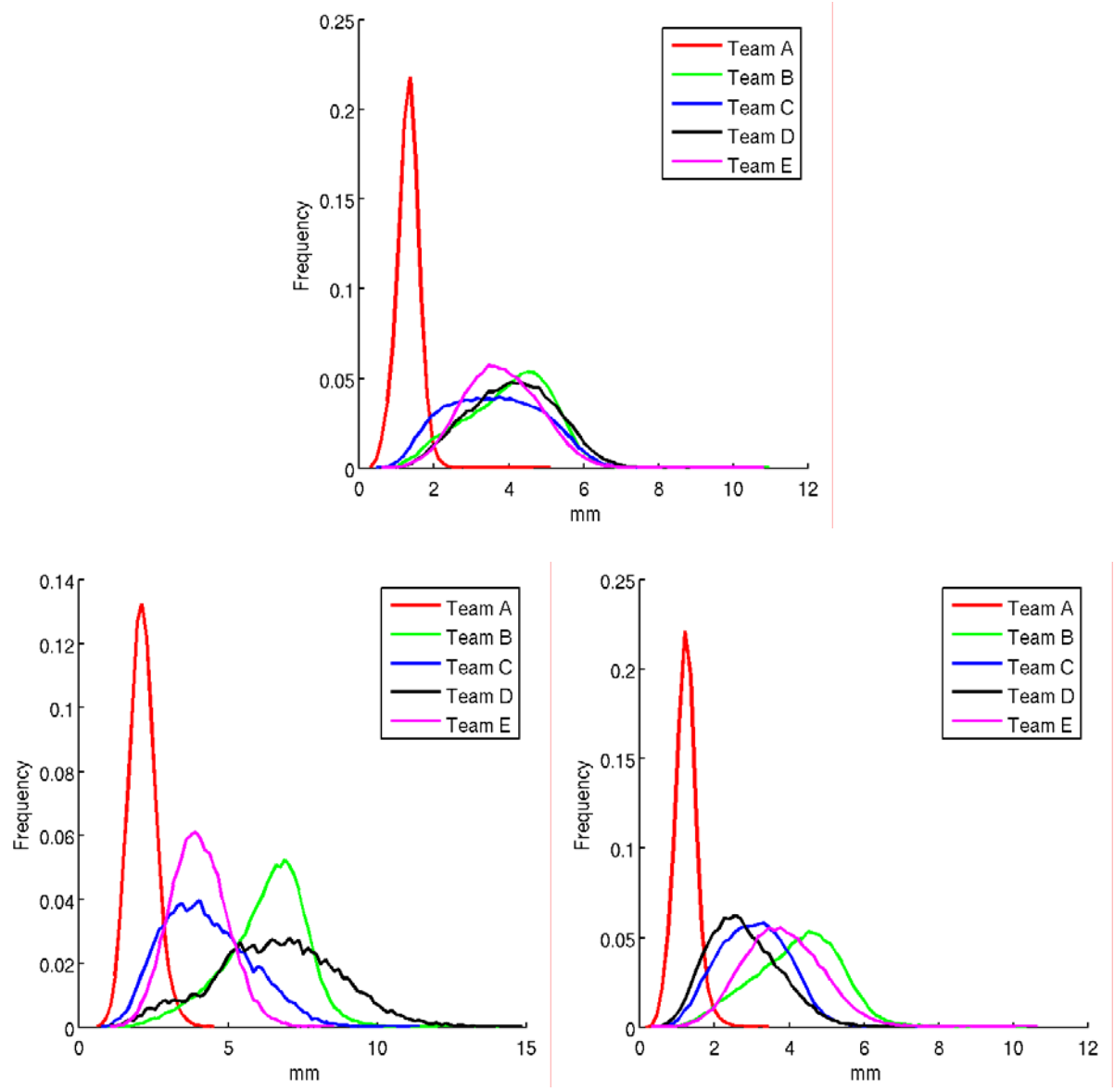

Figure 6.4: Normalized histograms of the spatio-temporal consistency (STC) [mm]. Above) Amsterdam, Envisat time series, left) Alkmaar, ERS time series, right) Alkmaar, Envisat time series.

7 CONCLUSIONS

7.1 Amsterdam

For Amsterdam, due to geo-coding errors (Section 4.1.1) it is not possible to make a perfect one-to-one comparison between scatterers and buildings. Therefore, in evaluating the results, intrinsic uncertainties due to for example geo-coding errors should be taken into account. Nevertheless, the main results for the Amsterdam case are as follows.

- Velocity validation.
  - The absolute standard deviation of the difference between PS velocity and tachymetry-based velocity ranges from 0.8 to 0.9 mm/yr.
  - The mean and median differences for all teams are close to zero.
  - All trend lines show a declination from the x=y axis, suggesting that PSI slightly underestimates deformation velocity with respect to tachymetry.
  - The absolute standard deviation of the double difference in velocity ranges from 1.0 to 1.2 mm/yr.

- Time series validation.
  - Average RMS errors of single deformation measurements range from 4.2 to 5.5 mm.

In general the PS data of all teams show a reasonably good correlation with the tachymetry data. However, all trend lines show a declination from the x=y axis, indicating a tendency for all teams to underestimate the deformation velocity with respect to data obtained by tachymetry in the higher velocity ranges. Furthermore, based on the current analysis, there is no significant difference in validation results between the four teams. All teams show similar results.

A suggestion for future work is to repeat the validation procedure when the construction of the N/S metro line is finished. It is expected that due to tunnel boring, which will commence in 2008, more buildings will be affected by deformation and also the amount of deformation will probably increase. This would result in a more evenly distributed velocity spectrum and a better comparison of velocities in the higher velocity range.

A second suggestion for future work is to evaluate the PSI algorithm and estimate the contribution of the various components like atmospheric phase screen, height-estimate and noise-component to the total error-budget.

7.2 Alkmaar

The main results for the Alkmaar case are as follows.

- Velocity validation.
  - After de-trending and removal of the bias between the PSI and the levelling datum, no systematic effects are found.
  - RMS error ranges from 1.0 – 1.5 mm/y for ERS and 1.3 – 1.8 mm/y for Envisat.

- Time series validation.
  - RMS error based on double differences (differences between PSI and levelling, and between measurement epochs) ranges from 4.2 – 5.9 mm for ERS and 4.6 – 6.1 mm for Envisat.

- Validation in the parameter space.
  - The approach overcomes the important intrinsic limitation of PSI validation, i.e. the fact that PSI and levelling do not measure the same point.
  - The analysis highlights a key PSI capability: the high number of available samples.
  - Even though the deformation signal is rather weak, the PSI vs. levelling comparison provides good results.
  - Even teams that have lower spatial point density have good results. This stresses the fact that it is not the absolute point density, but rather the sampling locations in relation to the deformation phenomenon that matters.

The main conclusion is that the results from PSI and levelling are comparable, with an RSME of 1.0 – 1.8 mm/y for the linear velocity rates. Although the scatter plots of the PSI results versus levelling suggests the absence of a signal when evaluating all observations, the plots of the spatial distribution of the PSI results indicate that the signal-to-noise ratio is sufficient to unambiguously detect the deformation signal of interest. The conclusion is confirmed by the validation in the parameter space. The PSI data of all teams enable the estimation of the signal of interest, despite the difference in PS density.

For future validation projects the availability of a profound quality description of the ground truth data is desirable. This would enable a more clear separation between the error budgets of PSI and ground truth. Moreover, the large variation in the distribution of the provided quality measured by the teams, in this case the temporal coherence values, is undesirable from an end-user perspective. Clarification of the cause of this variation would be a good step ahead. Possibly alternative quality measures, such as the spatio-temporal consistency, could be used as well.

8 REFERENCES

- Adam, N. and A. Parizzi (2007). Specification of validation approach. Part1: Process validation. GMES, TerraFirma, ESRIN/Contract no. 19366/05/I-E.
- Adam, N. and A. Parizzi (2008). Validation of existing processing chains in TerraFirma stage2: Process analysis report – Part 1: DLR process validation. GMES, TerraFirma, ESRIN/Contract no. 19366/05/I-E.
- Anderson, E.M.(1936). The dynamics of the formation of cone-sheets, ring-dykes, and caldron subsidence. *Proceedings of the Royal Society of Edinburgh 56*, 128-157.
- Born, M. and E. Wolf (1980). *Electromagnetic Theory of Propagation Interference and Diffraction of Light*. Pergamon Press, New York.
- Alessandro Ferretti, Claudio Prati, and Fabio Rocca. Permanent scatterers in SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 39(1):8-20, January 2001.
- Haak, H.W., B. Dost, and F. H. Goutbeek (November 2001). Seismische analyse van de aardbevingen bij Alkmaar op 9 en 10 september en Bergen aan Zee op 10 oktober 2001. KNMI Technical report TR-239. Technical report, KNMI.
- Mogi, K. (1958). Relations between eruptions of various volcanoes and the deformations of the ground surfaces around them. *Bulletin of the Earthquake Research Institute, University of Tokyo 36*, 99-134.
- Shearer, J.W. (1990). Accuracy of digital terrain models. In *Terrain Modelling in Surveying and Civil Engineering,* Petrie, G, Kennie, T.J.M., Eds., Thomas Telford, London, pp. 315-336.
- Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. SIAM, Philadelphia.
- Winningsplan Middelie area, dd. 15-11-2005.

A ANNEX

A.1 Results Amsterdam Envisat

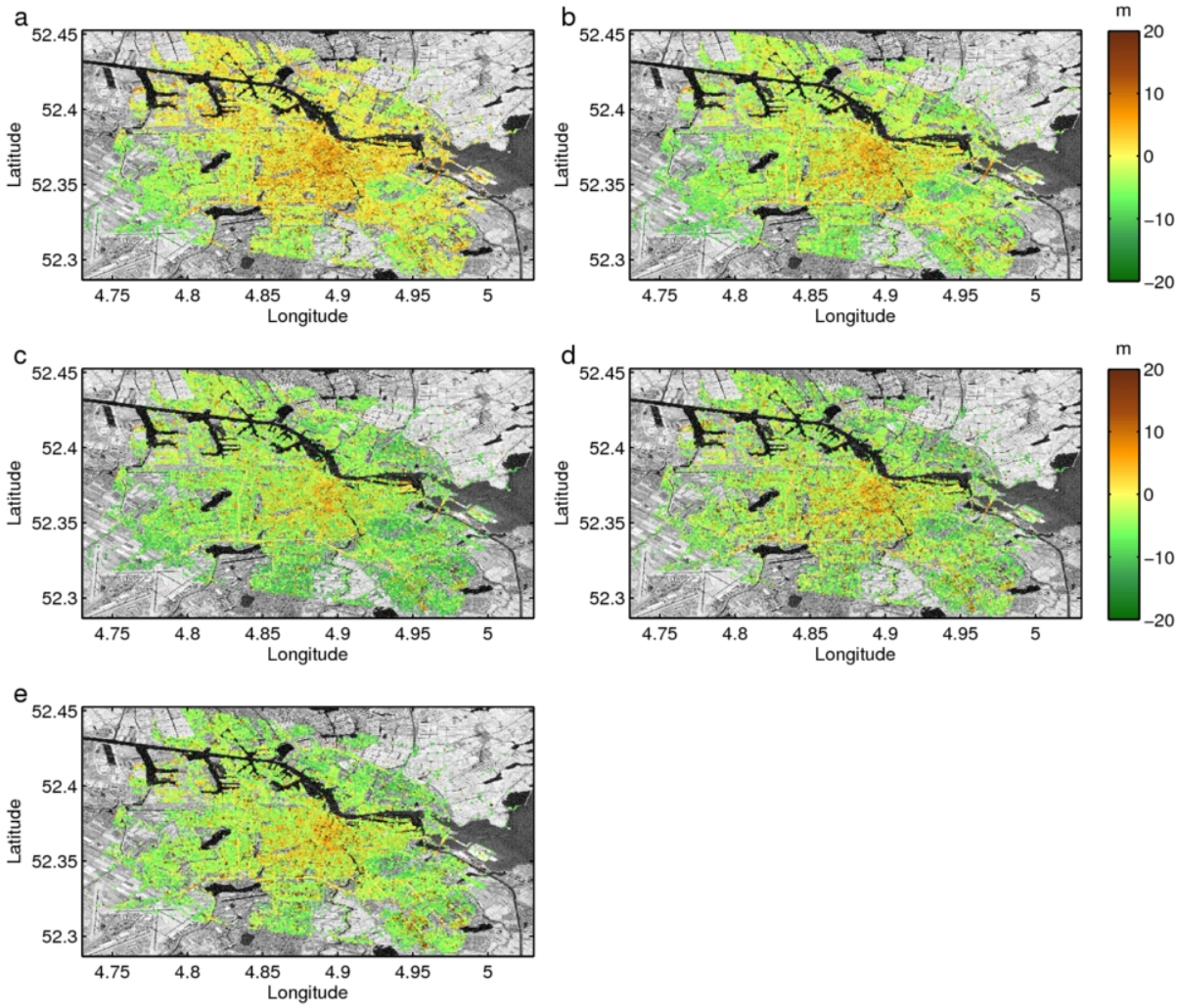

Figure A. 1: Amsterdam, Envisat. Relative height [m] in Amsterdam obtained from Envisat time series (Mar. 2003 – Mar. 2007). a) - d) OSP results, e) TU Delft result.

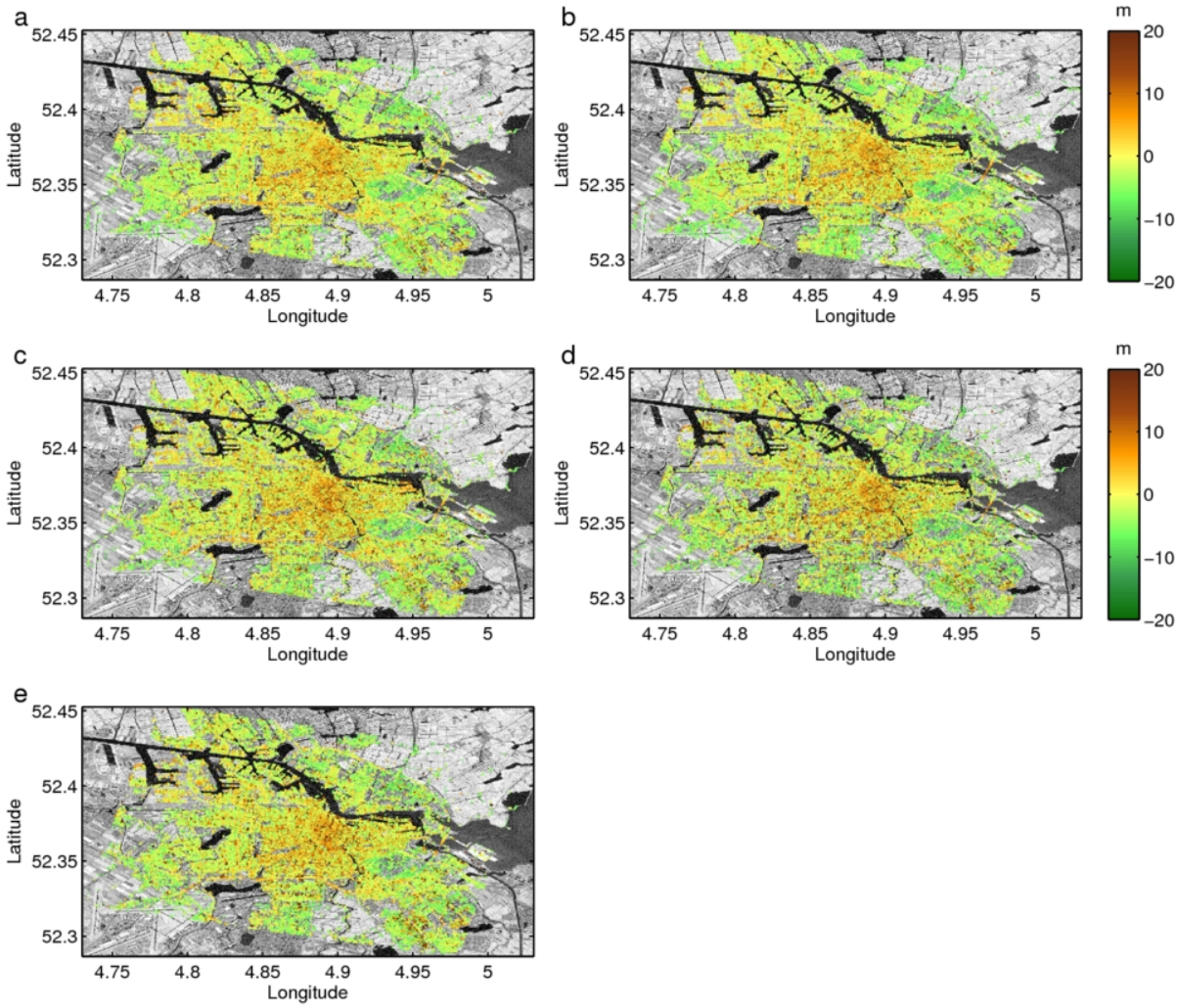

Figure A. 2: Amsterdam, Envisat. De-trended relative height [m] in Amsterdam obtained from Envisat time series (Mar. 2003 – Mar. 2007). a) - d) OSP results, e) TU Delft result.

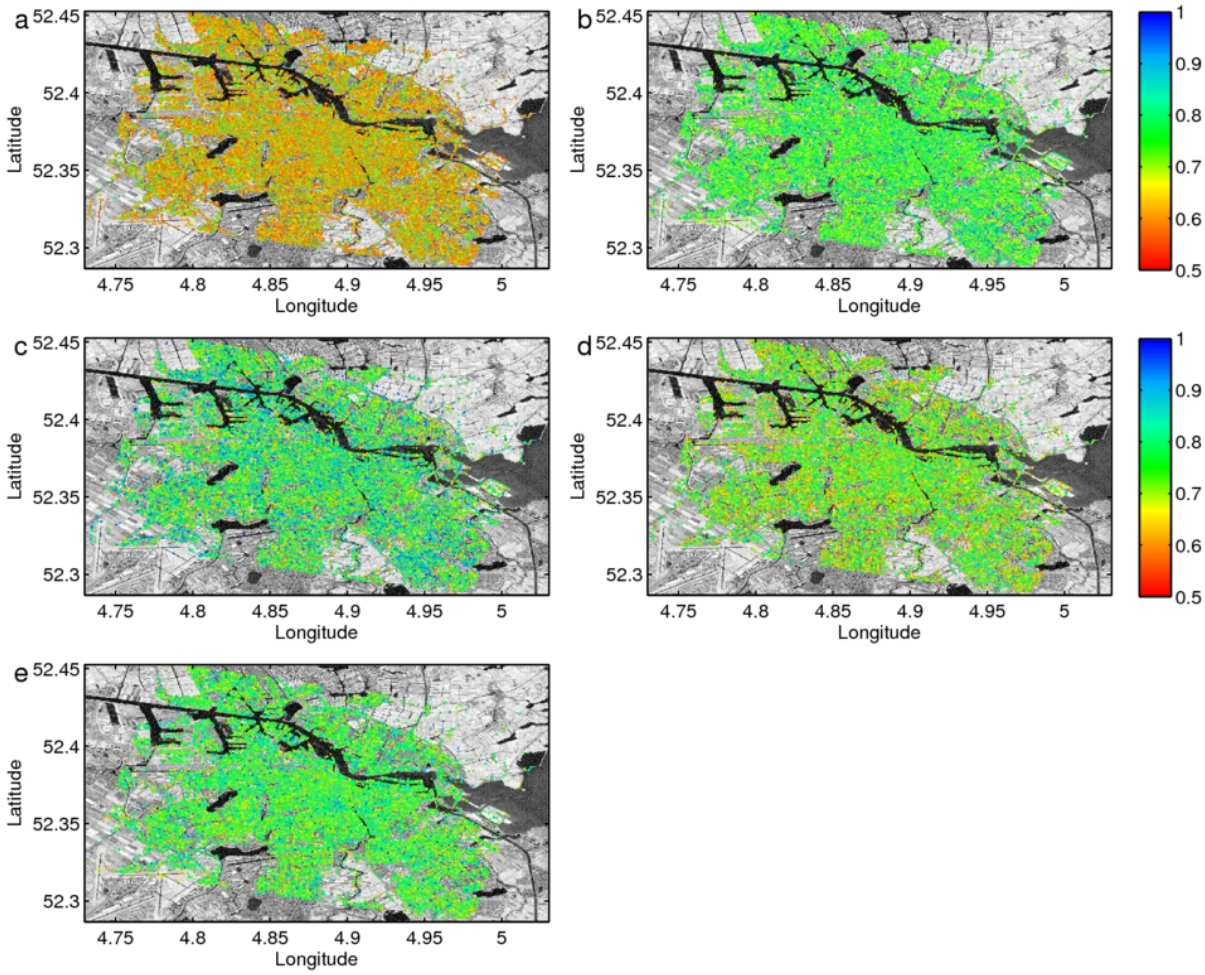

Figure A. 3: Amsterdam, Envisat. Temporal coherence estimates in Amsterdam for Envisat time series (Mar. 2003 – Mar. 2007). a) - d) OSP results, e) TU Delft result.

A.2 Results Alkmaar ERS

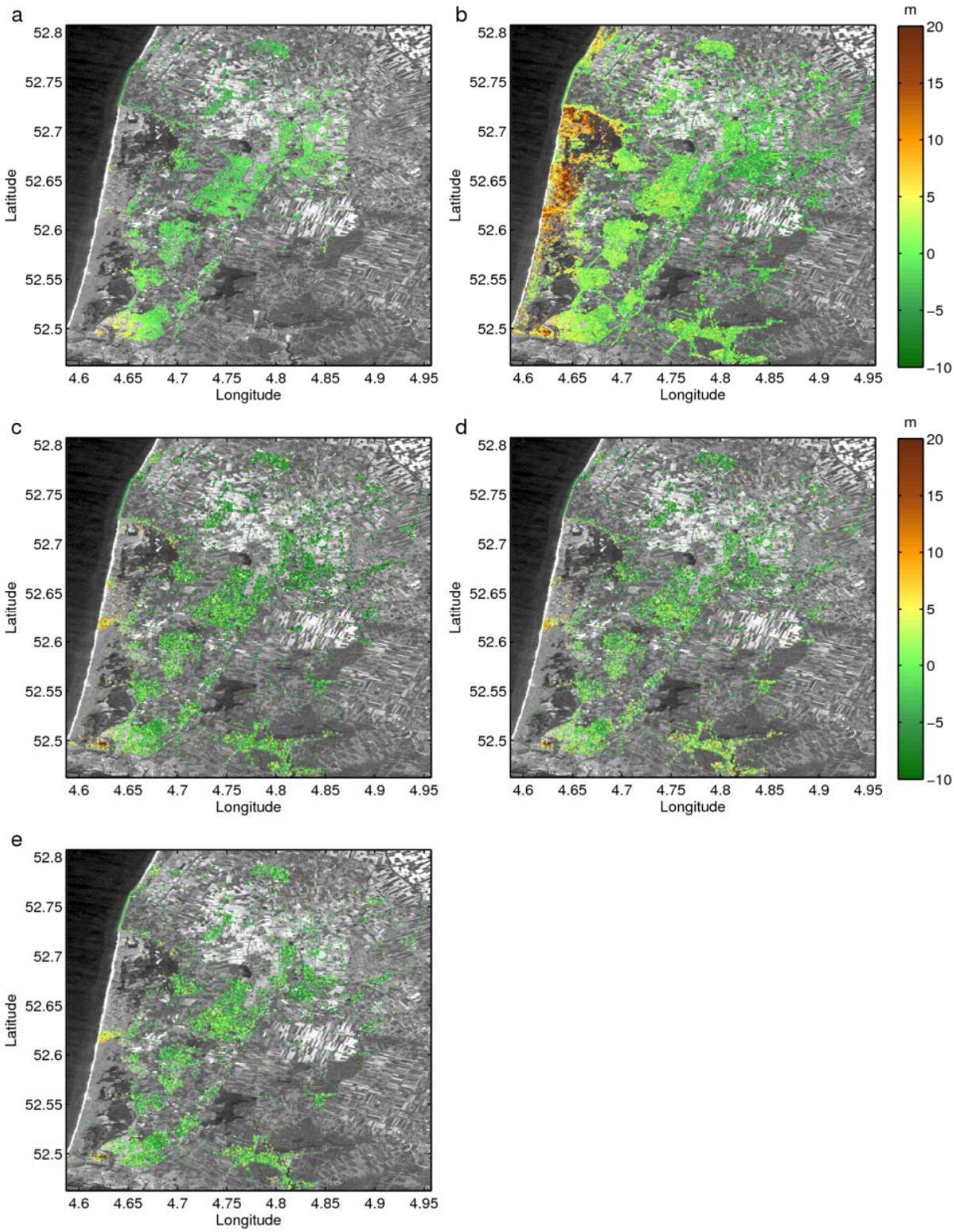

Figure A. 4: Alkmaar, ERS. Relative height [m] around Alkmaar obtained from ERS time series (Apr. 1992 – Sep. 2000). a) - d) OSP results, e) TU Delft result.

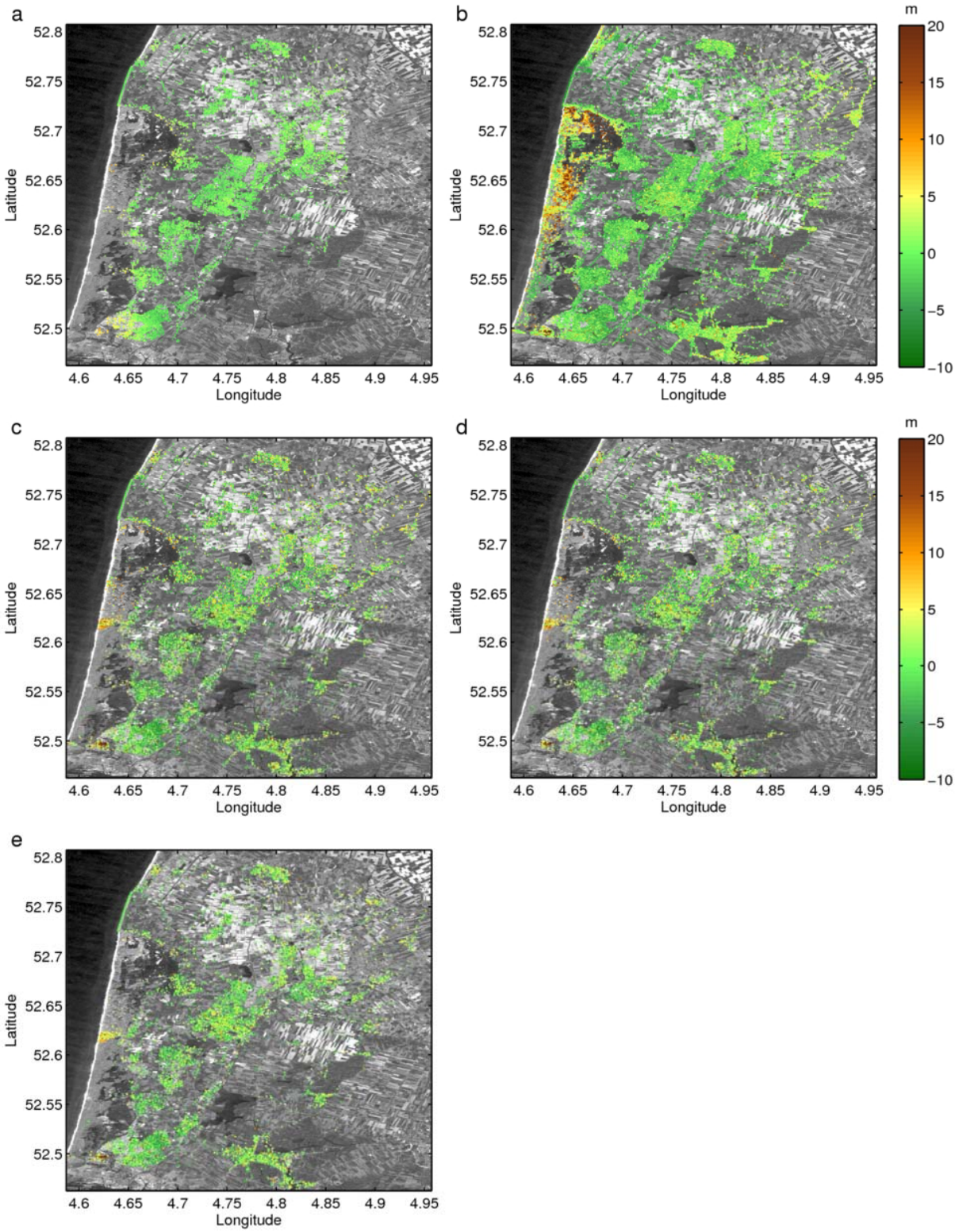

Figure A. 5: Alkmaar, ERS. De-trended relative height [m] around Alkmaar obtained from ERS time series (Apr. 1992 – Sep. 2000). a) - d) OSP results, e) TU Delft result.

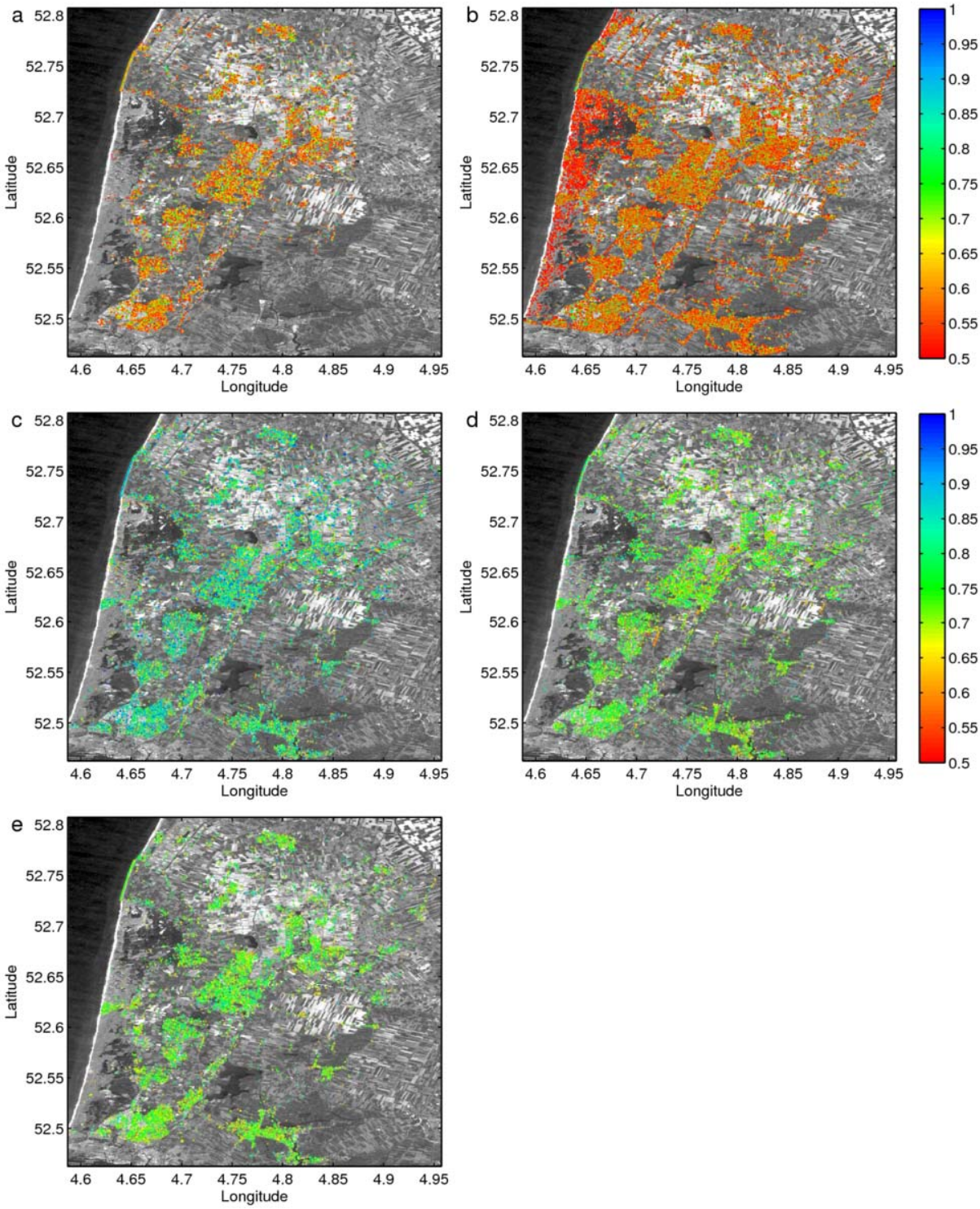

Figure A. 6: Alkmaar, ERS. Temporal coherence estimates around Alkmaar for ERS time series (Apr. 1992 – Sep. 2000). a) - d) OSP results, e) TU Delft result.

A.3 Results Alkmaar Envisat

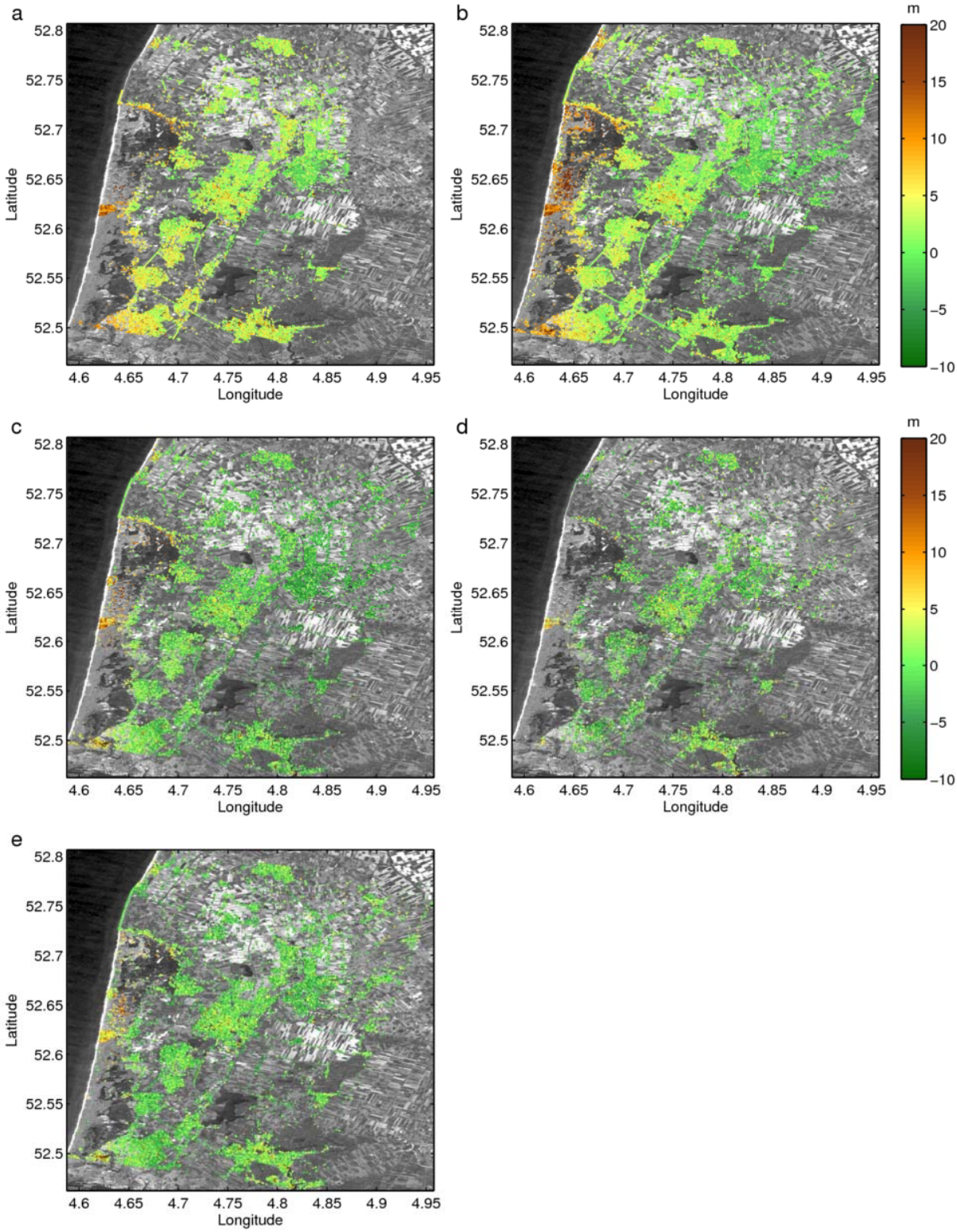

Figure A. 7: Alkmaar, Envisat. Relative height [m] around Alkmaar obtained from Envisat time series (Mar. 2003 – Mar. 2007). a) - d) OSP results, e) TU Delft result.

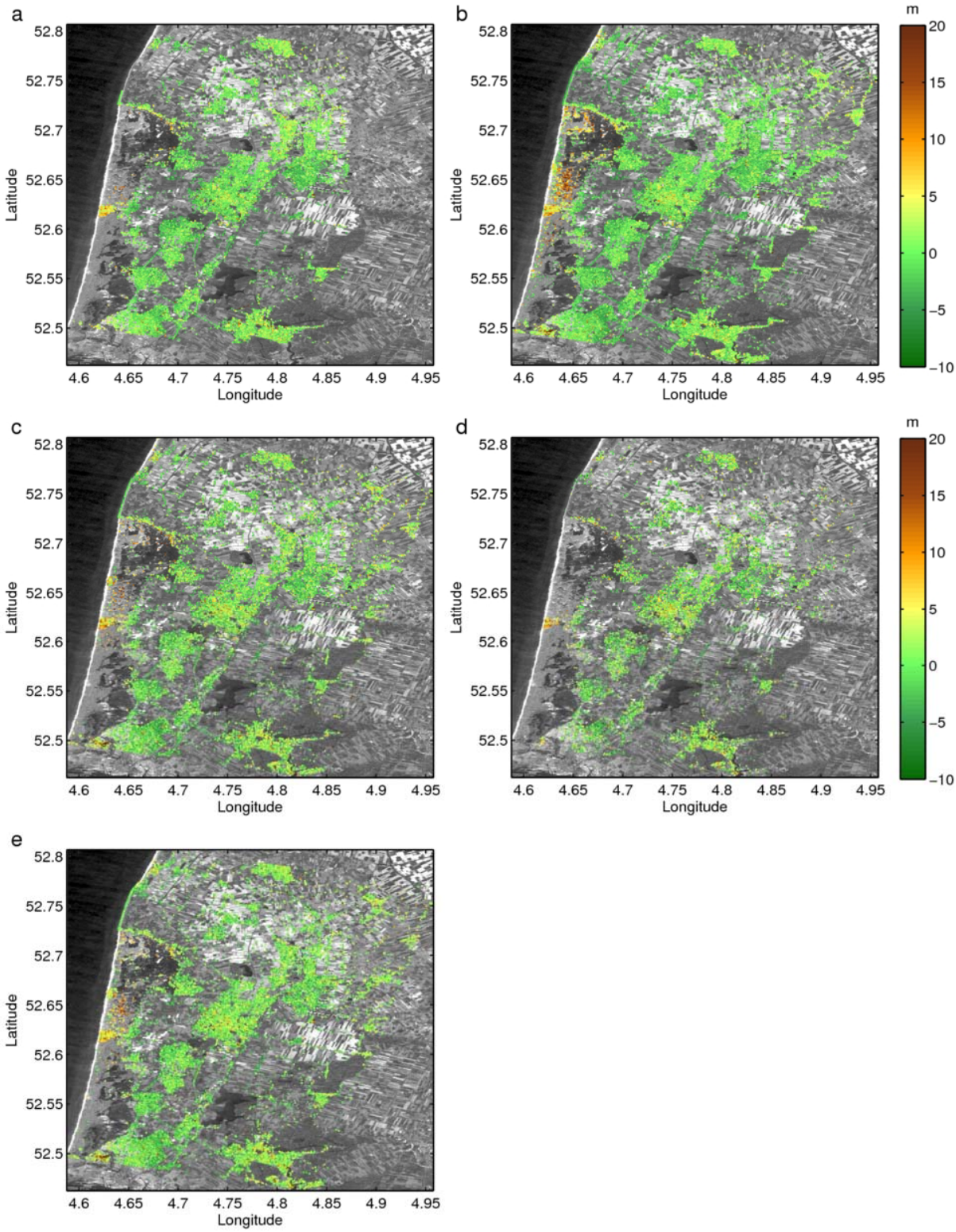

Figure A. 8: Alkmaar, Envisat. De-trended relative height [m] around Alkmaar obtained from Envisat time series (Mar. 2003 – Mar. 2007). a) - d) OSP results, e) TU Delft result.

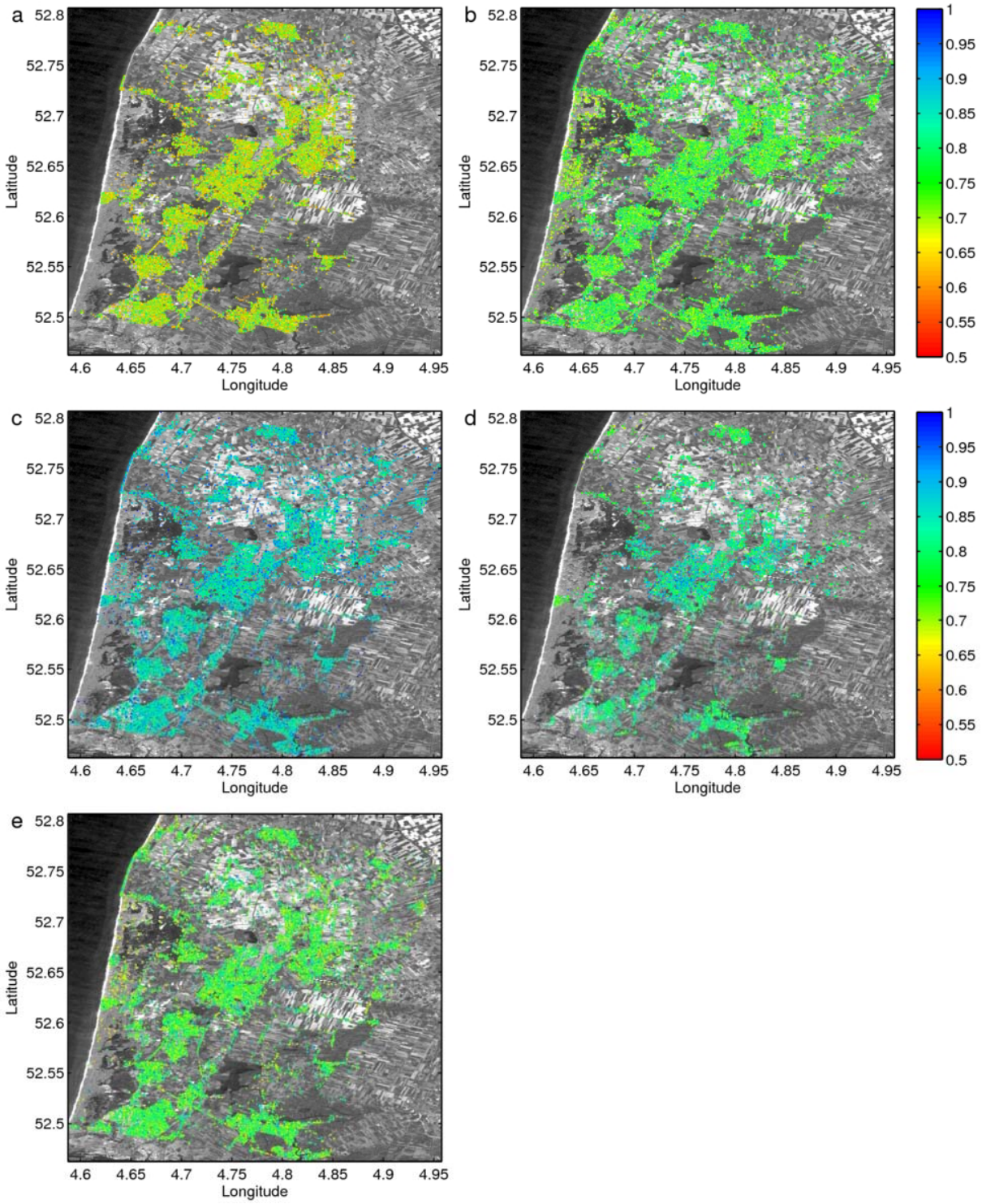

Figure A. 9: Alkmaar, Envisat. Temporal coherence estimates around Alkmaar for Envisat time series (Mar. 2003 – Mar. 2007). a) - d) OSP results, e) TU Delft result.

A.4 Quality classification based on STC

Table A. 1: RMSE of PSI versus levelling displacement velocities [mm/y] based on spatio-temporal consistency (STC) class.

	Team A	Team B	Team C	Team D	Team E
Alkmaar-ERS					
Top 20% stc					
Number of benchmarks	4	15	3	9	4
RMSE PSI-levelling [mm/y]	0.96	0.81	1.17	0.75	2.01
Top 40% stc					
Number of benchmarks	13	36	12	19	12
RMSE PSI-levelling [mm/y]	0.78	1.48	0.72	1.19	1.4
Top 60% stc					
Number of benchmarks	20	69	28	33	19
RMSE PSI-levelling [mm/y]	0.91	1.49	1.04	1.12	1.15
Top 80% stc					
Number of benchmarks	27	95	38	47	29
RMSE PSI-levelling [mm/y]	0.87	1.55	0.99	1.13	1.24
100% stc					
Number of benchmarks	34	150	54	51	41
RMSE PSI-levelling [mm/y]	0.93	1.54	1.02	1.25	1.12
Alkmaar-ASAR					
Top 20% stc					
Number of benchmarks	5	5	5	6	6
RMSE PSI-levelling [mm/y]	1.56	0.68	1.81	1.29	0.58
Top 40% stc					
Number of benchmarks	21	24	15	13	18
RMSE PSI-levelling [mm/y]	1.38	1.57	1.71	1.18	0.89
Top 60% stc					
Number of benchmarks	33	50	31	25	33
RMSE PSI-levelling [mm/y]	1.39	1.66	1.53	0.93	1.27
Top 80% stc					
Number of benchmarks	54	71	51	35	54
RMSE PSI-levelling [mm/y]	1.38	1.60	1.44	0.99	1.41
100% stc					
Number of benchmarks	73	114	76	45	80
RMSE PSI-levelling [mm/y]	1.44	1.81	1.48	1.11	1.38

Table A. 2: RMSE of PSI versus levelling displacement velocities [mm/y] based on spatio-temporal consistency (STC) value.

	Team A	Team B	Team C	Team D	Team E
Alkmaar-ERS					
STC <2 mm					
Number of benchmarks	10	0	0	0	0
RMSE PSI-levelling [mm/y]	0.78	-	-	-	-
STC <4 mm					
Number of benchmarks	34	3	14	3	16
RMSE PSI-levelling [mm/y]	0.93	0.55	0.68	0.47	1.24
STC <6 mm					
Number of benchmarks	34	29	40	18	40
RMSE PSI-levelling [mm/y]	0.93	1.47	0.99	0.98	1.13
STC <8 mm					
Number of benchmarks	34	124	53	40	41
RMSE PSI-levelling [mm/y]	0.93	1.54	1.03	1.19	1.12
STC <10 mm					
Number of benchmarks	34	150	54	49	41
RMSE PSI-levelling [mm/y]	0.93	1.54	1.02	1.20	1.12
Alkmaar-ASAR					
STC <2 mm					
Number of benchmarks	70	0	3	6	0
RMSE PSI-levelling [mm/y]	1.35	-	2.31	1.29	-
STC <4 mm					
Number of benchmarks	73	23	53	37	30
RMSE PSI-levelling [mm/y]	1.44	1.58	1.45	1.02	1.18
STC <6 mm					
Number of benchmarks	73	97	76	43	74
RMSE PSI-levelling [mm/y]	1.44	1.73	1.48	1.12	1.41
STC <8 mm					
Number of benchmarks	73	114	76	45	79
RMSE PSI-levelling [mm/y]	1.44	1.81	1.48	1.11	1.39
STC <10 mm					
Number of benchmarks	73	114	76	45	80
RMSE PSI-levelling [mm/y]	1.44	1.81	1.48	1.11	1.38

Table A. 3: RMSE of PSI versus levelling displacement time series [mm] based on spatio-temporal consistency (STC) class.

	Team A	Team B	Team C	Team D	Team E
Alkmaar-ERS					
Top 20% stc					
Number of double differences	14	65	9	32	11
RMSE PSI-levelling [mm]	4.83	5.14	5.28	4.86	9.46
Top 40% stc					
Number of double differences	41	150	44	79	39
RMSE PSI-levelling [mm]	4.75	6.17	3.54	4.76	6.42
Top 60% stc					
Number of double differences	56	257	103	109	66
RMSE PSI-levelling [mm]	4.53	5.82	3.73	5.38	5.84
Top 80% stc					
Number of double differences	86	359	148	154	119
RMSE PSI-levelling [mm]	4.52	5.84	3.68	5.07	5.48
100% stc					
Number of double differences	116	482	195	173	147
RMSE PSI-levelling [mm]	4.64	5.90	4.07	5.49	5.19
Alkmaar-ASAR					
Top 20% stc					
Number of double differences	18	10	8	8	9
RMSE PSI-levelling [mm]	4.61	6.53	5.03	4.98	2.27
Top 40% stc					
Number of double differences	36	33	24	19	23
RMSE PSI-levelling [mm]	4.01	5.30	4.40	4.04	2.76
Top 60% stc					
Number of double differences	47	62	42	32	38
RMSE PSI-levelling [mm]	4.57	5.06	5.08	3.73	4.19
Top 80% stc					
Number of double differences	57	84	53	37	58
RMSE PSI-levelling [mm]	4.49	4.89	4.61	3.57	4.12
100% stc					
Number of double differences	75	117	77	46	82
RMSE PSI-levelling [mm]	4.34	5.61	4.58	3.90	4.19

Table A. 4: RMSE of PSI versus levelling displacement time series [mm] based on spatio-temporal consistency (STC) value.

	Team A	Team B	Team C	Team D	Team E
Alkmaar-ERS					
STC <2 mm					
Number of double differences	32	0	0	0	0
RMSE PSI-levelling [mm]	4.49	-	-	-	-
STC <4 mm					
Number of double differences	116	8	47	10	46
RMSE PSI-levelling [mm]	4.64	4.30	3.51	6.08	6.26
STC <6 mm					
Number of double differences	116	122	156	73	143
RMSE PSI-levelling [mm]	4.64	5.66	3.64	4.64	5.24
STC <8 mm					
Number of double differences	116	426	190	148	147
RMSE PSI-levelling [mm]	4.64	5.93	4.12	5.11	5.19
STC <10 mm					
Number of double differences	116	482	195	168	147
RMSE PSI-levelling [mm]	4.64	5.90	4.07	5.34	5.19
Alkmaar-ASAR					
STC <2 mm					
Number of double differences	71	0	3	7	1
RMSE PSI-levelling [mm]	4.43	-	7.28	5.28	1.31
STC <4 mm					
Number of double differences	75	30	53	39	36
RMSE PSI-levelling [mm]	4.34	5.04	4.61	3.55	4.23
STC <6 mm					
Number of double differences	75	105	77	44	78
RMSE PSI-levelling [mm]	4.34	5.14	4.58	3.91	4.23
STC <8 mm					
Number of double differences	75	116	77	46	82
RMSE PSI-levelling [mm]	4.34	5.63	4.58	3.90	4.19
STC <10 mm					
Number of double differences	75	117	77	46	82
RMSE PSI-levelling [mm]	4.34	5.61	4.58	3.90	4.19