VGI

Österreichische Zeitschrift für

Vermessung & Geoinformation

91. Jahrgang 2003 Heft 1/2003

Organ der Österreichischen Gesellschaft für Vermessung und Geoinformation und der Österreichischen Geodätischen Kommission

A-1025 Wien, Schiffamtsgasse 1-3 Postgebühr bar bezahlt

AUSTRIAN CONTRIBUTIONS to the

XXIII. General Assembly of the
International Union of Geodesy and Geophysics
June 30 - July 11, 2003, Sapporo, Japan

VGI

Österreichische Zeitschrift für
VERMESSUNG & GEOINFORMATION

91. Jahrgang 2003 Heft 1/2003
ISSN 0029-9650

Schriftleiter: **Dipl.-Ing. Wolfgang Gold**
Stellvertreter: **Dipl.-Ing. Stefan Klotz**
Dipl.-Ing. Ernst Zahn
A-1025 Wien, Schiffamtsgasse 1–3

Organ der Österreichischen Gesellschaft für Vermessung und Geoinformation und der Österreichischen Geodätischen Kommission

Seite

Special Issue
XXIII General Assembly of the International Union of Geodesy and Geophysics (IUGG), Sapporo, Japan
ed. by H. Schuh, F.K. Brunner and H. Kahmen

Editorial of Special Issue

The Austrian Geodetic Commission (ÖGK) is the national organisation in Austria related to the International Association of Geodesy (IAG) which is one of the seven associations of the International Union of Geodesy and Geophysics (IUGG).

Every four years IUGG and its associations hold a General Assembly. The IAG uses these general assemblies to present the work performed, to review the structure of IAG, to appoint new officers and to organise a suite of symposia. The XXIII General Assembly of IUGG will take place in Sapporo, Japan, June 30 to July 11, 2003.

It has been a requirement that member countries of IAG prepare a National Report about the geodetic activities during the past four years for presentation at the general assemblies. However, quite often these National Reports provide only statistical information, rather than details about geodetic research work. Considering this situation, the ÖGK has decided to prepare a collection of geodetic research papers by Austrian institutions on the occasion of the forthcoming General Assembly.

The present issue of the Österreichische Zeitschrift für Vermessung & Geoinformation (VGI) may be considered a "snapshot" of the geodetic research work carried out in Austria. As a snapshot, it has the shortcoming that only limited areas of active research can be covered. I sincerely hope that additional special issues of VGI will help to broaden the picture about the impressive geodetic research activities. One paper of this issue of VGI provides a status report on the Austrian geoid, the other papers report on investigations of meteorology using geodetic space measurements, on modelling of GPS satellite clocks, on satellite navigation systems, on the development of a special calibration facility, and on applications of geodetic methods in engineering and in landslide studies. On behalf of ÖGK I would like to thank the authors for their efforts in preparing these very interesting papers.

The Austrian Geodetic Commission is proud to present this window on Austrian geodetic research work to the international and national geodetic community.

Fritz K. Brunner
President of ÖGK

The authors of this special issue of the Österreichische Zeitschrift für Vermessung & Geoinformation (VGI) represent the two Austrian universities, i.e., the Vienna University of Technology (TUW) and the Graz University of Technology (TUG), which offer the full programme in geodesy, and the Federal Office of Metrology and Surveying (BEV). Several of the papers are co-authored by scientists from abroad which shows the strong links of Austrian groups to the international geodetic community.

The main task of this special issue is to give the international geodetic research community an idea about the various topics Austrian geodesists are working on and also to show to those Austrian colleagues, who are usually involved in practical work, the new developments Austrian Scientists are involved in. The variety of research topics indicates how closely geodesy is related to its neighbouring disciplines such as geophysics, meteorology, space science, and computer science. Joint efforts in different fields are needed to reach the goals which are common in modern society, such as precise navigation on Earth or a thorough description of the various interactions in 'system Earth'.

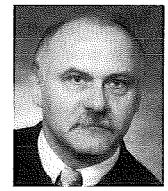
On behalf of all authors and of the two co-editors, I would like to thank the Austrian Society of Geodesy and Geoinformation for providing this platform for the publications. In particular I am grateful to Wolfgang Gold, editor-in-chief of the VGI, for taking care of the layout of the papers and solving all technical problems. Finally, it has to be mentioned that all papers of this issue were subject to an external reviewing process.

Harald Schuh
Main-editor of this special issue

The Austrian Geoid – Recent Steps to a New Solution

Erhard Erker, Norbert Höggerl, Erich Imrek, Wien and Bernhard Hofmann-Wellenhof, Norbert Kühtreiber, Graz





Abstract

A refined version of the Austrian geoid with the working title „GEOID 2000" will be released after the IAG General Assembly in Sapporo. The project was worked out in a cooperation of the Federal Office of Metrology and Surveying and the Technical University of Graz, Institute of Geodesy. The territory of Austria serves as an ideal test area for the different computational methods concerning usability and accessible precision as well as for the compatibility of the available datasets. An overview of the computation process as well as the key figur es of the new geoid are discussed.

Zusammenfassung

Eine neue verfeinerte Version des Österreichischen Geoids mit dem Arbeitstitel „Geoid 2000" wird während der Generalversammlung der IAG in Sapporo präsentiert. Die Berechnungen wurden in einer Zusammenarbeit des Bundesamtes für Eich- und Vermessungswesen mit dem Institut für Geodäsie der Technischen Universität Graz durchgeführt.

Die unterschiedliche Topographie, die von den Alpen im Westen bis zu den großen Becken im Osten reicht, macht Österreich zu einem idealem Testgebiet für eine Geoidbestimmung. Dabei können die Anwendung und die Genauigkeit von Berechnungsmethoden einerseits und die Übereinstimmung verschiedenartiger Datensätze andererseits ideal untersucht werden.

Die vorliegende Arbeit beschreibt in den einleitenden Abschnitten die der Neuberechnung zugrunde liegenden Daten. Seit der letzten hochauflösenden Geoidberechnung 1987 wurden mehrere Datensätze stark verbessert sowie zusätzliche Daten erschlossen. So liegt nun ein umfassender Datensatz von Schwereanomalien vor. Für die Reduktion der Messgrößen wurde vom Bundesamt für Eich- und Vermessungswesen ein neues hochauflösendes Höhenmodell (44 m x 49 m) bereitgestellt. Weiters wurde ein homogener Datensatz von GPS Punkten verwendet, bei dem besonderer Wert auf die Genauigkeit der Höhenkomponente gelegt wurde. Für alle GPS Punkte liegen hochgenaue orthometrische Höhen, die in das europäische UELN-95/98, version 13, eingebunden sind, vor.

Für die Geoidberechnung kommt ein „Remove-Restore" Prozess zur Anwendung. Die Geoidhöhe wird mittels Kollokation aus Schwere- und Lotabweichungsdaten bestimmt. Um eine Aussage über die erreichbaren Genauigkeiten sowie die Möglichkeit der Kombination von Schwere und Lotabweichungen zu erhalten, wurden ein astrogeodätisches Geoid (nur Lotabweichungen), ein gravimetrisches Geoid (nur Schwereanomalien) und eine Kombinationslösung (Lotabweichungen und Schwereanomalien) bestimmt. Für die Kombinationslösung ist dabei eine eingehendere Untersuchung der Gewichte der Lotabweichungen im Verhältnis zu den Schwereanomalien notwendig.

Die Genauigkeit der einzelnen Lösungen wurde durch den Vergleich der resultierenden Geoidhöhen mit Geoidhöhen, die aus orthometrischen Höhen und ellipsoidischen Höhen (GPS) abgeleitet wurden, überprüft. Dazu wurden 3D-Koordinaten mithilfe der Geoidhöhen und orthometrischen Höhen abgeleitetet und in das Referenzsystem der 3D-Koordinaten aus GPS (System ETRF89) transformiert. Die Restklaffen der Transformation sind ein Maß für die Genauigkeit der Berechnungen. Für alle Geoidlösungen (astrogeodätische, gravimetrische und kombinierte Lösung) können die Restklaffen in einen Trend und Abweichungen davon aufgespalten werden. Grundsätzlich zeigt sich eine gute Übereinstimmung von astrogeodätischer und gravimetrischer Lösung. Das beste Resultat zeigt die komibinierte Lösung. Die Abweichungen der Restklaffen vom Trend liegen dabei im Mittel bei $\pm$ 1.4 cm und bestätigen die hohe Genauigkeit der Lösung.

1. Introduction

The first determination of a high resolution geoid in Austria was performed in 1987 using a set of more than 650 deflections of the vertical [1] [2]. In 1998, when a reasonable amount of gravity data for Austria and the neighboring countries was available, a gravimetric geoid was computed [3]. The results of a combination of both datasets, supported by GPS and leveling data, using the remove-restore technique and collocation, was presented at the International

Meeting of the Gravity and Geoid Commission (IGGC 2002) of IAG Section III in Thessaloniki in August 2002 [4]. By combining gravity data and deflections of the vertical the precision of the recent geoid solution was increased to the cm-level. Nevertheless, the use of only 50 GPS/leveling points with an inhomogeneous distribution was unsatisfactory. Therefore an additional GPS-campaign was initiated in October 2002. The improved GPS-leveling results and its comparison with the geoid solution will be discussed in the following.

2. Data

2.1. Digital Height and Density Model

All investigations are based on the same digital terrain model with a uniform resolution of 44 m x 49 m [5] and a constant density value of 2,67 g/cm^3.

2.2. Gravity

The basic gravity data set is a subset out of 86000 Austrian gravity observations and gravity material of the neighboring countries. In an inner zone ($46.20° \le \phi \le 49.21°$ and $9.25° \le \lambda \le 17.25°$) 5796 gravity data points were selected, approximately representing a grid with 6 km grid spacing. Additional data with the approximate grid spacing of 12 km were added in the outer zone ($45.70° \le \phi \le 49.70°$ and $8.50° \le \lambda \le 18.20°$). More details about the basic data set can be found in [4].

2.3. Deflections of the Vertical

The set of more than 650 deflections of the vertical used in the 1987 solution is used without modification in the following investigation.

2.4. GPS

As mentioned above, the last geoid determination [4] was supported by the use of 50 GPS/leveling points. The resulting geoid heights were fitted with the help of 12 selected points. The remaining 38 points were used for an external check.

In some parts unexplainable discrepancies between GPS/leveling and the three solutions (gravity, astro, combined) occurred. Therefore an additional GPS-campaign was initiated in October 2002, including some new leveling connections and an examination of the stability and identity of the observed points. As a result of this campaign a set of 102 points in the frame of AGREF/AREF (the Austrian GPS reference frame), which represents the Austrian densification of EUREF/ETRS89, was obtained.

All points were measured with at least 24-hour sessions during the last decade.

At 50% of the points observations of two independent sessions were available. The high precision of the up-component by repeated GPS-measurements is shown in Fig. 1. Height differences of 4 to 6 cm are due to early GPS measurements in the beginning 1990ies which may be disturbed by geometry, multipathing and troposphere. Additional measurements are planned.

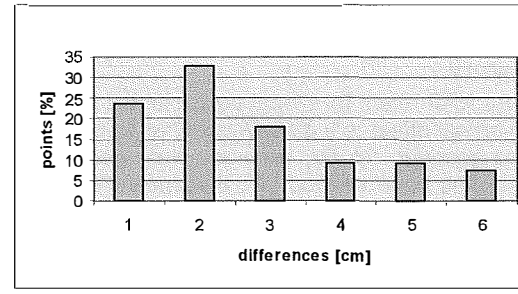


Fig. 1: Maximum height differences between double or multiple GPS-measurements

2.5. Leveling

2.5.1. Orthometric Heights

Orthometric heights refer to the geoid and are linked to ellipsoidal heights by the simple formula:

$$H_{ell} = N + H_{orth} \quad (1)$$

with

H_{ell} ellipsoidal height,
N geoid undulation,
H_{orth} orthometric height.

Orthometric heights can be derived with the formula

$$H_{orth} = C/g^* \quad (2)$$

where

$C = g \cdot \Delta H$ ... geopotential number
g^* ..... integrated mean value of the gravity between the surface and the zero-level
g ..... gravity value at the surface point
ΔH ..... levelled height difference

For more than ten years, the introduction of an orthometric height system in Austria has been

under way. So far orthometric height values are available for all precise leveling points (more than 30.000).

2.5.2. The European Vertical Reference System EVRS

In order to establish a homogeneous height system for the whole European continent, the EUREF[1]) subcommission initiated a new adjustment of the 1st order leveling network of all European countries by the use of ΔC-values. For the Austrian part in this project the solution UELN-95/98 vers. 13 [6] was used as a basis for the computation of C-values for all precise leveling points. The adjustment of the ΔC values in the UELN was performed as an unconstrained adjustment linked to the reference point NAP (Normaal Amsterdams Peil). In this adjustment, the geopotential value of NAP is

$C_{NAP} = 0$.

As NAP refers to a local equipotential surface, a vertical offset to a global geoid has to be taken into account. So far, no world-wide geoid surface has been defined yet, that means that the above mentioned offset is unknown.

2.5.3. The estimation of the precision of orthometric heights

2.5.3.1. The geopotential number C

In Austria a gravity network based on 30 absolute gravity points and on about 700 first order points has been established. Along the precise leveling lines 23.000 gravity points have been measured. Additionally 15.000 gravity points, evenly spread over the territory of Austria, exist. By the use of a digital terrain model (~50 m x 50 m) the interpolation of gravity values for surface points with a precision better than 1 mgal (1.10^{-5} m/s^2) is possible. This is sufficient to achieve the same level of precision for the ΔC-values as for the ΔH values.

If the Austrian part of the adjustment of the UELN95/98 vers. 13 is considered, the C-values show a standard deviation of $m_{0/1} = \pm$ 0.8 kgalmm/km while for the C-values with reference to NAP $m_{cNAP} = \pm$(10–12) kgalmm. If a central point in Austria is used as reference the standard deviation of the C-values is $m_{cA} = \pm$(4–8) kgalmm (internal precision).

2.5.3.2. The integrated mean values g of the gravity along the plumb line*

The computation of g*-values faces two problems:

- the determination of the integrated mean value of the gravity,
- the estimation of the influence of the varying mass density on the reduction process.

Intensive investigations have led to the following results. For the integration of the mean gravity value two cases are considered:

- For points with an altitude lower than 1400 m, the Kepler interpolation method is applied as it proved to give the best results. Consequently, the gravity for 3 points along the plumb line (with equal spacing) are used to calculate the mean value. For the weights the relationship 1-4-1 is chosen.
- For points above 1400 m altitude, Simpson's rule is applied. Five points along the plumb line (with equal spacing) are used, the weights for the gravity values of these points being 1-4-2-4-1.

The differences of the g* values computed by means of Kepler's method (or by Simpson's rule) in comparison with a g* value computed by the use of 20 intermediate points along the plumb line are smaller than 1.1 mgal [7].

The estimation of the influence of the varying density of masses can be summarised as follows:

An estimation of the influence of the varying density (2,8 g/cm^3 instead of 2.67 g/cm^3) on the orthometric height of a benchmark with an altitude of 2577m (Edelweißspitze/ Großglockner) which in fact is the highest precise leveling bench mark in Austria, shows a value up to 32 mm in maximum. On the other hand investigations done by Sünkel [8] show that larger density anomalies will be reduced by the fact that they are isostatically compensated.

It can be shown that for about 73% of the points used in this geoid investigation, the influence of the varying density is smaller than 8 mm (for a density variation of 15%); for 18% of the points the influence could be up to 15 mm.

Summarizing the above mentioned error influences on the determination of the orthometric heights, the following rough estimation can be given:

$m_{Horth} < 15$ mm (altitude $<$ 1000 m),
$m_{Horth} < 20$ mm (altitude 1000 – 1500 m),
$m_{Horth} < 25$ mm (altitude 1500 – 2000 m).

1) EUREF: IAG-Subcommission for Europe

3. The Geoid Computation

3.1. Remove-Restore

The geoid computation is done by the remove-restore procedure. The basic idea behind it is to take advantage of the fact that parts of the gravitational potential can be approximated by existing models. The long wavelength part is known through a given earth's gravitational model which is expressed in terms of a spherical harmonic expansion. The short wavelength part is a function of the mass (density) distribution of the topography and can be modeled by digital terrain and density models.

The remove-restore technique is applied in the following way. In the remove step residual gravity anomalies (gRES are computed by

$$\Delta g_{RES} = \Delta g - \Delta g_{EGM} - \Delta g_{DTM} \quad (3)$$

The two effects removed from the gravity anomalies Δg are Δg_{EGM}, the long wavelength part of the gravity anomalies, and Δg_{DTM}, the short to medium wavelength part of the gravity anomalies. By this remove step we gain Δg_{RES} which represent a smooth field with only local-to-regional structures.

Here the adapted EGM96 [9] was used to compute the long wavelength part in the remove-restore procedure. For the short to medium wavelengths, a topographic isostatic reduction was performed using the adapted technique and a detailed height model with the resolution 11.25″ x 18.75″. For the isostatic model an Airy-Heiskanen approach with a standard constant density of 2.67 g/cm^3, a normal crustal thickness T of 30 km and a crust-mantle density contrast of 0.4 g/cm^3 was used. Table 1 shows the statistics for the reduction process.

Mgal	min	max	mean	std.dev.
Δg	–154.1	187.2	9.8	±42.2
$\Delta g - \Delta g_{EGM}$	–204.3	224.0	–1.1	±47.6
Δg_{RES}	–72.0	85.4	0.6	±23.6

Tab. 1: Gravity reduction using the standard density value of 2.67 g/cm^3 and the adapted geopotential model EGM96. Statistics are based on 5796 points (6 km x 6 km set).

After the remove-step the geoid heights N_{RES} are modeled from the residual gravity anomalies Δg_{RES}. In the following the estimation was done by collocation. Details on the used covariance function are given in chapter 3.2.

Finally the removed effects are restored again

$$N = N_{RES} + N_{EGM} + \delta N_{DTM} \quad (4)$$

Here is the indirect effect which takes into account that removing the masses has changed the potential. N_{EGM} is computed using the spherical harmonic expansion of the earth's gravitational model.

This technique is commonly applied in local gravity field determination. Early computations done by this method are e.g. [2] and [11].

3.2. Covariance Function

The well-known Tscherning-Rapp covariance function model was used for the following LSC solutions. The global covariance function of the gravity anomalies $C_g(P,Q)$ given by Tscherning and Rapp ([12], p. 29) is written as

$$C_g(P,Q) = A\sum_{n=3}^{\infty}\frac{n-1}{(n-2)(n+b)}s^{n+2}P_n(\cos\psi) \quad (5)$$

where $P_n(\cos\psi)$ denotes the Legendre polynomial of degree n, ψ is the spherical distance between P and Q and A, B and s are the model parameters. A closed expression for (Equ. 5) is available in (ibid., p. 45).

The local covariance function of gravity anomalies C(P,Q) given by Tscherning-Rapp can be defined as

$$C(P,Q) = A\sum_{NN+1}^{\infty}\frac{n-1}{(n-2)(n+B)}s^{n+2}P_n(\cos\psi) \quad (6)$$

Modeling the covariance function means in practice fitting the empirically determined covariance function (through its three essential parameters; the variance C_0, the correlation length ξ and the variance of the horizontal gradient G_0) to the covariance function model. Hence the four parameters A, B, NN and s are to be determined through this fitting procedure. A simple fitting of the empirical covariance function was done using COVAXN-Subroutine (13).

The essential parameters of the empirical covariance parameters for 2489 gravity stations in Austria are 740.47 mgal^2 for the variance C_0 and 43.5 km for the correlation length ψ_1. The value of the variance for the horizontal gradient G_0 was roughly estimated as 100 E^2.

With a fixed value B = 24, the following Tscherning-Rapp covariance function model parameters were fitted: s = 0.997065, A = 746.002 mgal^2 and NN = 76. The parameters were used for the astrogeodetic, the gravimetric as well as the combined geoid solution.

3.3. Astrogeodetic Solution

The astrogeodetic solution by collocation is based on 659 deflections of the vertical uni-

formly distributed over Austria. After removing the long and short wavelength effects of the gravitational potential from the observations a geoid was estimated by LSC. Figure 2 shows the difference between the geoid solution by LSC and the GPS/leveling derived geoid. Be aware that the contour plot is based on the differences given at few selected GPS/leveling points. For the moment we are only interested in the long wavelength character of the differences which show a west-east trend of 1 m.

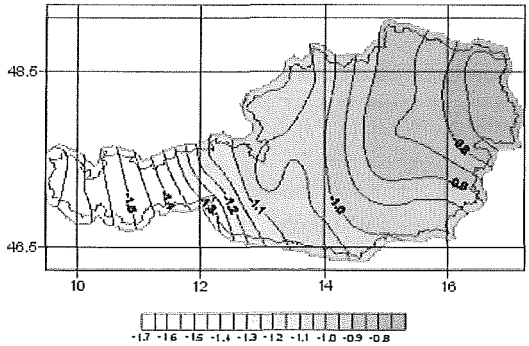


Fig. 2: Difference in geoid heights given in m, for the astrogeodetic geoid solution and the GPS/leveling geoid. Contour interval = 5 cm.

3.4. Gravimetric Solution

The gravimetric solution by collocation is based on the gravity anomaly data set mentioned in chapter 2.2 and presented in [4]. Figure 3 shows the difference between the LSC geoid solution and the GPS/leveling derived geoid. Once again the contour plot is based on the differences given at few selected GPS/leveling points. The differences are of the same magnitude as for the astrogeodetic geoid result (see Fig. 2). The differences show a high order polynomial trend with a west-east gradient of about 0.8 m over 600 km.

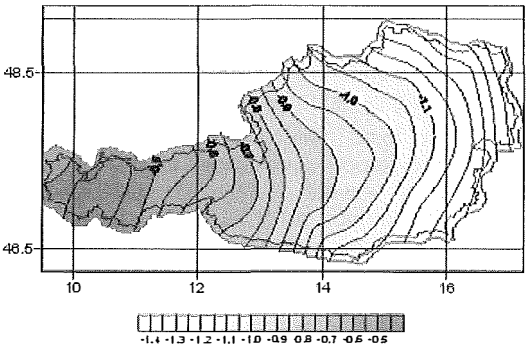


Fig. 3: Difference in geoid heights given in m, for the gravimetric geoid solution and the GPS/leveling geoid. Contour interval = 5 cm.

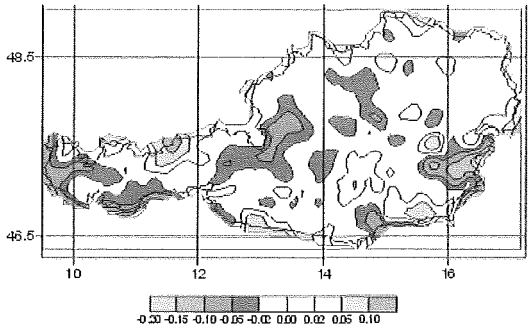


Fig. 4: Difference of the gravimetric solution fitted to GPS/leveling points minus the astrogeodetic solution fitted to GPS/leveling points. Difference given in cm.

Of particular interest is the comparison of the gravimetric solution and astrogeodetic solution. The differences are in most places less than ± 10 centimeters (see Fig. 4). One should notice that unequally spaced contour intervals were chosen to point out three different categories, namely ± 2 cm, ± 5 cm and ± 10 cm. The white pattern shows all points where the two solutions agree within ± 2 cm. Around 50% of the points fulfill this criterion. If we choose the category ± 5 cm, around 75% of the differences are within this range. A closer look at the largest differences reveals that no correlation with the topography exists. For instance the big differences along the Austrian border reflect the fact that the astrogeodetic solution is based on deflections points inside Austria only, while the gravimetric solution was computed on a more regional basis. The biggest difference is located in the eastern part of Austria.

We conclude that the difference probably depends on the distribution of the deflections of the vertical. The more homogeneous and dense the distribution of these points, the better the agreement between the two solutions. Of course, differences also depend on the smoothness of the residual gravitational potential. Therefore the residual gravitational field should be as smooth as possible and the used measurements should homogeneously cover the region in order to get a precise geoid solution.

3.5. Combined Solution

The first combined solution was done by Kühtreiber [14]. This combination of the gravimetric and astrogeodetic geoid was done by computing a simple arithmetic mean of the astrogeodetic and the gravimetric solution. In order to take the advantage of collocation as a method for

combining gravity anomalies and deflections of the vertical in one estimation process, investigations concerning the relative weighting of these two data types are needed.

An important point concerning a reasonable combination of deflections of the vertical and gravity anomalies in a collocation process is the choice of the standard deviation for the different data types. A case study was carried out where the geoid heights were computed by a combination of deflections of the vertical and gravity anomalies. Three different cases, each with a different standard deviation for Δg but a fixed standard deviation for the deflections of the vertical, are considered. Each of the combined solutions is compared to the pure astrogeodetic and the pure gravimetric geoid solution (see Fig. 5).

Let us consider the starting configuration. The standard deviation for Δg was given as ± 0.3 mgal while the standard deviation for the deflections ξ and η were given as $\pm 0.2''$ and $\pm 0.3''$ respectively. Comparing the combined solution with the astrogeodetic solution shows more or less a big west-east trend with regions which don't fit the trend at all (e.g. eastern part of Austria). The difference between the combined solution and the gravimetric solution is a pure trend, no deviation from the trend is visible. The conclusions we can draw from the first pair of plots in Fig. 5 are:

- the standard deviation of ± 0.3 mgal for the gravity anomalies is too small. Hence the deflections of the vertical don't contribute to the solution on a local basis,

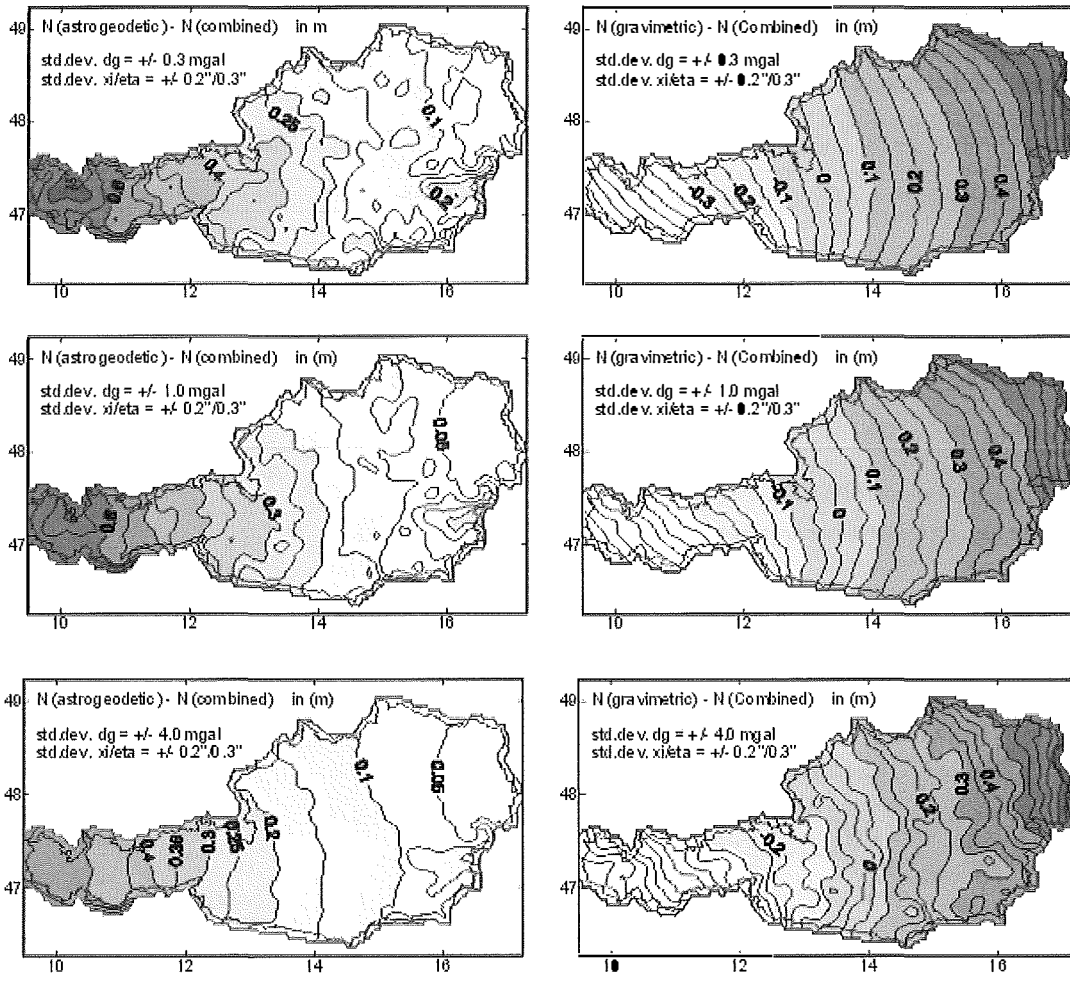


Fig. 5: Changing the standard deviation of the gravity anomalies in the combined solution and comparing the solution to the astrogeodetic and gravimetric solution.

- big discrepancies between the astrogeodetic and the combined solutions appear in regions with sparsely distributed deflections.

Increasing the standard deviation of the gravity anomalies and keeping the standard deviation of the deflections of the vertical fixed, should down-weight the influence of the gravity anomalies in the combined solution. Thereby the difference between the astrogeodetic and the combined solution should become a more or less pure regional trend, while the difference between the gravimetric and the combined solutions should become more irregular. The second and third pairs of plots in Fig. 5 prove this fact.

The poor results of the astrogeodetic solution in the east of Styria exist in all solutions. Even a very high weight for the deflections of the vertical relative to the gravity anomalies preserves the structure of the gravimetric geoid solution to some extent. This is clear as the deflections of the vertical are too sparse in this area to contribute to the combined solution.

As a conclusion of this study, the standard deviation of the gravity anomalies and the deflections of the vertical were chosen as ± 1.5 mgal for Δg and $\pm 0.2''$, $\pm 0.3''$ for ξ, η, respectively.

4. Comparisons with an extended GPS/leveling information

The comparison between the enlarged GPS/leveling set presented in chapter 2.4 and the geoid solutions of chapter 3 was performed in the following steps:

- *Interpolation of geoid undulations* (astrogeodetic, gravimetric and combined) for the available 102 GPS/leveling points by the use of Newton's interpolation algorithm for a regular grid. As a basis for the interpolation in each case 2847 grid points with a spacing of 3′ x 5′ were used. An individual point was calculated in the frame of the 16 adjacent grid points (degree 2 of Newton's interpolation).
- *Calculation of ellipsoidal heights* using the results of the above mentioned interpolation and leveled orthometric heights (chapter 2.5).
- *Calculation of 3D Cartesian coordinates* in combination with ETRF89-values (ϕ,λ) for each GPS-point.
- *Transformation into a best fitting position to the „real" GPS-derived ETRF89-values* by use of a 7-parameter Helmert transformation. The result of this transformation can be characterized by the following statistics:

solution	residual (mean value/cm)
ASTRO	4,0
GRAV	4,5
KOMB	3,7

- *Modeling the residuals:* The resulting residuals were modeled by use of Surfer 32 (Kriging, Point Griding in a regular grid of 10km x 10km) for the three given geoid solutions. (Fig. 6,7,8)

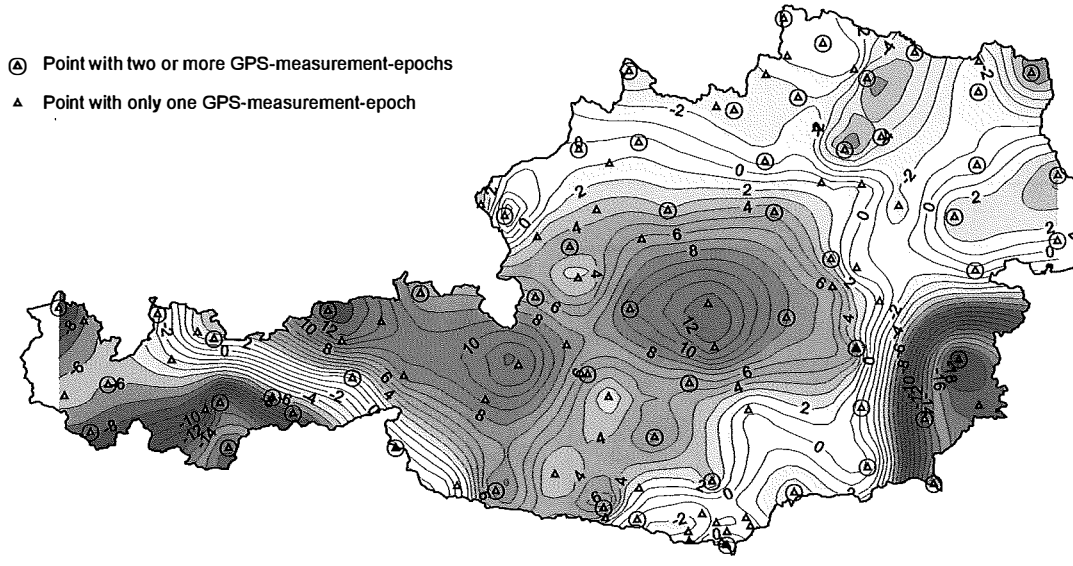


Fig. 6: Residuals GPS/Lev. minus ASTRO [cm]

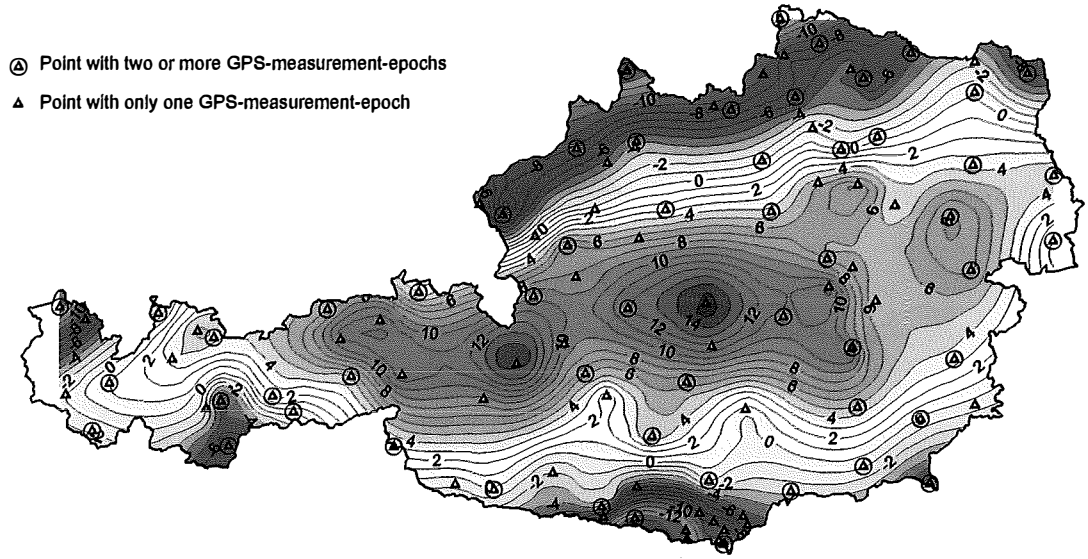


Fig. 7: Residuals GPS/Lev. minus GRAV [cm]

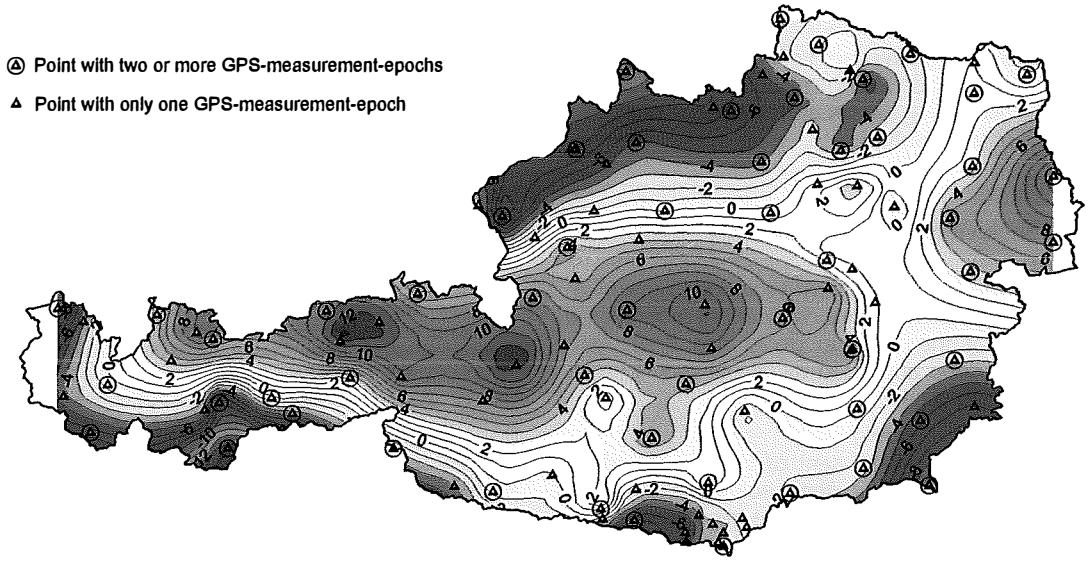


Fig. 8: Residuals GPS/Lev. minus COMBINATION [cm]

Generally the residuals show smooth long-wavelength distortions especially in the gravimetric solution. Again the residuals in the eastern part of the astrogeodetic solution are bigger than the overall trend. Further investigations in this area are needed.

5. The refined version of the Austrian geoid

By using and combining the astrogeodetic, the gravimetric data and the comprehensive GPS/leveling information presented above, a new refined version of the Austrian geoid could be performed. For this calculation the high precision of the available GPS-derived ellipsoidal heights as well as the high quality of orthometric heights recalculated within the new Austrian height system have to be taken into account.

Therefore the new geoid is based on the „Combined Solution" (chapter 3.5) and fitted to ETRS89 by use of the GPS/leveling information supplemented by the modeled residuals presented in Fig. 8.

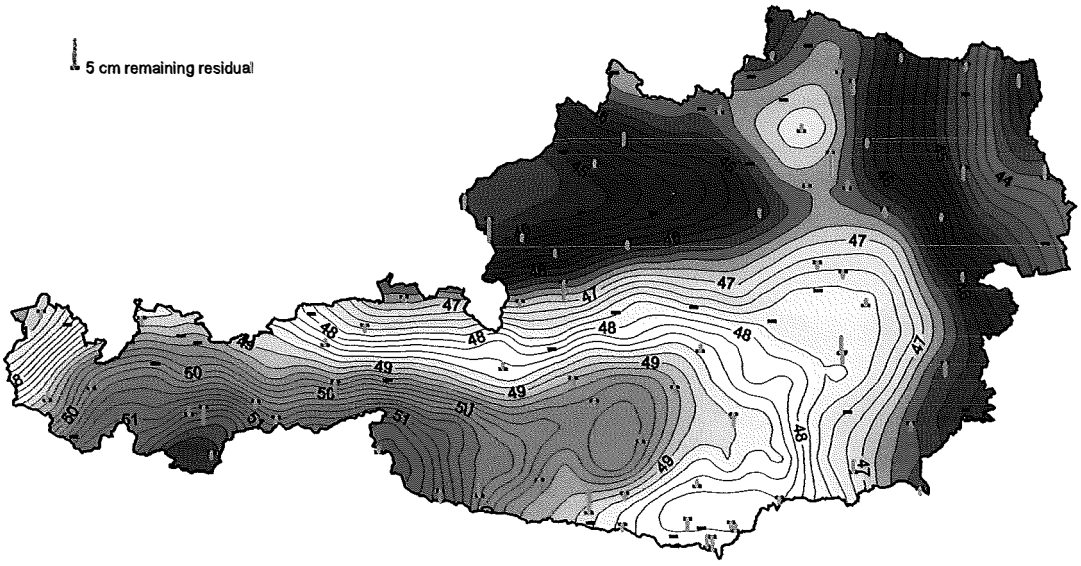


Fig. 9: Refined geoid after modeling the residuals

The resulting new geoid of Austria is presented in Fig. 9. The remaining residuals decrease to a mean value of $\pm$1,4 cm and mirror the high precision of the solution.

6. Conclusions

With the help of modern techniques (remove-restore and least square collocation) a combined solution of the Austrian geoid has been estimated, refined and fitted by use of precise GPS and leveling data.

The solution shows:

- an excellent agreement of the astrogeodetic and the gravimetric geoid solution
- a centimeter precision of the geoid solution as a result of the combined version,
- the influence of high precision GPS- and leveling data in the fitting procedure and in the possibility of modeling long wavelength distortions.

Finally it must be pointed out, that the new Austrian geoid solution is a wide step forward, but there are still several remaining problems, which have to be investigated. One of these problems is the local distorsion of the astrogeodetic solution in the eastern part of Austria. At least a remeasurement and densification of the deflections of the vertical especially in that region together with additional GPS-leveling points covering the total area of Austria is planned.

References

[1] *Erker, E. 1987.* The Austrian Geoid; Local Geoid Determination Using Modified Conservative Algorithms, In: The Gravity Field in Austria, Geodätische Arbeiten Österreichs für die Internationale Erdmessung, Neue Folge, Band IV, edited by the Austrian Geodetic Commission, Graz.

[2] *Sünkel, H., Bartelme, N., Fuchs, H., Hanafy, M., Schuh, W., Wieser, M. 1987.* The Gravity Field in Austria, In: The Gravity Field in Austria, Geodätische Arbeiten Österreichs für die Internationale Erdmessung, Neue Folge, Band IV, edited by the Austrian Geodetic Commission, Graz.

[3] *Kühtreiber, N. 1998.* Improved Gravimetric Geoid AGG97 of Austria, In: Forsberg, R., Feissel, M. and Dietrich, R., eds. Geodesy on the Move: Gravity, Geoid, Geodynamics and Antarctica, IAG Symposia, Rio de Janeiro, Brazil, 306–311.

[4] *Kühtreiber, N. 2002.* High Precision Geoid Determination of Austria Using Heterogeneous Data, In: Gravity and Geoid Commission (IGGC 2002) of IAG Section III, Proceedings (in press), Thessaloniki, Greece, August 2002

[5] *Graf, J. 1996.* Das digitale Geländemodell für Geoidberechnungen und Schwereduktionen in Österreich, Proceedings of the 7th International Meeting on Alpine Gravimetry, Vienna, Österr. Beiträge zu Meteorologie und Geophysik, 121–136.

[6] *Sacher, M., Ihde, J., Celms, A., Ellmann, A. 1999.* The first UELN Stage is Achieved. Report on the Symposium of the IAG Subcommission for the European Reference Frame (EUREF) held in Prague, 2–5 June 1999. Veröffentlichungen der Bayerischen Kommission für die Internationale Erdmessung der Bayerischen Akademie der Wissenschaften, Astronomisch-Geodätische Arbeiten Nr. 60, München 1999, S. 87–94.

[7] *Ruess, D., Höggerl, N. 2000,* Orthometric Height System in Austria by Refined Methods of Gravity Reduction. Presentation at the 8th International Meeting on Alpine Gravity, Leoben 2000.

[8] *Sünkel, H. 1986.* Konventionelle und moderne Verfahren zur Ableitung orthometrischer Höhen. Österreichische Zeitschrift für Vermessungswesen und Photogrammetrie. 74. Jg. Heft 2, 1986.

[9] *Abd-Elmotaal, H. and Kühtreiber, N. 2001.* Precise Geoid Computation Employing Adapted Reference Field, Seismic Moho Information and Variable Density Anomaly, Presented at the Scientific Assembly of the International Association of Geodesy IAG 2001, Budapest, Hungary, September 2–8, 2001.

[10] *Forsberg, R. 1984.* A Study of Terrain Reductions, Density Anomalies and Geophysical Inversion Methods in Gravity Field Modelling, Department of Geodetic Science, Ohio State University, Columbus, Ohio, 355.

[11] *Tscheming, C.C. and Rapp, R.H. 1974.* Closed Covariance Expressions for Gravity Anomalies, Geoid Undulations, and Deflections of the Vertical Implied by Anomaly Degree Variance Models, Department of Geodetic Science, The Ohio State University, Columbus, Ohio, 208.

[12] *Tscheming, C.C. 1976.* Implementation of Algol-Procedures for Covariance Computation on the RC 4000-Computer, The Danish Geodetic Institute, 12.

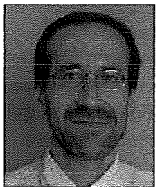
[13] *Kühtreiber, N. 1999.* Combining Gravity Anomalies and Deflections of the Vertical for a Precise Austrian Geoid, Bollettino di Geofisica teorica ed applicata, 40, 545–553.

Contact

Dr. Erhard Erker, Dipl.-Ing. Norbert Höggerl, Dipl.-Ing. Erich Imrek: Bundesamt für Eich- und Vermessungswesen, Grundlagenvermessung, Schiffamtsgasse 1–3, A-1025 Wien. email: erhard.erker@bev.gv.at, norbert.hoeggerl@bev.gv.at, erich.imrek@bev.gv.at
Univ.-Prof. Dr. Bernhard Hofmann-Wellenhof, Univ.-Ass. Dr. Norbert Kühtreiber: Technische Universität Graz, Abteilung für Positionierung und Navigation, Steyrergasse 30, A-8010 Graz.
email: hofmann-wellenhof@tugraz.at, kueh@phgg.tu-graz.ac.at

IVS Pilot Project – Tropospheric Parameters

Harald Schuh and Johannes Boehm, Wien




Abstract

In April 2002 the IVS (International VLBI Service for Geodesy and Astrometry) set up the Pilot Project – Tropospheric Parameters, and the Institute of Geodesy and Geophysics (IGG), Vienna, was asked to coordinate the project. Seven IVS Analysis Centers have joined the project until now and submitted their estimates of tropospheric parameters (wet and total zenith delays, horizontal gradients) for all IVS-R1 and IVS-R4 sessions since January 1st, 2002, on a regular basis. The individual submissions are combined by a two-step procedure to stable, robust and highly accurate tropospheric parameters with 1 h resolution. The zenith delays derived by VLBI (Very Long Baseline Interferometry) are compared with those provided by the International GPS Service (IGS). At collocated sites (VLBI and GPS antennas at the same station), almost constant biases are found between the GPS (Global Positioning System) and VLBI derived zenith delays, although the signals recorded by both techniques are subject to the same tropospheric delays. Possible reasons for these biases are discussed.

Kurzfassung

Ganz ähnlich wie das GPS-Verfahren ist auch die Radiointerferometrie auf langen Basislinien (Very Long Baseline Interferometry, VLBI) in der Lage, troposphärische Laufzeitverzögerungen in Zenitrichtung sehr genau zu bestimmen. Diese beinhalten unter anderem Informationen über den Feuchtegehalt der Troposphäre an den beteiligten VLBI-Stationen. Die Ergebnisse können nicht nur für meteorologische Zwecke verwendet werden, sondern spielen auch in der Klimaforschung eine Rolle. Wieder einmal zeigt sich, dass sozusagen ein Nebenprodukt geodätischer Messungen von großem Interesse für Nachbardisziplinen der Geodäsie sein kann. Zwar ist die globale Verteilung von VLBI-Stationen nicht so hoch wie bei GPS und eine Auswertung in Echtzeit ist noch nicht möglich, aber dennoch sind die troposphärischen Laufzeitverzögerungen der VLBI auf Grund ihrer hohen Genauigkeit von großer Bedeutung für Vergleiche mit Ergebnissen von GPS oder anderen Techniken, wie z. B. Wasserdampfradiometern. Außerdem können für einige VLBI-Stationen konsistente Zeitserien der troposphärischen Parameter von beinahe 20 Jahren ermittelt werden, die für klimatologische Studien herangezogen werden können. Aus diesen Gründen wurde im April 2002 durch den IVS (International VLBI Service for Geodesy and Astrometry) das 'Pilot Project – Tropospheric Parameters' eingerichtet, und das Institut für Geodäsie und Geophysik (IGG) der TU Wien wurde mit der Koordination des Pilotprojekts betraut. Mittlerweile nehmen sieben VLBI-Analysezentren teil und reichen regelmäßig ihre Schätzungen der troposphärischen Parameter (totale und feuchte Laufzeitverzögerung in Zenitrichtung, horizontale Gradienten) der IVS-R1 und IVS-R4 Experimente seit 1. Jänner 2002 ein. Die einzelnen Abgaben werden am IGG in einem zweistufigen Verfahren zu genauen und stabilen troposphärischen Parametern mit stündlicher Auflösung kombiniert. Diese Laufzeitverzögerungen in Zenitrichtung wurden mit den vom IGS (International GPS Service) ermittelten Werten verglichen. An Stationen mit VLBI- und GPS-Antennen treten konstante Differenzen zwischen den Laufzeitverzögerungen auf, obwohl beide Verfahren den gleichen troposphärischen Einflüssen unterliegen. Mögliche Gründe dafür werden diskutiert.

1. Introduction

In the last few years, the collaboration between geodesy and meteorology/climatology has become more and more intensive. GPS (Global Positioning System) has proved to be very important for meteorology, and because of the short delay between the GPS observations and the availability of tropospheric results, these can even be used for weather-forecasts. Tropospheric parameters determined by VLBI (Very Long Baseline Interferometry) are mainly useful for climatological studies. Since there is a long history of consistent VLBI sessions since 1984, they comprise accurate information about the long-term development of precipitable water above the VLBI sites. Furthermore, due to their high accuracy, the parameters derived by VLBI are of interest for the validation and calibration of parameters determined by GPS, WVR (water vapour radiometer) and other techniques.

In VLBI data analysis, tropospheric modeling is one of the major error sources. Therefore, a comparison of tropospheric parameters was part of the 2nd IVS (International VLBI Service for Geodesy and Astrometry) Analysis Pilot Pro-

ject in 2001. Ten time series submitted by nine Analysis Centers (ACs) were compared by the IVS Associate Analysis Center at the Institute of Geodesy and Geophysics (IGG) of the University of Technology, Vienna. The investigations showed that the series submitted by IVS ACs are consistent and of high quality (Boehm et al., 2002b, [2]). At the 7th IVS Directing Board meeting in Tsukuba (Feb. 2002) it was decided to set up an IVS Pilot Project on Tropospheric Parameters coordinated by IGG. This Pilot Project (PP) is a research and study project with a structure similar to the IVS Working Groups. After the call for participation by the IVS Analysis Coordinator in May 2002, six IVS ACs agreed to take part in the PP. In January 2003, the IVS AC at Onsala Space Observatory, Sweden, joined the project as the seventh AC. A Pilot Project Group (PPG) has been set up to coordinate all activities within the PP and to discuss all steps that should finally lead to operational products.

	IVS Analysis Center
BKG	Federal Agency for Cartography and Geodesy, Germany
CGS	Centro di Geodesia Spaziale, Italy
CNR	Istituto di Radioastronomia, Italy
GSF	NASA Goddard Space Flight Center, U.S.A.
IAA	Institute of Applied Astronomy, Russia
IGG	Institute of Geodesy and Geophysics, Austria
OSO	Onsala Space Observatory, Sweden

Table 1: IVS ACs taking part in the PP – Tropospheric Parameters. Onsala Space Observatory joined the PP in January 2003.

2. Submissions by the ACs

Most of the ACs have provided their tropospheric parameters beginning with January 2002. That allows the generation of a combined series since the start of the IVS-R1 and IVS-R4 sessions. Total and wet zenith delays as well as gradients are submitted by all ACs. GSF and IGG even apply a priori gradients calculated from numerical weather models. Most of the ACs use the CALC/SOLVE software package, only IAA and IGG apply the QUASAR and OCCAM software, respectively. About half of the ACs fix the ITRF2000, and all ACs use cutoff elevation angles at or below 5°. The Niell mapping functions (Niell, 1996, [5]) are used throughout – only IGG applies the isobaric mapping function of the hydrostatic part (Niell, 2001, [6]). Meteorological parameters can be extracted from the databases.

AC	a priori gradients	ITRF2000 fixed	software
BKG	no	yes	CALC/SOLVE
CGS	no	no	CALC/SOLVE
CNR	no	no	CALC/SOLVE
GSF	yes	no	CALC/SOLVE
IAA	no	yes	QUASAR
IGG	yes	yes	OCCAM
OSO	no	yes	CALC/SOLVE

Table 2: Features of the submissions. Two ACs use a priori gradients; four ACs fix ITRF2000.

The tropospheric parameters should be provided for every full hour, i.e. in equidistant time intervals of 60 minutes, starting at the first integer hour of the session. If other time intervals are used for the computation (e.g., longer time intervals for the gradients), all parameters have to be referred to the same hourly instants. More details about the Pilot Project – Tropospheric Parameters, the Pilot Project Group and the submissions of the ACs are described in Schuh et al. (2003, [8]).

3. Combination strategy for the total and wet zenith delays

Each AC that is taking part in the IVS Pilot Project – Tropospheric Parameters submits two files per week, namely one for the IVS-R1 and one for the IVS-R4 session. They are combined to weekly files in order to be comparable with results provided by the IGS, although most VLBI sites take part in one 24 h session per week only.

GPS week	IVS-R1 session	IVS-R4 session
1147	—	IVS-R4 001
1148	IVS-R1 001	IVS-R4 002
1149	IVS-R1 002	IVS-R4 003
1150	IVS-R1 003	IVS-R4 004
...	...	...

Table 3: The IVS-R1 and IVS-R4 sessions are combined to weekly files.

Before the combination, the data submitted by the ACs are edited using a limit of 30 mm for the formal errors. Estimates with larger formal errors are discarded. No interpolation has to be carried out to get the tropospheric parameters at the same time instants because the ACs were asked to provide their estimates at integer hours (see section 2). The combination itself is a two-step procedure which is carried out site by site, week by week and parameter by parameter (see Figure 1).

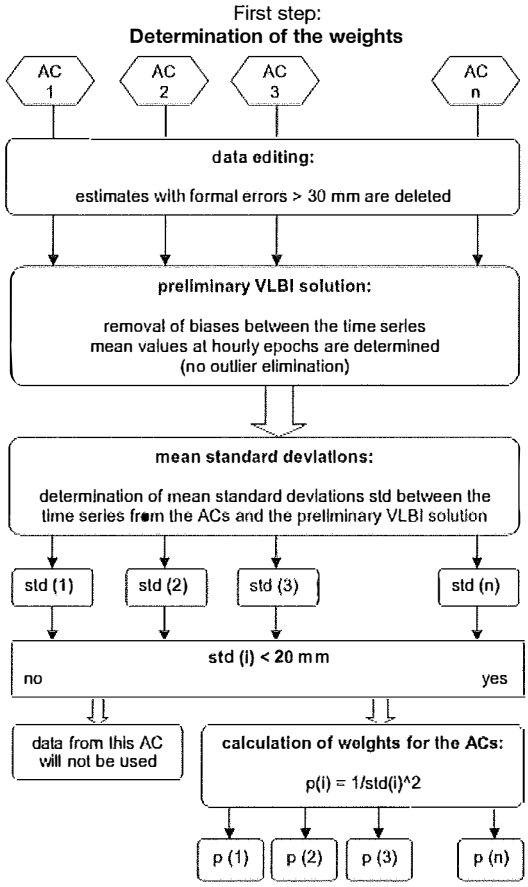


Figure 1a.: First step of the combination procedure. Weights for the individual ACs are determined and 'bad observations' are discarded.

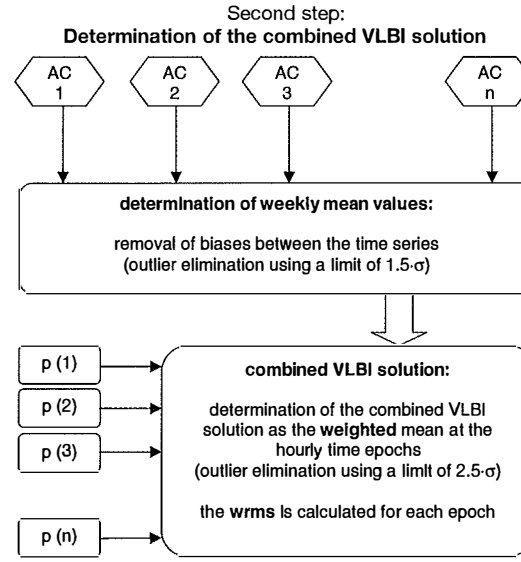


Figure 1b: Second step of the combination procedure. The combined VLBI solution is determined using outlier elimination.

In the first step preliminary VLBI time series of the total and wet zenith delays are produced. This combination comprises the removal of biases and the calculation of mean values at each time without any outlier elimination. Then the mean standard deviations between the preliminary VLBI time series and the time series of the ACs (shifted to the common mean) are computed for each week and each station. If a standard deviation is larger than 20 mm at a certain station, data from this AC will not contribute to

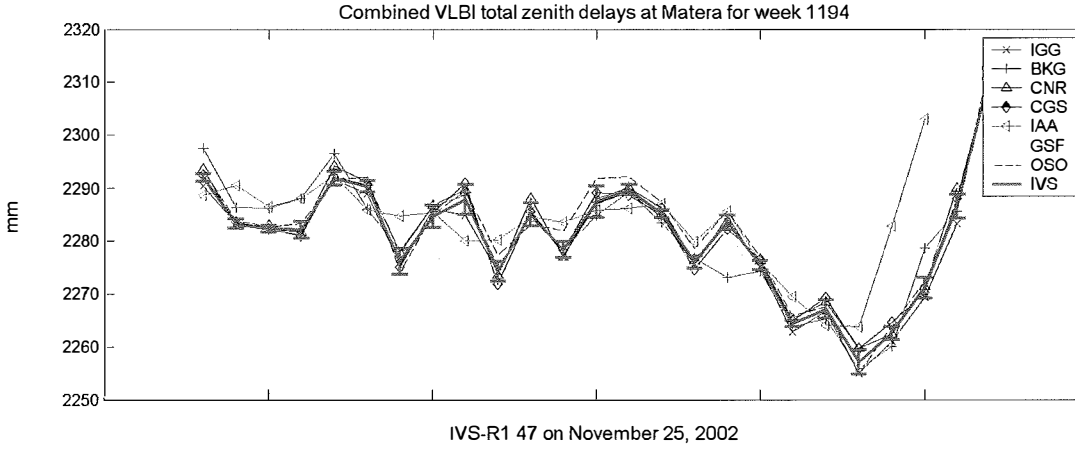


Figure 2: Submissions for the wet zenith delays at Matera by the various Analysis Centers (GPS week number 1194) and the combined VLBI solutions (red bold line with error bars). A rather good agreement between the time series can be seen. The mean of the standard deviations of the combined hourly results is ±1.5 mm.

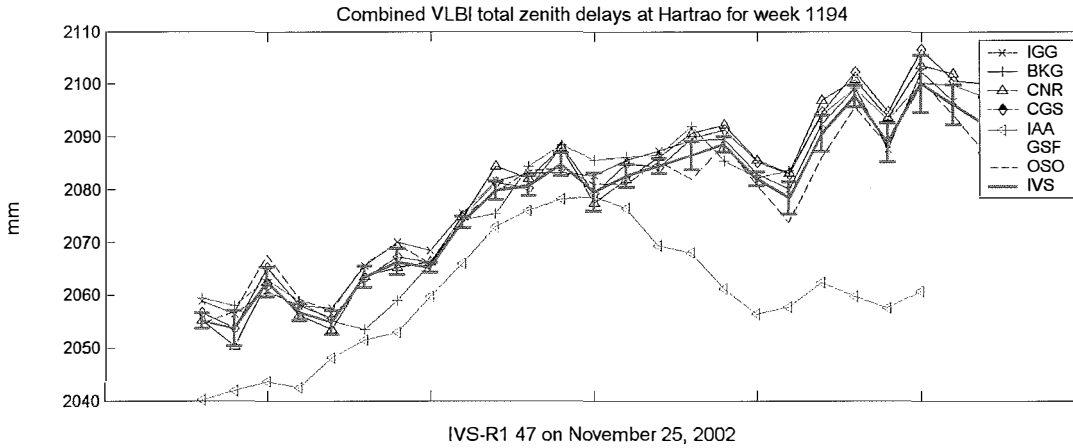


Figure 3: Submissions for the wet zenith delays at Hartrao by the various Analysis Centers (GPS week number 1194) and the combined VLBI solutions (red bold line with error bars). The mean of the standard deviations of the combined hourly results is ±2.5 mm.

the second step of the combination. Furthermore, a mean value of the standard deviations for all VLBI sites is determined for each AC. These mean standard deviations are used for assigning weights to the individual AC solutions in the final (second) combination.

In the second step the biases between the weekly time series are removed at each station using a limit of 1.5 σ (σ ... standard deviation). Then the VLBI values of the tropospheric parameters at each time are calculated as weighted means. Again, outliers are removed that exceed a limit of 2.5 σ.

With the approach described above, one VLBI time series is determined for the total and one for the wet zenith delays. Two examples with the wet zenith delays as submitted by the ACs and the combined solution can be seen in Figures 2 and 3. While Figure 2 (Matera) shows a rather good agreement between the ACs (±1.5 mm), the mean of the standard deviations of the combined hourly results in Figure 3 (Hartrao) is larger (2.5 mm). Anyway, the combined series is usually much smoother and thus probably more stable and robust than the individual submissions of the ACs. On the other side, short period variations of the zenith delays as for instance at Matera (Figure 2) seem to be reproduced by the combined values. In some sessions there were gaps in the observations at certain stations that have not been recognized by the ACs. For instance, if there were no observations in the middle of a 24 h session, the ACs might not be aware of this fact because they are using piecewise linear functions with constraints for the rates of the zenith delays. Another critical case occurs when no pressure data is available for a station and the ACs use adopted mean values for the pressure. Then the estimated wet delays are not used for the final product. To avoid these problems, IGG discards all combined estimates if there are no pressure data available in the database within one hour around the combination time.

Furthermore, so far a combined solution is only computed if there are at least data from three ACs contributing. Finally for cross checking, meteorological data are taken from the databases to compute the hydrostatic zenith delays at each station by the formula of Saastamoinen (1973, [7]). If the difference between the total and the hydrostatic plus wet delay of the combined solution is larger than 3 mm, the combined value at this time epoch is discarded.

4. Accuracy of the combined zenith delays

There are two kinds of accuracies that can be investigated. On the one hand, there is the accuracy of the absolute values. Apart from systematic errors due to the VLBI technique that might be inherent in the zenith delays submitted by all ACs, the weekly biases between the ACs should be a good criterion to evaluate the (remaining) absolute accuracy. Possible reasons for systematic biases in the VLBI estimates might be:

- errors of the terrestrial reference frame (at least for those solutions where the ITRF2000 is fixed),
- errors of the mapping functions,
- unmodelled effects (atmospheric loading, antenna deformation, ..)

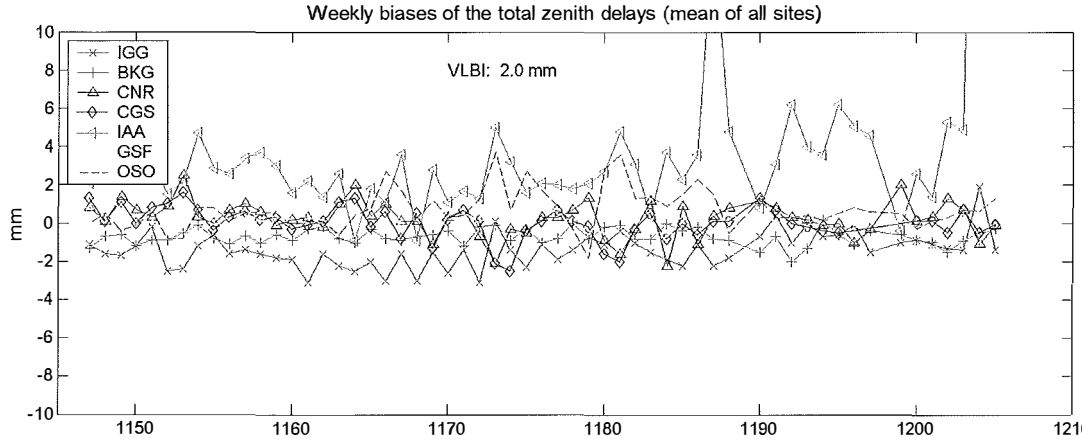


Figure 4: Weekly biases of the total zenith delays in 2002. The biases are within ±2 mm for most of the ACs.

On the other hand, relative accuracies can be determined after removing the weekly biases between the time series when the standard deviations at the hourly instants are evaluated.

4.1. Absolute accuracies

As can be seen in Figure 4, the weekly biases of the total (and wet) zenith delays are within (2 mm for most of the ACs. This indicates that – apart from systematic effects as described above – the accuracy of the absolute values of the zenith delays is at the 2 mm level, which is a mean value for all VLBI sites.

4.2. Relative accuracies

Relative accuracies can be calculated as the mean standard deviations at the hourly epochs after removing the weekly biases. Figure 5 shows the mean values (averaged per week) of the hourly standard deviations of the combined VLBI solution (red solid line) of the total zenith delays (mean of all sites). Additionally, the mean standard deviations of the hourly estimates of the individual time series against the combined VLBI solution are shown. Thus, the relative accuracy of the combined VLBI zenith delays is ~± 1.8 mm.

5. Comparison with tropospheric parameters determined by IGS

The IGS has produced tropospheric parameters for 150 IGS sites since 1997 (Gendt, 1996, [4]). This allows to compare at collocated sites (stations with VLBI and GPS antennas nearby) the combined total zenith delays derived by VLBI within the IVS-PP with those published by the IGS.

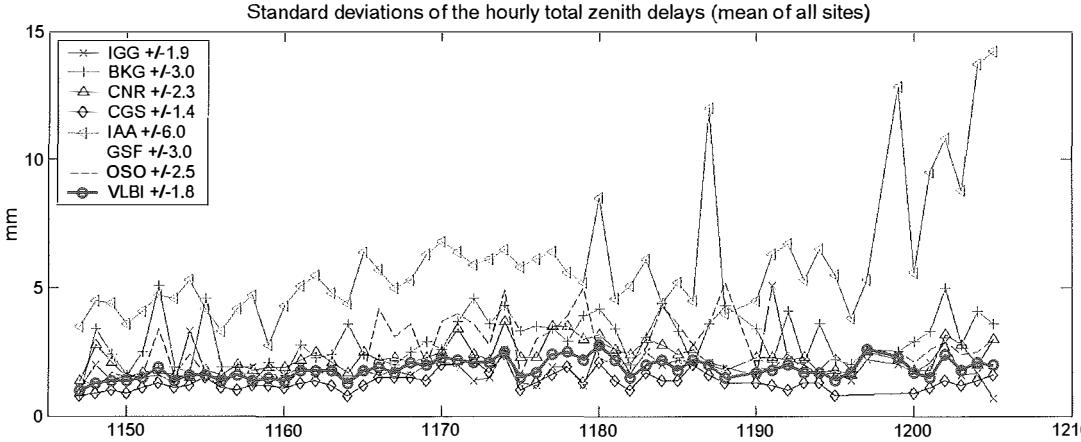


Figure 5: Mean values averaged per week of the standard deviations of the combined hourly zenith delays in 2002. Additionally, the mean values of the standard deviations for all stations are shown that were achieved by the individual ACs.

site	IVS acronym	IGS acronym	height diff. [m]	std. IVS [mm]	std. IGS [mm]
Algopark	ap	algo	23.0	±1.6	±2.2
Fortleza	ft	fort	3.3	±2.6	±4.4
Gilcreek	gc	fair	14.2	±1.5	±2.2
Hartrao	hh	hrao	2.3	±2.4	±3.1
Hobart26	ho	hob2	24.9	±2.4	±2.6
Matera	ma	mate	8.7	±1.8	±3.9
Medicina	mc	medi	18.1	±1.1	±1.3
Nyales20	ny	nyal	6.5	±1.4	±1.6
Seshan25	sh	shao	8.2	±1.8	±4.4
Wettzell	wz	wtzr	4.1	±1.5	±1.8
Onsala60	on	onsa	13.8	±1.0	±1.8

Table 4: Collocated sites with VLBI and GPS antennas. The 2-letter IVS acronyms and the 4-letter IGS acronyms are given as well as the height differences (VLBI – GPS) between the antennas. The last two columns show mean values of the hourly standard deviations for the combined IVS and IGS time series for identical epochs.

Because both services, IGS and IVS, use very similar combination strategies, a comparison of the mean values of the hourly standard deviations is possible. Table 4 shows these values for identical times at collocated sites. As mentioned before, the relative accuracy of the VLBI derived total zenith delays is at the ±2 mm level, and for most of the stations treated here it is slightly better than that from GPS.

In a second step, the biases and standard deviations between the IGS and IVS time series of the total zenith delays are determined. The height differences between the VLBI and GPS stations are accounted for by means of meteorological data recorded at the VLBI stations for the calculation of the differential hydrostatic and wet delays. Table 5 shows the mean biases between the time series and the standard deviations after removing these biases.

from the systematic effects for VLBI described above, there might be some problems with GPS observations as well:

- higher cutoff elevation angles applied in GPS (larger than 10 degrees),
- multipath effects,
- phase center variations of the antennas,
- errors of satellite ephemerides,
- same mapping function for the hydrostatic and wet delays

6. Results and conclusions

VLBI is capable of determining very accurate tropospheric zenith delays. Apart from systematic errors that might be inherent in the VLBI technique, the accuracy of the combined hourly VLBI results is at the 2–4 mm level. The first year of

site	bias	std	site	bias	std	site	bias	std
ap	7.1	±4.8	ft	13.5	±9.6	gc	4.2	±3.7
hh	5.2	±8.1	ho	3.2	±7.4	ma	3.9	±6.8
mc	1.4	±4.6	ny	4.1	±3.8	sh	1.5	±6.0
wz	2.4	±4.3	wf	4.8	±4.5			

Table 5: Biases (IGS minus IVS) and mean values of the hourly standard deviations in mm at collocated sites for the combined IVS and IGS time series. Although the height difference between the antennas is taken into account all biases are positive.

Although the standard deviations between the IVS and VLBI time series are at the ±5 mm level or even worse, it is noticeable that all mean values of the total zenith delays derived by GPS are larger than those derived by VLBI. The positive biases are between +1.4 mm (Medicina) and +13.5 mm (Fortaleza). This confirms first results reported by Boehm et al. (2002a, [1]). Apart the Pilot Project clearly showed that comparing and combining the results of several ACs which use different VLBI software or apply different analysis strategies allows

- to give feedback to the individual AC in case of any problems,
- to determine stable, robust and highly accurate final IVS products with standard devia-

tions that are usually significantly smaller than those of the individual submissions.

Zenith delays derived by VLBI can be compared to those derived by GPS and WVR. The always positive and almost constant biases between the GPS and VLBI time series at collocated sites need to be investigated in more detail.

The other field of application for zenith delays derived by VLBI is the contribution to climatological studies, at least when the time series cover a longer time interval. First results are reported by Boehm et al. (2003, this issue, [3]).

Acknowledgements

The authors would like to thank all IVS Analysis Centers who contributed to the Pilot Project and to all members of the Pilot Project Group.

References

[1] *Boehm, J., H. Schuh and R. Weber:* Influence of tropospheric zenith delays obtained by GPS and VLBI on station heights. In: Proceedings of the IAG Symposium on Vertical Reference Systems, Cartagena, Colombia, Feb. 2001, edited by H. Drewes, A. Dodson, L. Fortez, L. Sanchez, P. Sandoval, Springer Verlag Berlin-Heidelberg, 2002a.

[2] *Boehm, J., E. Messerer and H. Schuh:* Comparison of Tropospheric Parameters Submitted to the 2nd IVS Analysis Pilot Project. In: IVS 2002 General Meeting Proceedings, edited by N. R. Vandenberg and K. D. Baver, NASA/CP-2002-210002, 340–344, 2002b.

[3] *Boehm, J., H. Schuh, V. Tesmer and H. Schmitz-Hübsch:* Determination of tropospheric parameters by VLBI as a contribution to climatological studies, this issue, 2003.

[4] *Gendt, G.:* Comparisons of IGS tropospheric estimates. Proceedings IGS Analysis Center Workshop, 19-21 March 1996 Silver Spring, Maryland USA, Eds. R. E. Neilan, P. A. Van Scoy, J. F. Zumberge, 151–164, 1996.

[5] *Niell, A.E.:* Global mapping functions for the atmosphere delay at radio wavelength, J. Geophys. Res., 101 (B2), 3227-3246, 1996.

[6] *Niell, A.E.:* Preliminary evaluation of atmospheric mapping functions based on numerical weather models, Phys. Chem. Earth, 26, 475–480, 2001.

[7] *Saastamoinen, J.:* Contributions to the Theory of Atmospheric Refraction, Part II, Bulletin Geodesique, Vol. 107, 13–34, 1973.

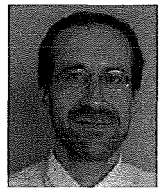
[8] *Schuh, H. and J. Boehm:* Status Report of the IVS Pilot Project - Tropospheric Parameters in International VLBI Service for Geodesy and Astrometry 2002 Annual Report, edited by N. R. Vandenberg and K. D. Baver, NASA/TP-2003-211619, 2003.

Contact

Univ.-Prof. Dr. Harald Schuh, Dipl.-Ing. Johannes Boehm: Institute of Geodesy and Geophysics, Gußhausstraße 27–29, A-1040 Wien. email:
hschuh@luna.tuwien.ac.at, jboehm@luna.tuwien.ac.at

Determination of Tropospheric Parameters by VLBI as a Contribution to Climatological Studies

Johannes Boehm and Harald Schuh, Wien, Volker Tesmer and Harald Schmitz-Huebsch, München

Abstract

As consistent VLBI observations at various stations over the whole globe have been carried out since 1984, it is possible to determine long time series not only of baseline vectors and Earth orientation parameters, but also of tropospheric parameters. Time series of wet zenith delays provide information about trends and periodic variations of the amount of water vapour in the troposphere. At Wettzell (Germany) there is a trend of ~+0.7 mm/year in the wet zenith delay which corresponds to ~+0.1 mm/year precipitable water vapour. Additionally, periodic variations in the time series are revealed by Fourier and wavelet analyses, and information about the precipitable water provided by the ECMWF (European Centre for Medium-Range Weather Forecasts) is used to evaluate the VLBI estimates.

Kurzfassung

Erst in den letzten Jahren wurde erkannt, dass die troposphärischen Laufzeitverzögerungen, denen die Signale der VLBI (Very Long Baseline Interferometry) und GPS unterworfen sind, nicht nur Störgrößen bei der Bestimmung geodätischer Parameter (Stationskoordinaten, Erdorientierungsparameter, ..) sind, sondern auch wertvolle Informationen für Meteorologie und Klimatologie liefern können. Zum Beispiel lässt sich aus dem feuchten Anteil der Laufzeitverzögerung in Zenitrichtung der Wasserdampfgehalt über der Station mit hoher Genauigkeit bestimmen. Im Gegensatz zu GPS ist eine Auswertung der VLBI-Experimente in genäherter Echtzeit noch nicht möglich; andererseits aber überdecken die zur Verfügung stehenden konsistenten VLBI-Reihen troposphärischer Parameter einen erheblich längeren Zeitraum. Für manche Stationen existieren Zeitserien seit Beginn der 80er Jahre. Daraus können langzeitliche Trends bestimmt werden und somit auf eine Zu- oder Abnahme des Feuchtegehalts der Troposphäre geschlossen werden. An der Station Wettzell (Bayerischer Wald, Deutschland) wurde der Trend für die letzten 20 Jahre zu ~+0.7 mm/Jahr bestimmt, was einer Zunahme des ausfällbaren Wassers von ~+0.1 mm/Jahr entspricht. Dies stimmt wiederum sehr gut mit der am Boden gemessenen durchschnittlichen Temperaturzunahme von +0.13 °C/Jahr an der Station Wettzell überein, da eine höhere Temperatur der Troposphäre auch eine erhöhte Speicherung von Wasserdampf erlaubt. Zusätzlich werden periodische Variationen in den Zeitserien mit Fourier- und Waveletanalysen ermittelt. Dabei zeigen sich neben den zu erwartenden saisonalen Schwankungen auch andere Perioden, die je nach Station unterschiedlich stark ausgeprägt sind. Schließlich werden die VLBI-Ergebnisse der Feuchte mit Daten des ECMWF (European Centre for Medium-Range Weather Forecasts) verglichen, wobei eine sehr gute Übereinstimmung zu erkennen ist.

1. Introduction

The total path delay for an observation at the elevation angle (consists of the hydrostatic and the wet part. Each of these parts is the product of the delay in zenith direction and the corresponding mapping function. Assuming azimuthal symmetry at a VLBI station, the total path delay in the neutral atmosphere (L(() can therefore be modelled as:

$$\Delta L(\varepsilon) = HZD \cdot mf_h(\varepsilon) + WZD \cdot mf_w(\varepsilon) \quad (1)$$

HZD hydrostatic zenith delay
WZD wet zenith delay
$mf_h(\varepsilon)$ hydrostatic mapping function
$mf_w(\varepsilon)$ wet mapping function

In standard VLBI analyses, the wet zenith delay (WZD) is estimated, while the other three parameters (HZD, mfh, mfw) are assumed to be known. Since consistent VLBI observations have been carried out for about 20 years, long time series of the wet zenith delays at various stations can be determined and used for climatological studies. Table 1 gives an overview of the VLBI stations that have been used for these investigations. On the average, 24 h geodetic VLBI sessions have been performed every 4th to 5th day, which yields a temporal coverage between 19% and 25%.

Station	latitude	temporal coverage	1st observation in the year
Wettzell, Germany	49°	25%	1984
Fortaleza, Brazil	−4°	20%	1993
Westford, Mass., U.S.A.	43°	19%	1984
Kokee Park, Hawaii, U.S.A.	22°	24%	1993
Gilcreek, Alaska, U.S.A.	65°	21%	1984

Table 1: Overview of selected VLBI stations, their latitudes, temporal coverage by VLBI sessions and the year of the first observation that was included in the analyses.

In VLBI analysis, wet zenith delays are estimated in the least-squares fit for each station of the observing session with a temporal resolution of 1 or 2 hours. The accuracy level of the absolute values is at about ± 5 mm [1]. In contrast to GPS, meteorological parameters are recorded at all VLBI stations, which is very valuable if we want to separate the hydrostatic and wet delays.

The contribution of GPS-derived wet zenith delays to climatology derived from nearly continuous GPS observations since 1994 has been reported recently [3], [2]. Since the spatial coverage of these observations is much denser than that of VLBI, it allows also regional studies. However, GPS-derived wet zenith delays can suffer from antenna phase center variations, multipath effects and the replacement of antennas or radomes. Thus, a comparison with wet zenith delays determined by VLBI at selected stations seems adviseable, in particular because a better long-term stability of the latter can be assumed due to the higher stability of the celestial and terrestrial reference frames used in VLBI.

2. Accuracy of the terrestrial reference frame

In order to detect significant trends in the wet zenith delays, the terrestrial reference frame has to be sufficiently accurate. This requirement is above all due to the high correlation of about −0.4 between station heights and zenith path delays, i.e. if a station height is wrong by +10 mm, the zenith path delay at this site will be shifted by about −4 mm (see Figure 1). If one assumes that the station coordinates of Fortaleza are error-free, that both stations are fixed in the analysis and that the observation is taken in zenith direction, an error of the station height coordinate of Wettzell (vertical arrow) will be fully transferred (with opposite sign) into an error of its wet zenith delay. As in typical VLBI sessions the observations are taken at elevations down to ~5°, the correlation decreases from −1 to about −0.4.

As this paper focuses on linear trends and periodic variations of the wet zenith delays rather

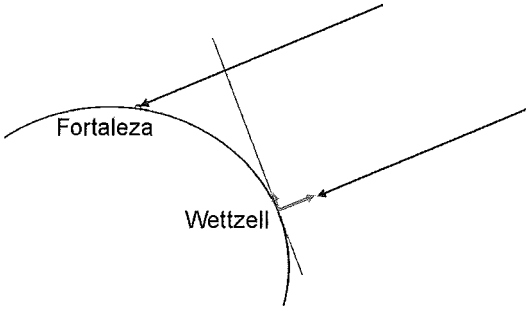


Figure 1: Geometry of a VLBI observation. The arrows at Wettzell mark horizontal and vertical errors in the station coordinates (see text).

than on absolute values, the station velocities and their standard deviations are of primary importance in this context. Two different terrestrial reference frames were applied to check the impact of their differences on the trends observed in wet zenith delays. In addition to the ITRF2000, which is a combined solution of VLBI, GPS, SLR, and DORIS measurements, a terrestrial reference frame purely determined by VLBI (DGFI02R02) was used for the analyses of the VLBI sessions. While the imprecision of the DGFI02R02 velocities is about ± 0.1 mm/year and that of the ITRF2000 is about ± 0.5 mm/year, the differences in station height velocities between both realizations do not exceed 0.8 mm/year for the subset of stations treated here (Table 2). Thus, ± 0.8 mm/year can be considered as a rough estimate of the inaccuracy of the terrestrial reference frame.

Gilcreek	Kokee Park	Westford	Fortaleza	Wettzell
0.5	−0.5	0.8	0.8	0.1

Table 2: Differences in station height velocities between ITRF2000 and DGFI02R02 in mm/year (ITRF2000 minus DGFI02R02).

The maximum deviation of 0.8 mm/year in station height velocity corresponds to about −0.3 mm/year in the wet zenith delay. Thus, if the linear trend in WZD exceeds ~0.3 mm/year,

it can be assumed as significant as far as the accuracy of the reference frame is concerned. To check this statement, different analysis strategies were compared: Fixing the coordinates to ITRF2000 and DGFI02R02 and calculating free network solutions with respect to both terrestrial reference frames yields similar trends for the wet zenith delays which will be presented in the following section.

3. Analysis and results

For this investigation, all 24 h geodetic VLBI sessions were analyzed that have been carried out since 1984. The VLBI software package OCCAM V 5.1 (Titov et al., 2001, [4]) was applied using the Gauss-Markov model for the least-squares adjustment. The wet zenith delays were estimated as 1 h piecewise linear functions, the elevation angle cutoff was set to 8°, and the ITRF2000 was fixed.

3.1. Linear long-term trends in wet zenith delays

Six-hour values of the wet zenith delays were extracted by interpolating between the two closest hourly estimates (Figure 2a) to allow comparison with meteorological data from numerical weather models (Figure 4a,b). In these data, e.g. at Wettzell (Germany) a linear trend was estimated to 0.83 mm/year and a big seasonal variation can be seen ranging from 0 mm (on some winter days) to 200 mm (on some summer days).Then mean seasonal values were determined. On the basis of these seasonal values, the overall rate of the wet zenith delays was estimated to +0.7 mm/year at Wettzell (Figure 2b). It is slightly different from the trend of the original time series, due to the different averaging processes within the computation of the seasonal values. The trend at Gilcreek (Alaska) for the time period 1989–2001 was determined to +0.3 mm/year (Figure 2d). Following the conclusion of section 2, the trend at Wettzell is significant, i.e. above the possible influence of the chosen reference frame. Multiplication of the observed rate by the length of the time series yields a change of 12.6 mm in 18 years. For the other VLBI stations the determination of reliable linear trends was not possible because either the time series were too short (Fortaleza, Kokee Park) or the seasonal wet zenith delays were too noisy (Westford). Figure 2c shows the averaged winter

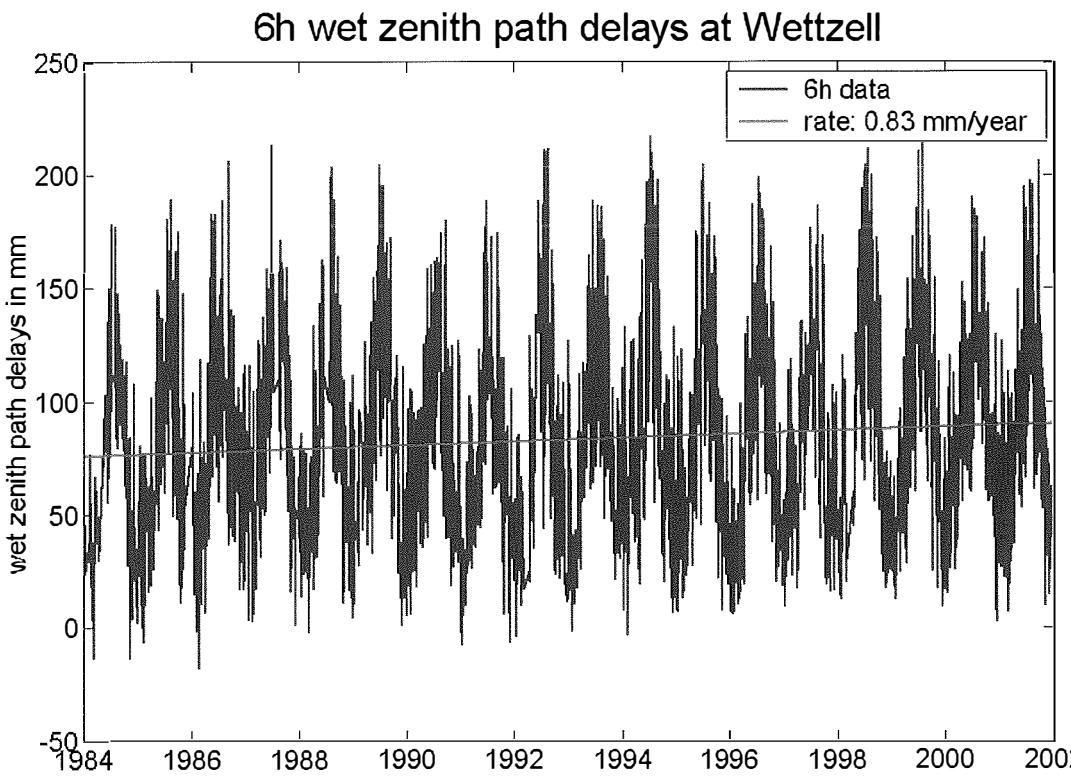


Figure 2a: 6 h wet zenith delays at Wettzell since 1984.

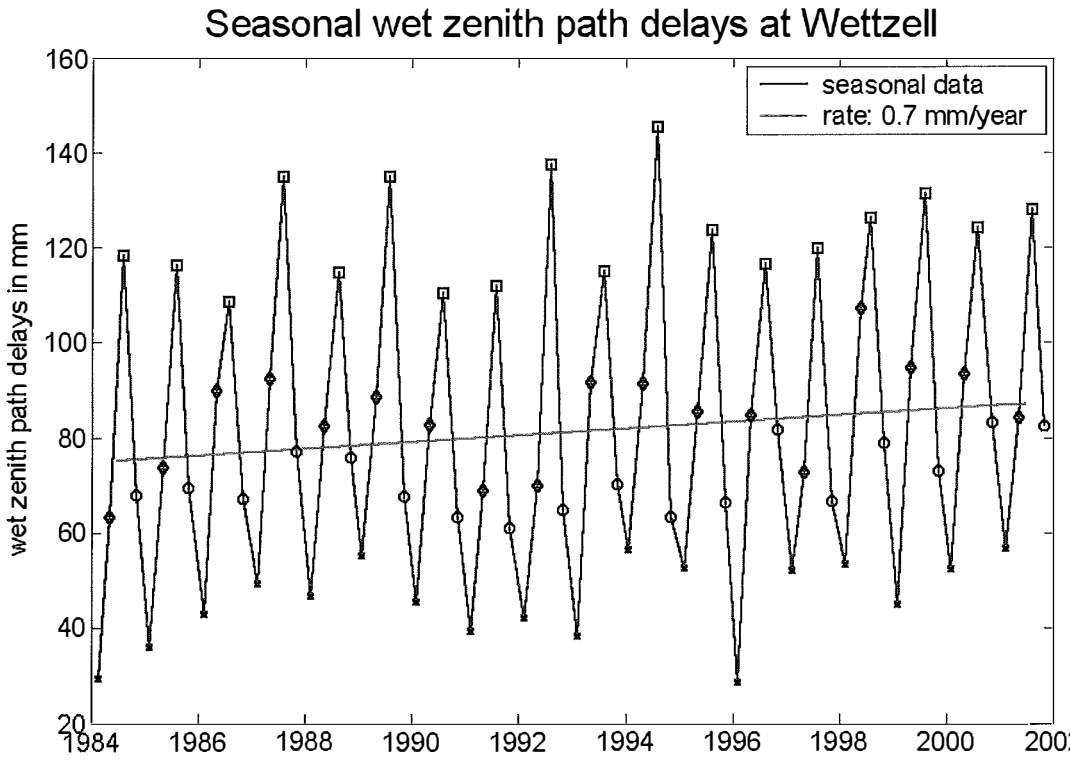


Figure 2b: Mean seasonal values of the wet zenith delays at Wettzell. The linear trend is estimated to 0.7 mm/year. Different markers are used for the seasons (◊ spring, □ summer, ○ autumn, × winter).

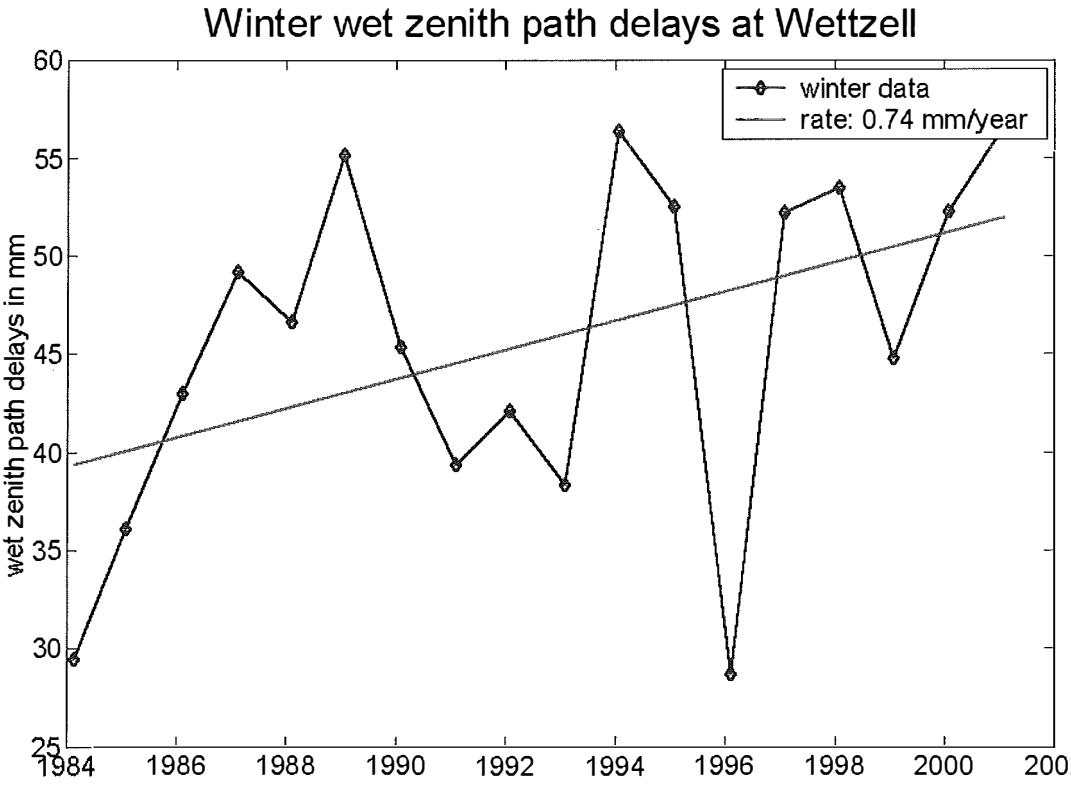


Figure 2c: Wet zenith delays at Wettzell in winter. In 1983/84 and 1995/96 the winters were extraordinarily dry (= cold).

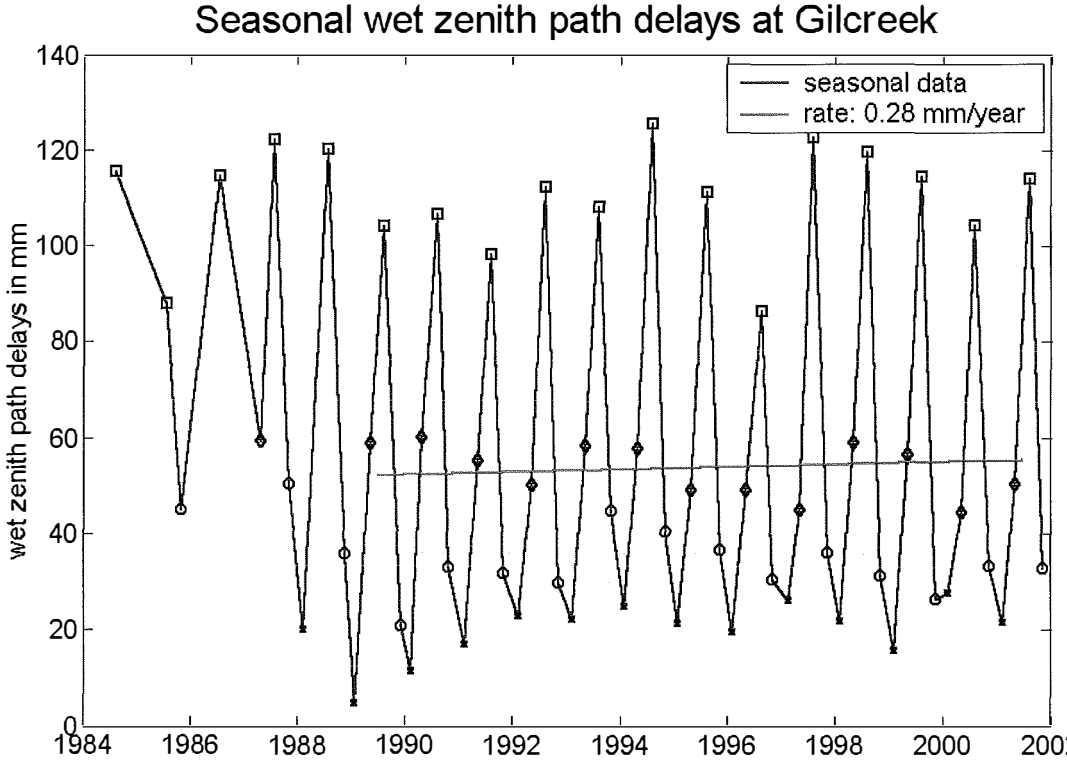


Figure 2d: Mean seasonal values of the wet zenith delays at Gilcreek. The linear trend is estimated to 0.3 mm/year. Different markers are used for the seasons (◊ spring, □ summer, ○ autumn, × winter).

WZD values at Wettzell with a linear trend of 0.74 mm/year. The lowest average WZD were obtained for winter 1983/84 and 1995/96. Meteorological records at Wettzell station confirm that these winters were extraordinarily cold and dry.

3.2. Climatological interpretation of trends in the wet zenith delays

Although wet zenith delays cannot be directly derived from meteorological data recorded at a site, there are equations that yield approximate values, e.g. by Moran et al. (2001, [6]):

$$WZD \approx \frac{e}{T^2}\ [m], \qquad (2)$$

where e is the water vapour pressure in hPa and T is the temperature in K. The VLBI databases comprise information about the temperatures and the relative humidities recorded close to each radiotelescope. At Wettzell, since 1984 the relative humidity has been rather constant at about 80% whereas the temperature has increased by about 0.13 K/year. The relative humidity f is defined by

$$f = \frac{e}{E(T)}. \qquad (3)$$

As the saturated water vapour pressure E(T) is increasing with rising temperature, the water vapour pressure e is increasing with rising temperature, too, if f is kept constant. Although (2) implies that the wet zenith delays are decreasing with increasing temperature, the influence of the increase in e (see (3)) is dominating over this effect. Using a mean temperature of 15 °C and a mean relative humidity of 80%, (3) and (2), applied for an increase in the temperature of 0.13 K per year, yield an increase in the wet zenith delay of 0.9 mm/year, what is close to the results from VLBI (0.7 mm/year to 0.8 mm/year).

t in °C	e in hPa (see (3))	T^2 in K^2	WZD in mm (see (2))
15.00	13.635	83030	124.8
15.13	13.750	83105	125.7

Table 3.:Change in the WZD after one year, when the temperature is rising by 0.13 K and the relative humidity is constant at 80 %.

3.3. Periodic variations in the wet zenith delays

Classical Fourier analyses and wavelet transformations of the six-hour time series were per-

formed to find periodic variations of the wet zenith delays. The Fourier spectra show wide peaks at the annual periods for tropical stations (Fortaleza, Kokee Park) (Figure 3b, plots at left side) and sharp peaks for stations in mid-latitudes (Wettzell, Westford). This is due to the fact that there are no pronounced differences between the seasons in the tropics, while the large seasonal differences of the temperatures for continental stations are mirrored in the strong annual variations of the wet zenith delays.

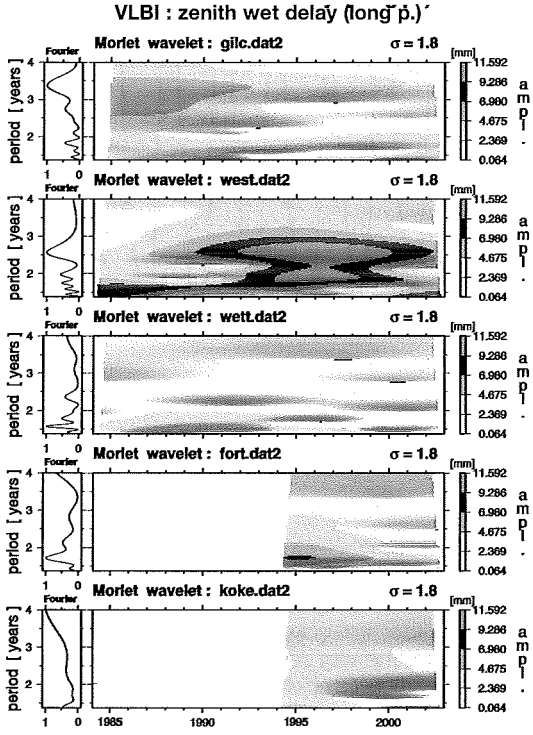


Figure 3a: Fourier and Morlet wavelet spectra for periods between 500 and 1460 days for the VLBI stations Gilcreek (gilc.dat2), Westford (west.dat2), Wettzell (wett.dat2), Fortaleza (fort.dat2) and Kokee Park (koke.dat2).

The wavelet analyses do not only provide information about the main periods of the wet zenith delays but also about their temporal variations:

- strong annual periods at all stations with variable amplitudes (Figure 3b),
- irregular variations at 1.6–1.7 years (Figures 3a),
- irregular variations with periods between 30 and 90 days (Figure 3c);

the strongest of these variations occurred at Westford, in particular with periods shorter than 50 days.

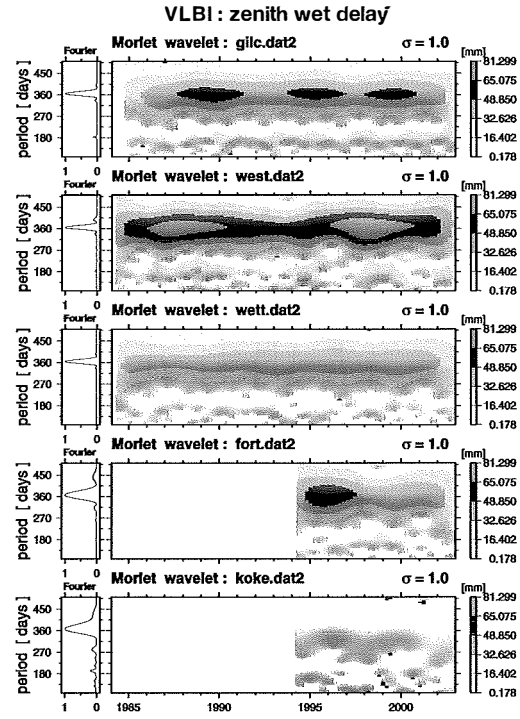


Figure 3b: Fourier and Morlet wavelet spectra for periods between 100 and 500 days for the VLBI stations Gilcreek (gilc.dat2), Westford (west.dat2), Wettzell (wett.dat2), Fortaleza (fort.dat2) and Kokee Park (koke.dat2).

4. Comparison with ECMWF data

The European Centre for Medium-Range Weather Forecasts (Reading, UK) provides meteorological data at six-hour intervals. The precipitable water is the parameter that is comparable to the wet zenith delay WZD. Firstly, the wet zenith delay has to be transformed into the integrated water vapour IWV (units kg/m^2):

$$IWV = WZD \cdot \Pi \tag{4}$$

The parameter Π is as follows:

$$\Pi = \frac{10^6 \cdot M_w}{\left(k_2' + \frac{k_3}{T_m}\right) \cdot R} \tag{5}$$

where

$$M_w = 18.0152 \frac{kg}{kmol}$$

$$k_2' = 17 \pm 10 \frac{K}{hPa}$$

$$k_3' = 373900 \pm 1200 \frac{K^2}{hPa}$$

$$R = 8314.34 \frac{J}{kmol \cdot K}$$

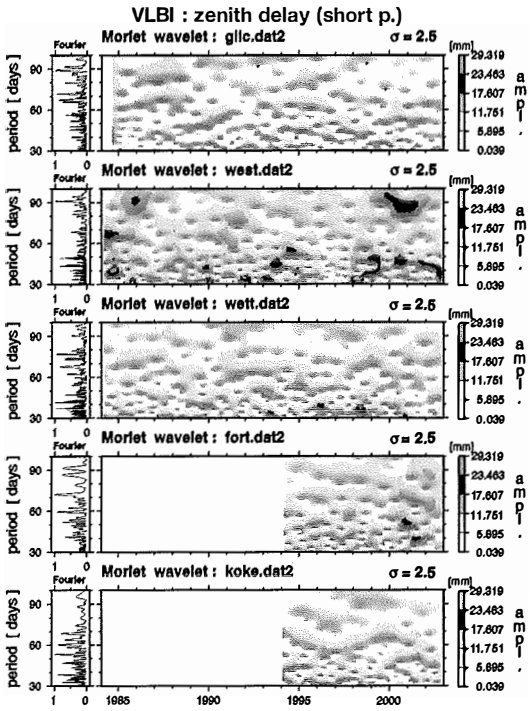


Figure 3c: Fourier and Morlet wavelet spectra for periods between 30 and 100 days for the VLBI stations Gilcreek (gilc.dat2), Westford (west.dat2), Wettzell (wett.dat2), Fortaleza (fort.dat2) and Kokee Park (koke.dat2).

M_w is the molar mass of water, k_2' and k_3 are empirically determined coefficients, T_m is the mean temperature above the station, and R is the general gas constant. With the density of liquid water ρ_w, the precipitable water PW (units: m) can be determined:

$$PW = IWV/\rho_w \quad (6)$$

The precipitable water can be approximated by the formula

$$PW \approx 0.15 \cdot WZD. \quad (7)$$

The comparison between precipitable water from VLBI and ECMWF, which is available since 1994, shows a very good agreement at the level of ± 1.85 mm corresponding to a zenith delay of ~12 mm (Figures 4a,b). Compared to the standard deviation of ± 1.85 mm (precipitable water) the bias between the time series is very small (0.44 mm PW or 3 mm WZD).

5. Conclusions and outlook

The investigations presented here reveal a systematic increase of the wet zenith delays at Wettzell in the past two decades. This trend is significantly above the potential influence of the chosen terrestrial reference frame. Thus, the results obtained from VLBI might be useful for cli-

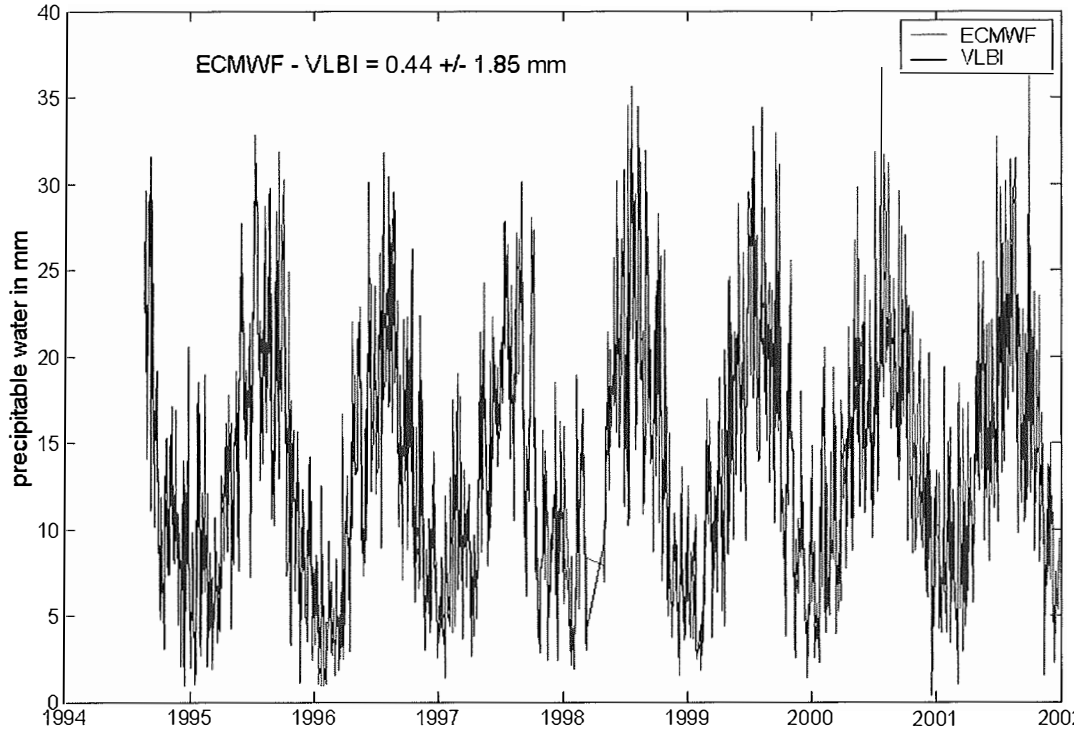


Figure 4a: Precipitable water from ECMWF and VLBI at Wettzell (1994–2002).

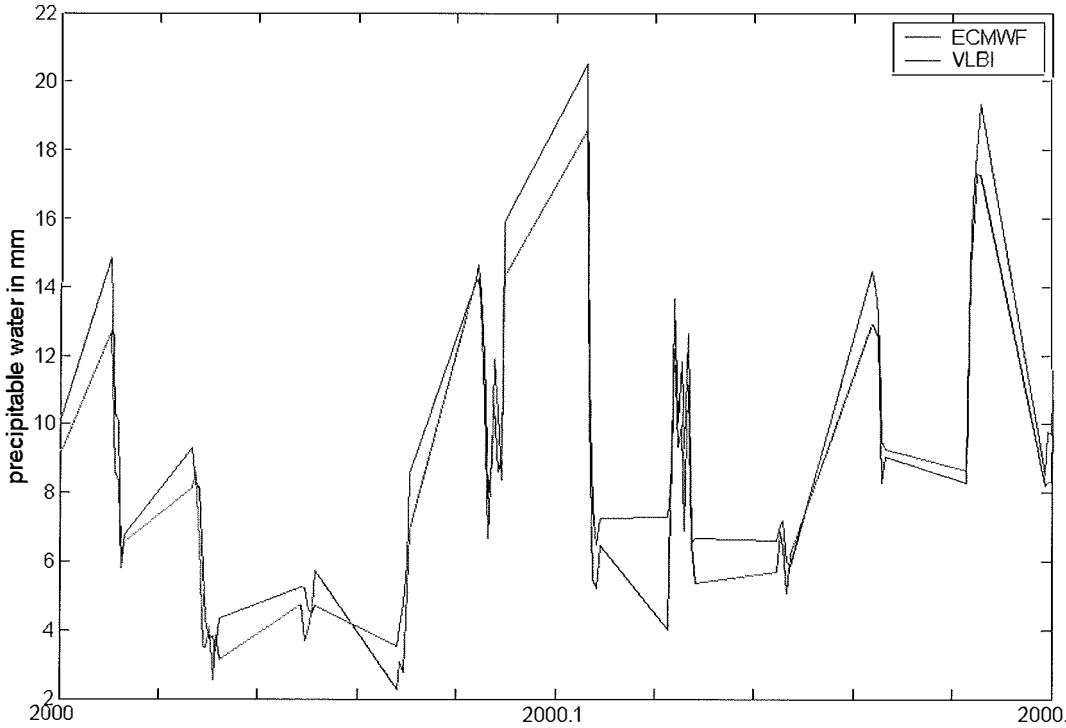


Figure 4b: Precipitable water from ECMWF and VLBI at Wettzell (2000.0 – 2000.2). ECMWF values are plotted only when VLBI values are available.

matological studies. A closer look remains to be taken at the other VLBI sites to possibly detect similar features in the time series of the tropospheric parameters. Similarily to the comparison with data from ECMWF, the tropospheric zenith delays can be compared with those provided by IGS [5]. Moreover, GPS-derived zenith delays can be used to fill the gaps between the results of VLBI and to finally obtain a robust combined time series.

Acknowledgements

We would like to thank the Zentralanstalt für Meteorologie und Geodynamik (ZAMG), Austria, for giving us access to the ECMWF data and to Burghard Richter (DGFI, Munich) for giving valuable hints to this publication.

References

[1] *Niell, A.E., A.J. Coster, F.S. Solheim, V.B. Mendes, P.C. Toor, R.B. Langley, C.A. Upham,* Comparison of Measurements of Atmospheric Wet Delay by Radiosonde, Water Vapor Radiometer, GPS and VLBI, Journal of Atmospheric and Oceanic Technology, 18, pages 830 – 850, 2001.

[2] *Gradinarsky L.P., J.M. Johansson, H.R. Bouma, H.-G. Scherneck, G. Elgered,* Climate Monitoring using GPS, Physics and Chemistry of the Earth, 27, pages 335 - 340, 2002.

[3] *Eyer M.,* Der Einsatz des GPS zur Wasserdampfmodellierung in der Klimatologie, Satelliten-Beobachtungsstation Zimmerwald, Bericht Nr. 28, 1999.

[4] *Titov O., V. Tesmer, J. Boehm,* OCCAM Version 5.0 Software User Guide, AUSLIG Technical Report 7, 2001.

[5] *Schuh, H., J. Boehm,* Determination of Tropospheric Parameters within the IVS Pilot Project, this issue, 2003.

[6] *Thompson, A.R., J.M. Moran, G.W. Swenson jr.,* Interferometry and Synthesis in Radio Astronomy, John Wiley and Sons, 2001.

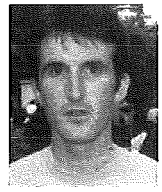
Contact

Dipl.-Ing. Johannes Boehm, Univ.-Prof. Dr. Harald Schuh: Institute of Geodesy and Geophysics (IGG), Vienna University of Technology, Gusshausstrasse 27–29, A-1040 Vienna. email: jboehm@luna.tuwien.ac.at, hschuh@luna.tuwien.ac.at

Dipl.-Ing. Volker Tesmer, Dr. Harald Schmitz-Huebsch: Deutsches Geodaetisches Forschungsinstitut (DGFI) Marstallplatz 8, D-80539 Munich, Germany. email: vtesmer@dgfi.badw.de, schmihue@dgfi.badw.de

VLBIONOS – Probing the Ionosphere by Means of Very Long Baseline Interferometry

Thomas Hobiger, Johannes Boehm and Harald Schuh, Wien


Abstract

In geodetic Very Long Baseline Interferometry (VLBI) the observations are performed at two distinct frequencies (2.3 and 8.4 GHz) in order to determine ionospheric delay corrections. This allows information to be obtained from the VLBI observables about the sum of electrons (total electron content – TEC) along the ray path through the ionosphere. Due to the fact that VLBI is a differential technique, only the differences in the behavior of the propagation media over the stations determine the values of the observed ionospheric delays. However, in a first simple approach, an instrumental delay offset per baseline shifts the TEC measurements by a constant value. This offset is independent of the azimuth and elevation of the observed radio source and allows separation of the ionospheric parameters for each station from the instrumental delay offsets per baseline in a least-squares adjustment. In first tests of this method Fourier coefficients up to the 4th order plus a constant value and a linear trend were estimated to represent the vertical TEC (VTEC). Slant TEC (STEC) values are converted into VTEC values by a mapping function. A disadvantage of this approach is the assumption that these values are assigned to the station coordinates but not to the geographical coordinates of the intersection point of the ray path and the infinitely thin ionospheric layer. The precision of the estimated values is about +/– 5 to 7 TEC units (TECU). The results obtained from VLBI agree with a standard deviation of +/– 10 TECU with other techniques like GPS, rarely exceeding 20 TECU. A second approach, developed at the TU Vienna, using piece-wise linear functions (VTM – Vienna TEC model) was also tested.

Kurzfassung

Aufbau, Beschaffenheit, geographische und zeitliche Veränderungen der Ionosphäre (grob gesagt der Bereich der Erdhülle zwischen 50 km und 1000 km) sind für Meteorologen und Klimaforscher ein wichtiger Untersuchungsgegenstand. Aber auch für die Geodäsie spielt die Ionosphäre eine immer bedeutendere Rolle. Einerseits werden Signale im Radiofrequenzbereich durch die Ionosphäre derart abgelenkt bzw. verzögert, daß z.B. hochgenaue GPS-Messungen nur durch Beobachtung auf zwei Frequenzen möglich sind und es trotzdem während der ungefähr alle 11 Jahre auftretenden Perioden starker Sonnenaktivität zu spürbaren Genauigkeits- und Qualitätseinbußen der GPS-Ergebnisse kommt. Andererseits ist es heutzutage möglich, aus global verteilten GPS-Messungen wie sie z.B. im Rahmen des International GPS Service (IGS) durchgeführt werden, Informationen über den Zustand und die kurz- und langfristigen Veränderungen der Ionosphäre zu gewinnen. Im vorliegenden Artikel wird erstmals gezeigt, daß dies auch mit dem Verfahren der Very Long Baseline Interferometry (VLBI) möglich ist. Es sollen erste Ergebnisse präsentiert werden, die im Rahmen eines vom österreichischen Fonds zur Förderung der wissenschaftlichen Forschung (FWF) seit März 2003 unterstützten Forschungsprojekts erzielt wurden. Bei dem Verfahren der Radiointerferometrie auf langen Basislinien (VLBI) wird auf zwei unterschiedlichen Frequenzbändern (2.3 und 8.4 GHz) beobachtet, um die Laufzeitverzögerung zu bestimmen, die durch die Ionosphäre verursacht wird. Dadurch können Rückschlüsse auf den Gesamtelektronengehalt (TEC) entlang des Ausbreitungsweges der Welle gezogen werden. Allerdings lassen sich in einem einfachen Ansatz nur die Differenzen in der Beschaffenheit der Ionosphäre über den einzelnen Stationen bestimmen, und die Beobachtungen sind zusätzlich noch um einen durch instrumentelle Einflüsse hervorgerufenen konstanten Wert verfälscht. Da jedoch in unterschiedlichen Azimuten und Elevationen beobachtet wird, gelingt mittels spezieller Methoden durch eine Parameterschätzung nach der Methode der kleinsten Quadrate eine Trennung von den instrumentellen Einflüssen und somit eine Bestimmung der absoluten ionosphärischen Parameter. In ersten Analysen wurde der vertikale Gesamtelektronengehalt in Form von Fourieransätzen (bis Grad 4) geschätzt. Dabei wurde vereinfachend angenommen, daß alle Beobachtungen in Zenitrichtung durchgeführt wurden. Die innere Genauigkeit der VLBI-Ergebnisse wird zu +/– 5-7 TEC Units (TECU) geschätzt. Trotz der erwähnten Approximation stimmen die Ergebnisse auch mit denen von GPS innerhalb von +/– 10 TECU überein mit maximalen Abweichungen von 20 TECU. Ebenfalls erprobt wurde ein zweiter, an der TU Wien entwickelter Ansatz mit stückweise linearen Funktionen (VTM – Vienna TEC Model).

1. The Earth's ionosphere

The Earth's ionosphere is defined as that part of the upper atmosphere where free electrons occur in sufficient density to influence the propagation of electromagnetic radio frequency waves. The ionization depends primarily on the Sun and its activity. Ionospheric structures vary strongly with time corresponding to the sunspot cycle and seasonal and diurnal cycles and with geographical location (polar, auroral, mid- latitude and equatorial regions). Certain further ionospheric disturbances can be related to the Sun. The major part of the ionization is produced by solar X-ray and ultraviolet radiation and by corpuscular radiation from the Sun. The most noticeable effect is seen as the Earth rotates with respect to the Sun. Ionization increases in the sunlit hemisphere and decreases on the shadowed side. Although the Sun is the largest contributor to the ionization, cosmic rays are the source of a small contribution, too.

The ionosphere is a dynamic system depending on many parameters, including acoustic motions of the atmosphere, electromagnetic emissions, and variations of the geomagnetic field. Any atmospheric disturbance affects the distribution of the ionization, and, because of its extreme sensitivity to atmospheric changes, the ionosphere can be used as a sensor of atmospheric events.

1.1. The structure of the ionosphere

1.1.1. Vertical profile of the ionosphere

Figure 1 gives schematically an overview of the vertical structure of the ionosphere, which can be vertically divided into four broad regions called D, E, F, and topside. These regions may be further divided into several regularly occurring layers, such as F1 or F2. The region between about 75 and 95 kilometers above the Earth, where the relatively weak ionization is mainly responsible for absorption of high frequency radio waves, is called the D region. The E region extends from 95 to 150 kilometers above the surface of the Earth. Other subdivisions isolating separate layers of irregular occurrence within this region are also labeled with an E prefix, such as the thick layer E2, and the highly variable, thin and sporadic Es layer. Ions in this region are mainly of type O_2^+ (ionized oxygen).

The F region is located about 150 kilometers above the E layer. In this region, the important reflecting layer F2 can be found. Other layers within this region are also described by the prefix F, such as a temperate latitude regular stratification F1. Ions in the lower part of the F layer are mainly of the type NO^+, in the upper part mainly O^+. The F layer is the region of primary interest for radio communications because this region is responsible for the reflection of radio waves that allows travel over great distances that cannot be achieved by direct radio waves due to the curvature of the Earth. Topside is the part of the ionosphere starting at the height of the maximum density of the F2 layer and extends upwards with decreasing density to a transition height where O^+ ions become less numerous than H^+ and He^+ ions. The transition height, which is usually above 1000 kilometers, varies by time but seldom drops below 500 kilometers at night or 800 kilometers in the daytime. Above the transition height the weak ionization has small influence on transionospheric radio signals.

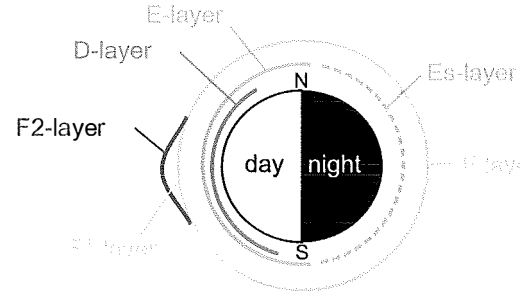


Figure 1: Ionospheric layers

1.1.2. Main geographical regions of the ionosphere

The ionosphere can be divided into two major geographical regions corresponding to the two principal regimens of magnetospheric circulation [1]: the low and mid latitude regions (less than 60° geomagnetic latitude) and the high latitudes. Figure 2 schematically shows these main geographic regions, although there is no well-defined boundary between these two parts. In the figure the day/night boundaries and the continents are represented for 12 h UT. The low latitude or equatorial region is characterized by high numbers of free electrons and by large spatial gradients. In this region the geomagnetic anomaly takes place. The mid latitude regions are those that present the more regular and predictable variations, although magnetic storms can cause nearly daily changes of about 30% in the total electron content (see section 1.2). The high latitudes or polar regions are characterized by spatial and temporal variations that are un-

predictable. The reason is that in these regions the magnetic field intensity is three times larger than at the equator and the direction of its magnetic force is almost vertical. Therefore, solar wind particles that are mainly protons are accelerated down to very low altitudes, ionizing atmospheric atoms and molecules by collision.

spots are relatively cold regions of the Sun's photosphere. Due to the fact that their temperature is lower than that of the adjacent regions, they appear as dark areas in optical observations. Sunspots are associated with the most impressive events that take place on the Sun, the flares. These events occur more frequently dur-

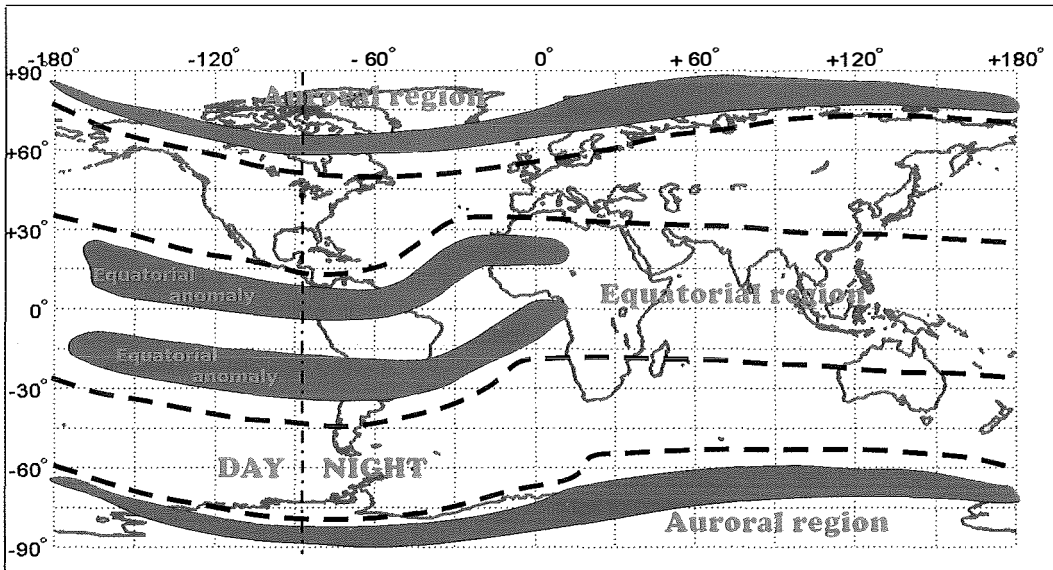


Figure 2: Main geographical regions of the ionosphere

1.1.3. Temporal variations

Every year the Sun, after crossing the equator from South to North or vice versa, reaches a maximum declination of about 23° in the solistices. This motion of the Sun, of which the most perceptible effect is the seasonal climatic change, produces in the ionosphere a variation of the free electron distribution from South to North of the equator or vice versa. This is the so-called seasonal variation. The most important ionospheric temporal variation is caused by the solar cycle, although along this cycle the radiation intensity does not fluctuate by more than 1%. The sunspot number (more precisely, the Wolf number, defined as $R = k(f + 10g)$ where f is the total number of sunspots, g is the quantity of sunspot groups, and k is a constant that depends on the sensitivity of the observational instrument) has been utilized as an indicator of solar activity. Historical records dealing with the sunspot numbers show a cycle with an average period of 11.1 years, with fluctuations running from 7 years for the shortest cycles up to 17 years for the longest periods. The most recent sunspot maximum happened in 2001/2002. Sunning the solar cycle maxima, and they are less frequent during the minima. Additional quantities of protons, alpha particles, and ultraviolet and soft X-rays are emitted by the Sun and produce so-called magnetic storms. The X-rays immediately arrive at the Earth causing a rapid increase in the free electron density all over the illuminated hemisphere, which is known as sudden ionospheric disturbance. The other parts of the ejected matter arrive later and cause irregularities in the behavior of the ionosphere.

There are two other phenomena that have to be explained here: traveling ionospheric disturbances and scintillations. Traveling ionospheric disturbances are structures with a free electron density much greater than in the surrounding volume. Their size can vary from 50 to 500 km, and their velocity can reach several hundreds of kilometers per second with respect to the Sun-fixed coordinate system. This fact causes temporal variations which may last from a few minutes to several hours. Traveling ionospheric disturbances are more frequently found at mid-latitudes during the solar cycle maxima. Scintillation is a particular ionospheric disturbance consisting

of a quick change in the free electron distribution, both in space and time, which affects the electromagnetic wave propagation within the ionosphere and produces a fast and irregular variation of its phase and amplitude. This effect is similar to the one that can be found when the brightness of a star is being observed through a turbulent atmosphere.

Ionospheric variations can happen on long time scales from years to decades and also on short time scales of a few hours or even seconds. The investigation of the ionospheric behavior is a quite important field of research, e. g. because the ionosphere influences the propagation characteristics of radio signals used in space geodetic techniques.

1.2. Impact on space geodetic techniques

One of the problems for all space geodetic techniques operating with electromagnetic waves is the determination of the propagation velocity of the signals. If these waves propagated in vacuum, the traveled distance would be just the product of the propagation time between emitter and receiver and the speed of light in vacuum. When signals travel through the ionosphere, the interaction between the electromagnetic field and the free electrons influences both the speed and the propagation direction of the signal, an effect known as ionospheric refraction [2]. The ratio between the propagation speed of a wave in vacuum and the propagation speed in a given medium is known as the refractive index

$$n = \frac{c}{v_{ph}}, \tag{1.1}$$

The phase velocity is related to the angular frequency and the wave number

$$v_{ph} = \frac{\omega}{k} \tag{1.2}$$

A medium is called to be dispersive when the wave number, and therefore the phase velocity, depends on the angular frequency of the wave. A signal consisting of a modulated carrier wave can be considered to be the result of superimposing a group of different wavelengths centered at the frequency of the carrier. In a dispersive medium the phase propagates with different velocity depending on its frequency. The group velocity v_g is given by:

$$v_g = \frac{d\omega}{dk} = v_{ph} + k\frac{dv_{ph}}{dk} \tag{1.3}$$

The group refractive index n_g of the medium is defined as

$$n_g = \frac{c}{v_g} = c\frac{dk}{d\omega} = n + \omega\frac{dn}{d\omega} \tag{1.4}$$

The Appleton-Hartree theory [2] allows the calculation of the refractive index for a single wave which propagates through a plasma (= ionized medium) according to

$$n^2 = 1 - \frac{X}{1 - jZ - \left[\frac{Y_T^2}{2(1-X-jZ)}\right] \pm \sqrt{\frac{Y_T^4}{4(1-X-jZ)^2} + Y_L^2}} \tag{1.5}$$

where

$$j^2 = -1, X = \frac{\omega_p^2}{\omega^2}, Y_L = \frac{\omega_{SL}^2}{\omega^2}, Y_T = \frac{\omega_{ST}^2}{\omega^2}, Z = \frac{\omega_c^2}{\omega^2}, \tag{1.6}$$

In these equations ω represents the carrier frequency, ω_p is the frequency ω_p of the electron plasma, ω_s the synchrotron frequency of the electrons, and ω_c the so-called collision frequency. The indices L and T stand for the longitudinal and transverse components of the magnetic field with respect to the propagation direction of the wave. The plasma frequency can be written as

$$\omega_p = \sqrt{\frac{N_e e^2}{\varepsilon_0 m_e}} \tag{1.7}$$

where N_e is the free electron density in the medium, ε_0 the vacuum permitivity, m_e and e are the mass and charge of the electron. Therefore the plasma frequency ω_p in the ionosphere is approximately $5 \cdot 10^7\ s^{-1}$. The synchrotron frequency in the ionosphere is proportional to the Earth's magnetic field and inversely proportional to the mass of an electron

$$\omega_s = \frac{eB}{m_e} \tag{1.8}$$

The synchrotron and collision frequencies are much smaller than the carrier frequencies of space geodetic techniques such as VLBI, GPS, GLONASS, DORIS, and TOPEX/POSEIDON. Neglecting these two values, the Appleton-Hartree formula can be reduced to

$$n \approx \sqrt{1-X} = \sqrt{1 - \left(\frac{\omega_p}{\omega}\right)^2} = 1 - \frac{1}{2}\left(\frac{\omega_p}{\omega}\right)^2\left[1 + \frac{1}{4}\left(\frac{\omega_p}{\omega}\right)^2 + \ldots\right] \tag{1.9}$$

Dropping the terms of orders higher than $\left(\frac{\omega_p}{\omega}\right)^2$ we obtain a new approximation

$$n \approx 1 - \frac{1}{2}\left(\frac{\omega_P}{\omega}\right)^2 = 1 - \frac{1}{2}\frac{N_e e^2}{\varepsilon_0 m_e \omega^2} = 1 - \frac{40.28 N_e}{f^2} \tag{1.10}$$

where is f the carrier frequency expressed in Hz ($\omega = 2\pi f$). Using this result the group refractive index n_g can be determined by

$$n_g \approx 1 + \frac{1}{2}\frac{N_e e^2}{\varepsilon_0 m_e \omega^2} = 1 + \frac{40.28 N_e}{f^2} \tag{1.11}$$

Now we can calculate the carrier phase delay, d_{ph}, to obtain the influence of the ionized med-

ium on the propagation delay, expressed in SI units (meters)

$$d_{ph} = \int_S (n_{ph} - 1) dS = -\frac{40.28}{f^2} \int_S N_e dS = -\frac{40.28 \cdot 10^{16}}{f^2} STEC \quad (1.12)$$

The integral of the electron density along the signal path is usually called (Slant Total Electron Content). This quantity can be interpreted as the total amount of free electrons in a cylinder with a cross section of 1 m^2 of which the axis is the slant signal path. is measured in Total Electron Content Units (TECU), which is equivalent to 10^{16} electrons $\cdot$ m^{-2}. The effect of the ionized medium on group propagation can be expressed by

$$d_g = \int_S (n_g - 1) dS = \frac{40.28}{f^2} \int_S N_e dS = \frac{40.28 \cdot 10^{16}}{f^2} STEC \quad (1.13)$$

The last two expressions show how the electron content in the ionosphere can influence measurements ranging from outer space to Earth-based stations. If the behavior of the ionosphere is known, these effects can be computed and can be used to correct measurements on radio frequencies. If ionospheric parameters are not available, group delay observations have to be carried out on two different frequencies,

$$d_1 = d_0 + \frac{const}{f_1}, \; d_2 = d_0 + \frac{const}{f_2} \quad (1.14)$$

where d_0 is the distance in vacuum and const = $40.28 \cdot 10^{16}$ STEC. Combining the two equations (1.14) yields to the so-called ionosphere-free linear combination

$$d_0 = d_1 \frac{f_1^2}{f_1^2 - f_2^2} - d_2 \frac{f_2^2}{f_1^2 - f_2^2} \quad (1.15)$$

Thus, observing at two different radio frequencies allows the elimination of ionospheric influences.

2. Using space geodetic techniques to probe the ionosphere

2.1. GPS and the ionosphere

The NAVigation Satellite Timing and Ranging Global Positioning System (NAVSTAR GPS or shortly GPS) operates on two different frequencies (f_1, f_2) which are derived from the fundamental frequency f_0 = 10.23 MHz

$$f_1 = 154 \cdot f_0 = 1575.42 \, MHz, f_2 = 120 \cdot f_0 = 1227.60 \, MHz \quad (2.1)$$

Both code and phase measurements are affected by the dispersive behavior of the ionosphere, but with different leading signs which can be assigned to equations 1.12 and 1.13. Besides the ionosphere-free linear combination which eliminates the influence of the ionosphere, the so-called geometry free linear combination (also called L4-combination)

$$d_4 = d_1 - d_2 \quad (2.2)$$

allows an investigation of the total electron content along the ray path [3]. This term contains only the slant total electron content and the satellite and receiver differential code biases (DCBs), which are interpreted as the differential, inter-frequency hardware delays [4]. Calculating the geometry free linear combination, using both code and phase measurements, allows separation of the DCBs from the STEC values in a least-squares adjustment process.

The International GPS Service (IGS) has initiated a working group for the development of global ionospheric gridded data representing the total electron content over the whole globe. Several analysis centers deliver their results of vertical TEC values (VTEC) and DCBs in the IONospheric Exchange (IONEX) format [5], which represents the ionosphere as an infinitesimal shell in time intervals of two hours. Currently discrepancies exist in the solutions of the different analysis centers, and a combined solution is still missing [6].

2.2. A new field of research – ionospheric investigations by geodetic VLBI

2.2.1. How VLBI works

Very Long Baseline Interferometry (VLBI) is a space geodetic technique that allows measurement of distances of thousands of kilometers between stations of a global network with an error of a few millimeters. Radio signals from extragalactic radio sources (quasars or radio galaxies) are observed that have traveled several billion light years to the Earth. The VLBI technology provides precise station coordinates, and repeated measurements covering many years reveal plate tectonic rates and small changes in the Earth rotation with high accuracy. Radio sources are located almost infinitely far from the Earth, and the radio waves that leave these objects reach the Earth as plane wave fronts. If the signals are recorded at two radio telescopes at different positions, A and B (see figure 3), there is a time difference in receiving the same signal. This time difference, also called „delay", is measured with an accuracy of the order of 10^{-11} seconds. The product of this delay and the velocity of the radio wave (equal to the velocity of light) is called path difference, which allows calcula-

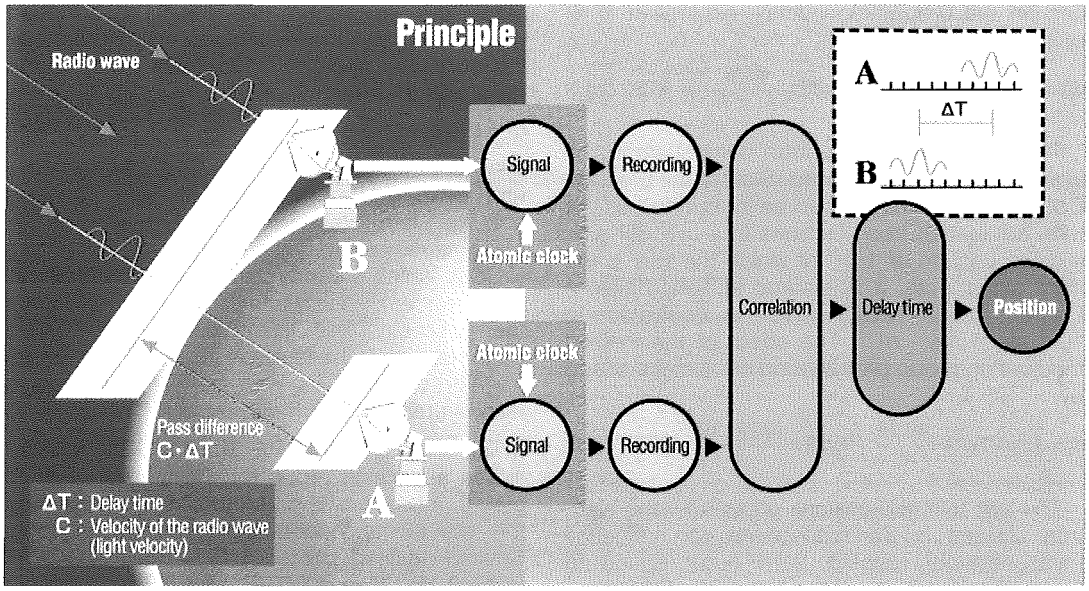


Figure 3: Principle of VLBI

tion of the distance between the two stations in the direction of the radio wave. The relative positions of A and B can be calculated by consecutively conducting these measurements in three or more directions.

2.2.2. VLBIonos

The project 'VLBIonos', supported by the Austrian Science Fund (FWF), started on March 1st, 2003, and aims at the investigation of the ionosphere using geodetic VLBI. VLBI observations are performed at two different frequencies (2.3 and 8.4 GHz, S- and X-band) in order to determine the ionospheric delay. As shown by 7] this information can be used to model the ionosphere above each station. It has to be mentioned that only the differences in the ionospheric delay between the two stations are measured. Unfortunately, instrumental offsets at each station bias these measurements, which leads us to

$$\tau_{model}(t) = \tau_{ion,1}(1) - \tau_{ion,2}(t) + \tau_{offset,1} - \tau_{offset.2} \qquad (2.3)$$

The ionospheric delay at X-band over station can be modeled as

$$\tau_{Ion,i}(t) = \frac{1.34 \cdot 10^{-7}}{f_x^2} \cdot S(E_i) \cdot VTEC_i(t) \qquad (2.4)$$

where

$$S(E_i) = \frac{1}{\cos\left\{\arcsin\left[\frac{R \cos E_i}{R+h}\right]\right\}} \qquad (2.5)$$

and $VTEC_i(t)$ represents the vertical TEC value at the intersection point of the ray path with the infinitesimally thin ionospheric layer assumed to be at the height h. The radius of the (spherical) Earth is abbreviated with R. Under the assumption that horizontal gradients in the ionosphere can be neglected within a range of about 300 kilometers, $VTEC_i(t)$ can be assigned to the geographical coordinates of the VLBI station. In our first investigation we concentrate on modeling the behavior of the ionosphere over the stations by two different approaches.

- Model proposed by Kondo [7]
  In this model $VTEC_{Kondo}(t)$ for station i and time t is calculated as suggested by Kondo [7] using a constant offset, sine and cosine functions, and a daily rate

$$VTEC_{Kondo,i}(t) = a_{i0} + \sum_{k=1}^{4}\left[a_{ik}\cos\left(\frac{kt\pi}{12}\right) + b_{ik}\sin\left(\frac{kt\pi}{12}\right)\right] + c_i t \qquad (2.6)$$

- Vienna TEC Model (VTM)
  In this model $VTEC_{VIENNA}(t)$ is calculated as a piece-wise linear function

$$VTEC_{VIENNA,i}(t) = offset_i + rate_{i1}(t_1 - t_0) + rate_{i2}(t_2 - t_1) + \ldots + rate_{in}(t_n - t) \qquad (2.7)$$

  and $t \le t_n$

  After calculating the partial derivatives

$$\left(\frac{\partial VTEC_{Kondo,i}}{\partial a_{ij}}, \frac{\partial VTEC_{Kondo,i}}{\partial b_{ij}}, \frac{\partial VTEC_{Kondo,i}}{\partial c_i} \text{ or } \frac{\partial VTEC_{VIENNA,i}}{\partial offset_i}, \frac{\partial VTEC_{VIENNA,i}}{\partial rate_{ij}}\right)$$

a least-squares adjustment allows separation of the VTEC values from the constant instrumental offset $\tau_{offset,i}$. This is possible because observations are performed at different elevation angles which prevents the design matrix from becoming singular. The VTM approach also includes constraints on the VTEC-rate of about +/– 30 TECU per hour for two reasons: to get physically reasonable values and to get a non-singular design matrix even if there are gaps in the data.

3. Results

3.1. Ionospheric values and maps derived by GPS

Data relevant to this work stored in IONEX format on the IGS web server are provided by the following analysis centers:

- Center for Orbit Determination in Europe (CODE), University of Berne, Switzerland,
- Geodetic Survey Division of Natural Resources Canada – formerly Energy, Mines and Resources Canada (EMR), Ontario, Canada,
- European Space Operations Centre (ESOC) at the European Space Agency (ESA), Darmstadt, Germany,
- Jet Propulsion Laboratory (JPL), Pasadena, U.S.A.,
- Group of Astronomy and Geomatics, Universidad Politecnica de Catalunya (gAGE/UPC), Barcelona, Spain.

Figure 4 shows VTEC values for Kokee Park (Hawaii) as calculated by the different analysis centers. Figure 5 shows the CODE solution (CODE, http://www.aiub.unibe.ch/ionosphere.html, [8]) for March 4th, 2003, 14:00 UTC plotted as a (two-dimensional) contour plot.

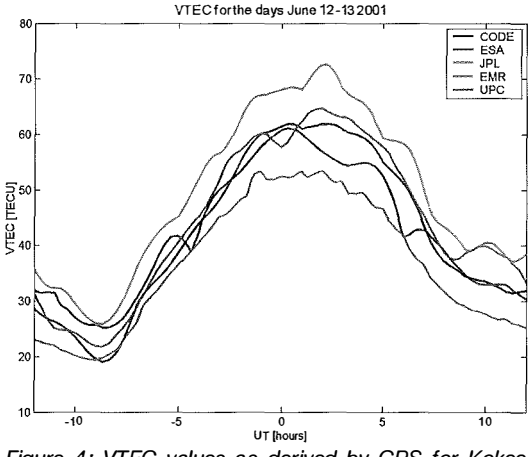


Figure 4: VTEC values as derived by GPS for Kokee Park (Hawaii).

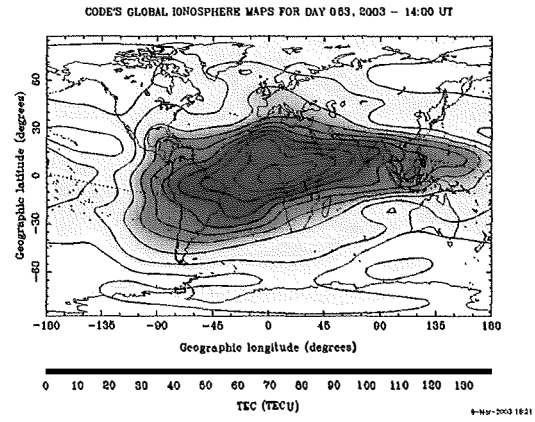


Figure 5: A global VTEC model derived by the CODE analysis center for March 4th, 2003, 14:00 UTC.

3.2. Values derived by VLBI

3.2.1. Results of the Kondo approach

Figure 6 shows VTEC values determined from the NEOS-A session on January 16th, 2001 (all contributing stations of this session are plotted in figure 7) over the station Fortaleza (solid line) compared to TEC numbers derived by CODE GPS solution (dashed line). In figure 8 the amplitudes and their precision of the harmonic terms, i.e. for periods of 24h, 12h, 8h, and 6h, for all stations involved in this experiment are plotted according to this model approach. Stations that are closer to the equator show larger amplitudes for daily periods. These results agree with the properties of the geographic regions as mentioned above. The correlation between amplitude and pole distance (90° minus absolute latitude)

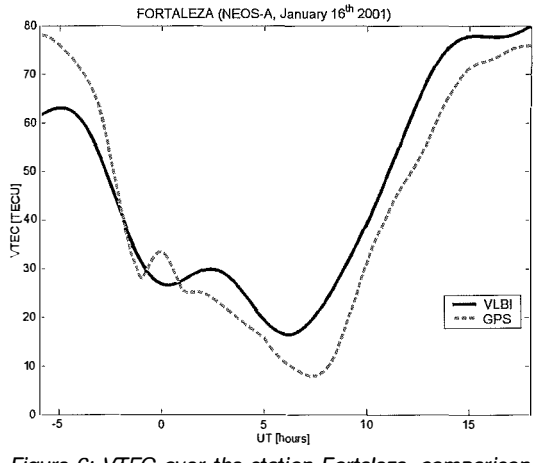


Figure 6: VTEC over the station Fortaleza, comparison between VLBI (Kondo approach) and GPS.

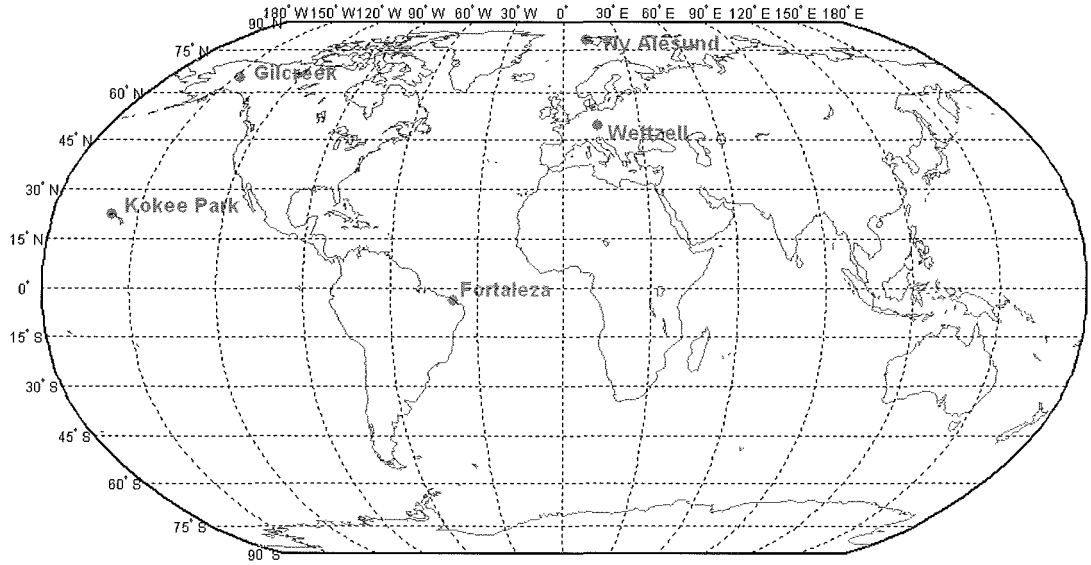


Figure 7: NEOS-A session on January 16th, 2001 – contributing stations

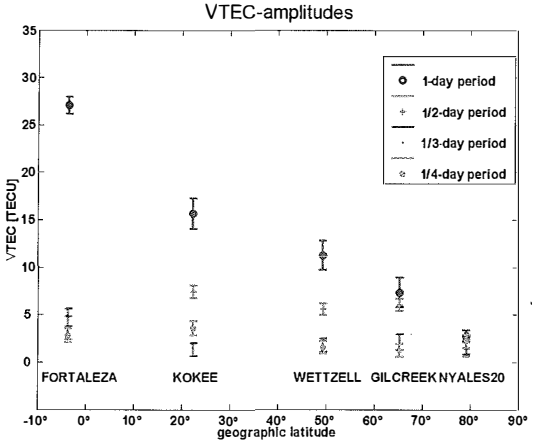


Figure 8: Amplitudes and their precision from the NEOS-A session on January 16th, 2001.

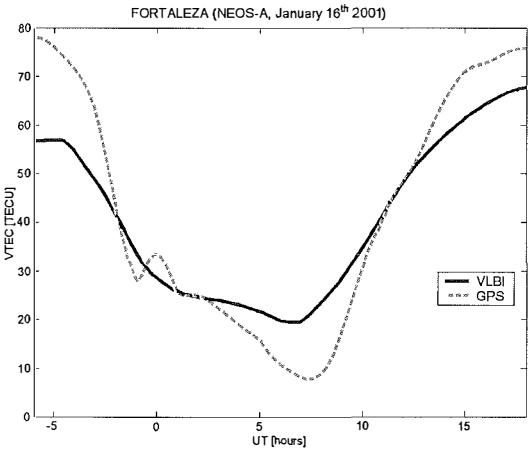


Figure 9: VTEC over station Fortaleza, comparison between VLBI (VTM approach) and GPS.

cannot be detected for sub-daily periods. From the error bars of the amplitudes we estimate the overall precision of the VLBI TEC values to +/– 5 to 7 TECU.

3.2.2. Results of the VTM approach

The same session as in 3.2.1 was analyzed now with the VTM. Two examples of the results are given here for the stations Fortaleza (figure 9) and Gilcreek (figure 10) compared to the CODE GPS computations.

For stations Fortaleza and Gilcreek generally good agreement can be seen between values derived by GPS and VLBI. The Fortaleza plot shows that the values derived by VLBI are much smaller during high ionospheric activity, i.e. around noon, local time.

3.2.3. Comparison of the VTM approach with all IGS analysis center solutions

The solutions of all analysis centers are plotted in figure 11 as dashed lines and compared with the VTM result (thick line). If for all stations in the NEOS-A network (the same as in 3.2.1 and 3.2.2) the differences between the VLBI and GPS solutions are calculated every 0.1 hour and all values are plotted in one histogram figure 12, is obtained. This should give an idea about the accuracy of the VTM approach.

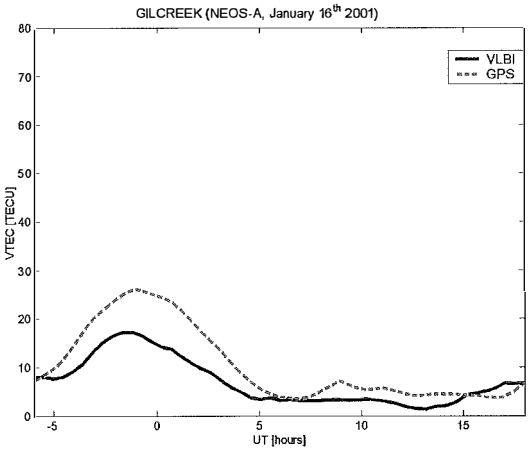


Figure 10: VTEC over station Gilcreek, comparison between VLBI (VTM approach) and GPS.

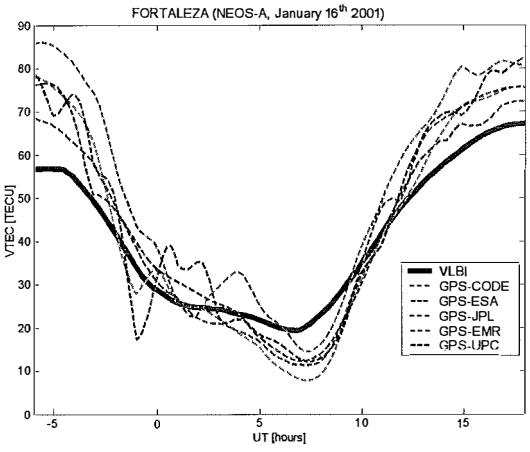


Figure 11: VTEC over station Fortaleza, comparison between VLBI (VTM approach) and GPS solutions of all analysis centers.

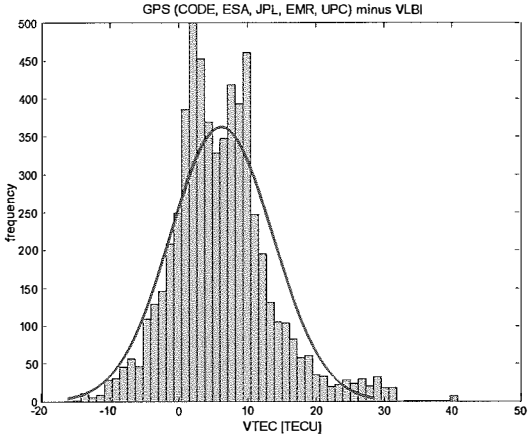


Figure 12: Histogram of the differences between VLBI and GPS solutions.

A normal distribution curve was fitted to the histogram which shows a mean difference ('bias') between VTEC values derived by VLBI and GPS of about 7 TECU.

4. Outlook

First investigations have shown that VLBI is able to deliver information about the spatial distribution and temporal variation of the ionosphere. Further research should deal with the problem that the calculated TEC values are not vertically located above the station. Another proposed goal of this project is the development of ionospheric maps like those published by IGS (IONEX). Since differences between VTEC values derived by GPS and TOPEX/Poseidon are usually in the range of 3 to 7 TECU, further comparisons with other techniques which deliver ionospheric information are necessary to evaluate which techniques are most accurate.

Acknowledgements

The research project P161366-N06 "Investigation of the ionosphere by geodetic VLBI" (VLBIonos) is funded by the Austrian Science Fund (FWF). We are also grateful to Dr. Arthur Niell from MIT Haystack Observatory for a thourough review of this paper.

References

[1] *Hargreaves, J.K.,* The solar-terrestrial environment, Cambridge atmospheric and space science series, Cambridge University Press, 1992

[2] *Hartmann, G.K., Leitinger, R.,* Range Errors Due to Ionospheric and Tropospheric Effectsfor Signals above 100 MHz, Bulletin Géodesique, Nr. 58, 109–136, 1984

[3] *Brunini, C.,* Global ionospheric models from GPS measurements, Tesis Doctoral, Facultadde Ciencias Astronomicas y Geofisicas, Universidad Nacional de La Plata, 1997

[4] *Schaer, S.,* Mapping and Predicting the Earth's Ionosphere Using the Global Positioning System, Inauguraldissertation der Philosophisch-naturwissenschaftlichen Fakultaet der Universitaet Bern, Bern, 1999

[5] *Schaer, S., Gurtner, W., Feltens, J., IONEX:* The Ionosphere Map EXchange Format Version 1, Proceedings of the IGS AC Workshop, Darmstadt, Germany, 9.–11. Februar, 1998, ftp://igscb.jpl.nasa.gov/igscb/data/format/ionex1.pdf

[6] *Feltens, J.,* Resumé of Three Years of Iono_WG Activities and Status of comparisons & Validations, IGS/IAACs Presentation, Collection of Presentations of the 2002 IGS/IAACs Ionosphere Workshop, ESOC, Darmstadt, Germany, January 17–18, 2002

[7] *Kondo T.,* Application of VLBI data to measurements of ionospheric total electron content, Journal of the Communications Research Laboratory, Vol. 38, No. 3, 613–622, Tokyo, Japan, 1991

[8] Center for Orbit Determination in Europe (CODE), University of Berne, Switzerland, http://www.aiub.unibe.ch/ionosphere.html

Contact

Dipl.-Ing. Thomas Hobiger, Dipl.-Ing. Johannes Boehm, Univ.-Prof. Dr.-Ing. Harald Schuh: Abteilung Höhere Geodäsie, Institut für Geodäsie und Geophysik, TU Wien, Gusshausstraße 27–29, A-1040 Wien. email: thobiger@luna.tuwien.ac.at, jboehm@luna.tuwien.ac.at, hschuh@luna.tuwien.ac.at

Results of Modelling GPS Satellite Clocks

Veronika Bröderbauer and Robert Weber, Wien


Abstract

The IGS (International GPS Service) Analysis Centers (ACs) provide GPS satellite clock offsets to GPS-Time (GPST) in the form of standard ephemeris in sp3-format or clock-RINEX files on a daily basis. These clock offsets, used mainly in GPS post-processing software along with consistent precise satellite ephemeris, are output to a least squares estimation process based on tracking data of the global IGS network.

Besides, to serve real-time applications, ACs have to forecast orbits and clock behaviour over a limited time span. The clock prediction models in use differ considerable both in terms of degree of the underlying polynomial as well as in the amount of observation data which enters a priori to fit the polynomial coefficients. First, we investigate the quality of the submitted clock offsets with respect to (w.r.t.) the observed combined IGS Rapid solution. Second, based on the satellites' clock-type specific behaviour, we try to set up a new model and to explore the stability and expected prediction errors of our approach.

Zusammenfassung

Für die Positionsbestimmung und Zeitübertragungsaufgaben mittels GPS benötigt der Nutzer Informationen über die Satellitenbahnen und -uhren. Die Analysis Centers (ACs) des IGS (International GPS Service) stellen die Bahnkoordinaten sowie die Abweichungen der GPS Satellitenuhren zu GPS-Zeit im sp3-Format zur Verfügung. Diese Files sind jeweils am folgenden Tag über einen freien ftp-Server erhältlich. Die Bahnen- und Uhren-Offsets sind das Ergebnis einer Parameterschätzung (vermittelnder Ausgleich nach der Methode der kleinsten Quadrate) auf Basis der Beobachtungsdaten des IGS-Stationsnetzes.

Für Echtzeit- oder beinahe Echtzeit-Anwendungen ist es notwendig, die Satellitenbahnen und -uhren für einen begrenzten Zeitraum vorauszuberechen. Die ACs verwenden für diese Prädiktion der Uhren verschiedene mathematische Modelle, die sich sowohl im Grad des Basispolynoms als auch im Umfang der verwendeten Eingangsdaten beträchtlich unterscheiden. In einem ersten Schritt beurteilen wir mittels einer groben Abschätzung die Qualität dieser Uhren-Offsets. Als Vergleichsdaten werden die Rapid Lösungen des IGS verwendet. Später wird versucht, ein eigenes verbessertes Modell für die Uhren-Prädiktion zu entwickeln. Nach der Bestimmung der Parameter eines Basispolynoms 2. Grades und gegebenenfalls einer additiven Sinusschwingung gelingt es schließlich, die GPS Satellitenuhren für 12 Stunden mit einer Genauigkeit von besser als ı2 ns vorherzusagen, was einem radialen Distanzfehler von ca. 50 cm entspricht.

1. Introduction

Precise point determination by means of GPS relies to a considerable extent on the quality of available satellite orbits and clock offsets with respect to GPST. GPST itself differs from the International Coordinated Time (UTC) by an amount of up to 40 ns (nanoseconds), ignoring leap-second differences. For many geodetic applications using differencing schemes the broadcast ephemeris, currently issued with an accuracy of about ± 2 m and a corresponding clock rms of ± 5 ns, are sufficient. To achieve highest precision, especially over baseline lengths larger than 10 km, the user is well-advised to take advantage of precise ephemeris provided by the IGS, [5]. Moreover, when analysing pseudorange and phase data, these products allow users to determine consistent coordinates and clock values even for an isolated GPS receiver with an internal accuracy of a few centimetres (precise point positioning), [4].

The IGS provides so-called 'IGS Final Orbits' of all GPS satellites on a weekly basis since the end of 1993. Compared to broadcast orbit information these ephemeris are more accurate by a factor of about 100, i.e. a few centimetres with a satellite clock rms of less than ± 0.1 ns. They are available for post processing applications with a delay of 13 days (counted from the last day of the week which is contained in the orbit-file). IGS combined clocks are based on a linear alignment to GPST separately for each day. So while the internal stability of ± 0.1 ns is quite good, the day-to-day stability of this reference

is poor. Besides, the IGS also provides so-called 'Rapid' solutions with a slightly lower quality which are available at 17.00 UTC the following day. IGS Final and Rapid solutions are available from the IGS website at [5] free of charge.

It is worth mentioning that a joint project of the IGS and the BIPM (Bureau International des Poids et Mesures) aims to develop and demonstrate the operational capabilities of satellite navigation systems for time transfer, [1], [2]. In this context concepts to steer the IGS time scale to UTC were investigated, [3]. This will allow for a general dissemination of accurate and easily accessible UTC in the near future.

Moreover, since November 2000, the IGS distributes 'Ultra-Rapid' products (IGU) comprising precise GPS satellite orbits and satellite clocks for real-time or near real-time applications. This solution, issued twice daily, contains both an observed and a predicted part. Both cover a period of 24 hours. While the orbits of the predicted part are output to an integration of the well-known force field the clocks have to be extrapolated by means of a sophisticated prediction model.

2. Ultra-Rapid clock solutions

Official IGS products are the result of a weighted averaging process, based on individual submissions of up to 8 IGS Analysis Centres. This statement applies for both, IGU orbits and clocks. In the first step we are interested in a rough estimate of the overall quality of the individual ACs clock submissions to the IGS Ultra-Rapid combination. Thus we start with a raw comparison of the observed and the predicted clock-offsets w.r.t. the combined IGS Rapid clock solution which serves as reference in all calculations. Later on we will extract some statistical information from these clock-differences. Our calculations are based on the clock information given in the sp3-product files with a time resolution of 15 minutes. This kind of comparisons are carried out at the Department of Advanced Geodesy (TU Vienna) since GPS week 1151, February 2002 and reported regularly at [7].

Raw clock differences usually reflect the clock offset and the clock drift of reference clock 2 w.r.t. reference clock 1 (see figure 1). In contrast to the combined IGS Rapid clock product (linearly aligned to GPS-time) the reference clocks used in AC solutions are steered to a very stable clock at one of the tracking sites or to a weighted assembly of hydrogen masers located in timing laboratories around the world. A clock-offset and the clock-drift are common to all reported satellite clocks. In addition clock-differences may reflect radial orbit differences (per satellite) of the corresponding ephemeris, which propagate into the clock solutions. For the observed 24 hours part these differences induced by the

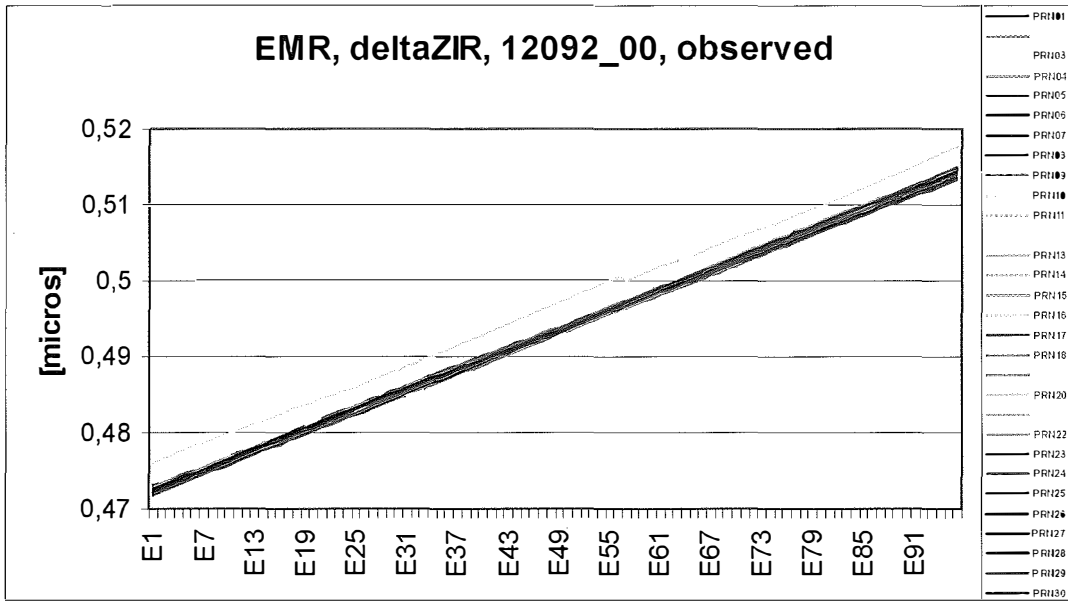


Figure 1: EMR satellite clock solution w.r.t. combined IGS Rapid clocks / GPS-week 1209, day 2. EMR ... Natural Resources Canada, Ottawa

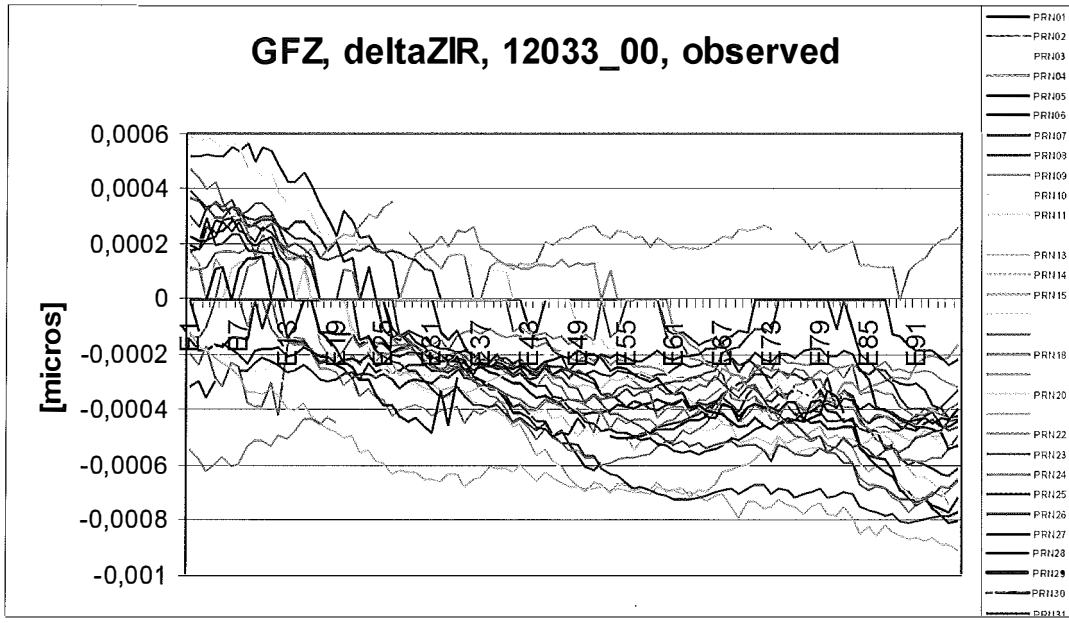


Figure 2: GFZ satellite clock solution w.r.t. combined IGS Rapid clocks / GPS-week 1203, day 3.
GFZ ... Geoforschungszentrum, Potsdam

orbits usually range up to a few tenth of a nanosecond (1 ns = 30 cm), see figure 2. In all diagrams of this paper label 'ZIR' stands for the IGS Rapid solution. Due to the 15-minutes binning, the time axis covers 96 epochs during a day (E1-E96).

Figure 3 demonstrates a common jump of about 1 ns around epoch E30. The graphic points to a phase-jump of the reference clock used in the ESA solution whereas a reset of an individual satellite clock would affect only that particular satellite. Phase jumps of the reference

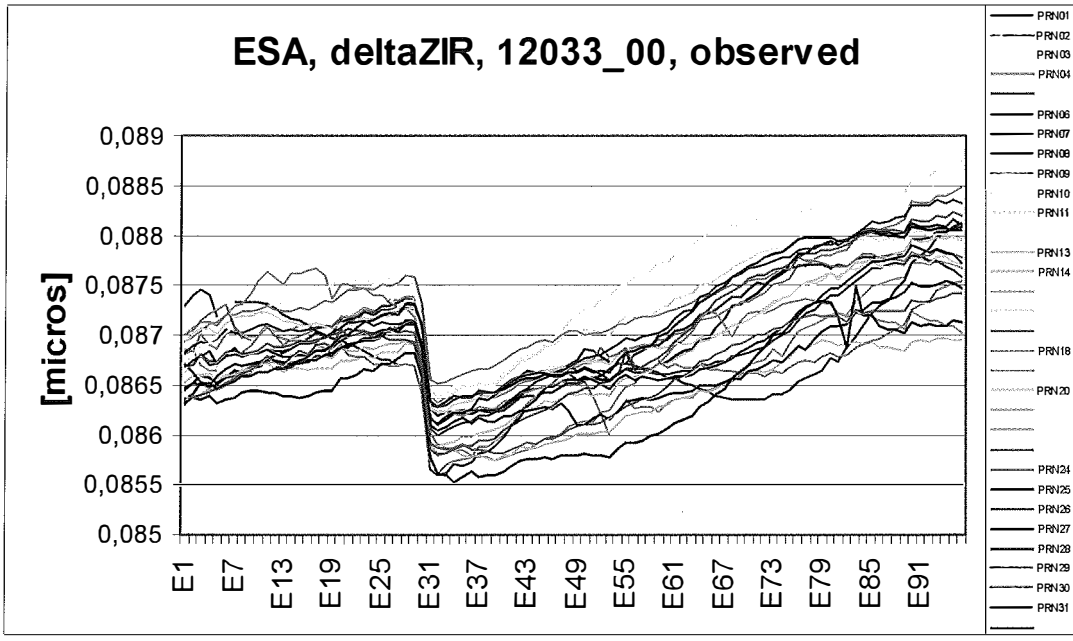


Figure 3: ESA satellite clock solution w.r.t. combined IGS Rapid clocks / GPS-week 1203, day 3.
ESA ... European Space Agency, Darmstadt

clock can be rather large, e.g., up to a couple of microseconds.

If the reference clock fails, the AC might switch to another reference clock of compatible quality. Currently about 30 stations within the IGS network are equipped with hydrogen masers and a few of them provide a stability of better than $2x10^{-15}$ for one day. It might be of interest to the reader that switching to another reference clock does not harm positioning as far as all satellite clock-offsets to GPS-time given for a single epoch refer to the same reference oscillator.

When inspecting the 24 hours period of clock prediction we find a complete different scenario. While the clock-differences of the observed part normally populate a small band of 1–2 ns, the values within the predicted part diverge substantially, see figure 4. Another outcome of the diagram is, that obviously some satellite clocks are more difficult to predict then others. Usually clock predictions over 12 hours are good to $\pm$3–5 ns, but depending on the stability of the satellite clock the prediction might be wrong by 10 ns over the same period. Figures 4a,b cover the same time slot in GPS-week 1204. Two IGS centres, namely USN (US Naval Observatory, Washington) and again ESA tried to predict the satellite clock behaviour during that period. Obviously the centres obtain their results from different prediction models. Ignoring a common drift and offset w.r.t. the reference clock (ZIR), which is insignificant for positioning, the USN predictions diverge after 24 hours by about 70 ns (satellites 10, 27) while the ESA predictions diverge by 60 ns.

In a second step we calculate the clock rms of the offset and drift reduced clock differences both for the observed and the predicted part. In detail, the ACs clocks are reduced for a satellite individual offset and drift w.r.t. to the combined IGS Rapid clocks. Thus, the remaining residuals Δ ZIR reflect solely periodical deviations which leads of course to very optimistic rms-estimates:

$$rms_i = \sqrt{(\sum(\Delta ZIR_i)^2)/n} \qquad (1)$$

with ΔZIR .. difference between the reduced ACs Ultra-Rapid clock solution and the IGS Rapid clock solution
i ..... satellite
n ..... number of epochs

As demonstrated in figure 5 the rms of observed satellite clocks typically range from 0.1 ns to about 0.4 ns. This result might be a little disappointing when compared to IGS Final or IGS Rapid clocks which are of a higher quality by a factor of 3. However, we should keep in mind that Ultra Rapid products are based on a relatively small quantity of immediately available tracking data.

For the predicted part, the clock rms is calculated in different intervals as shown in figures 6a,b. The intervals start at 0.00 GPST with the first predicted clock offset and last for 3, 6, 9, 12, and 24 hours, respectively. Again the clock differences have been reduced for an offset and a drift in advance. The satellite specific clock

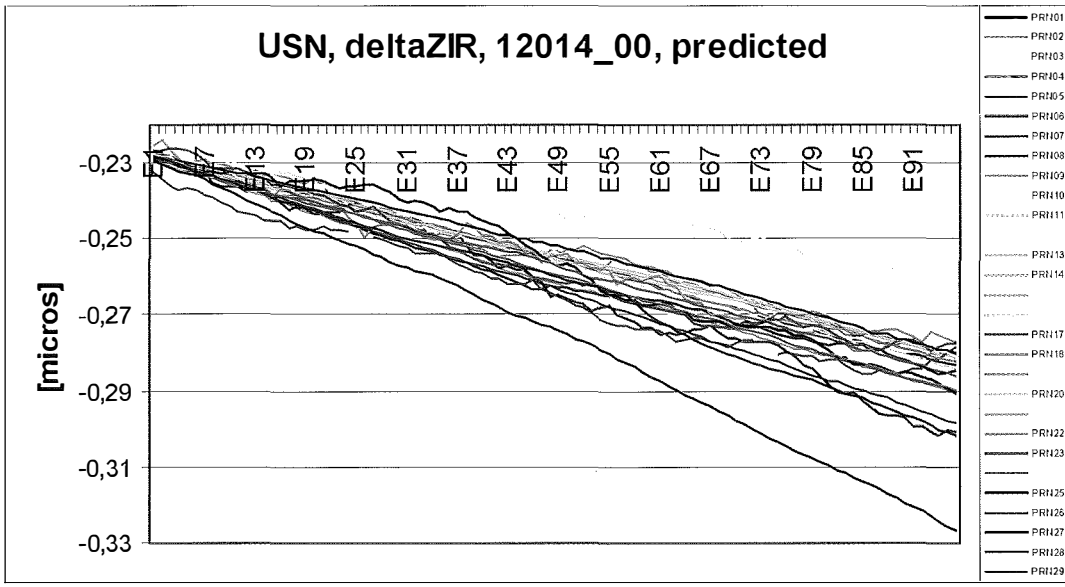


Figure 4a: USN satellite clock prediction w.r.t. combined IGS Rapid clocks / GPS-week 1201, day 4.

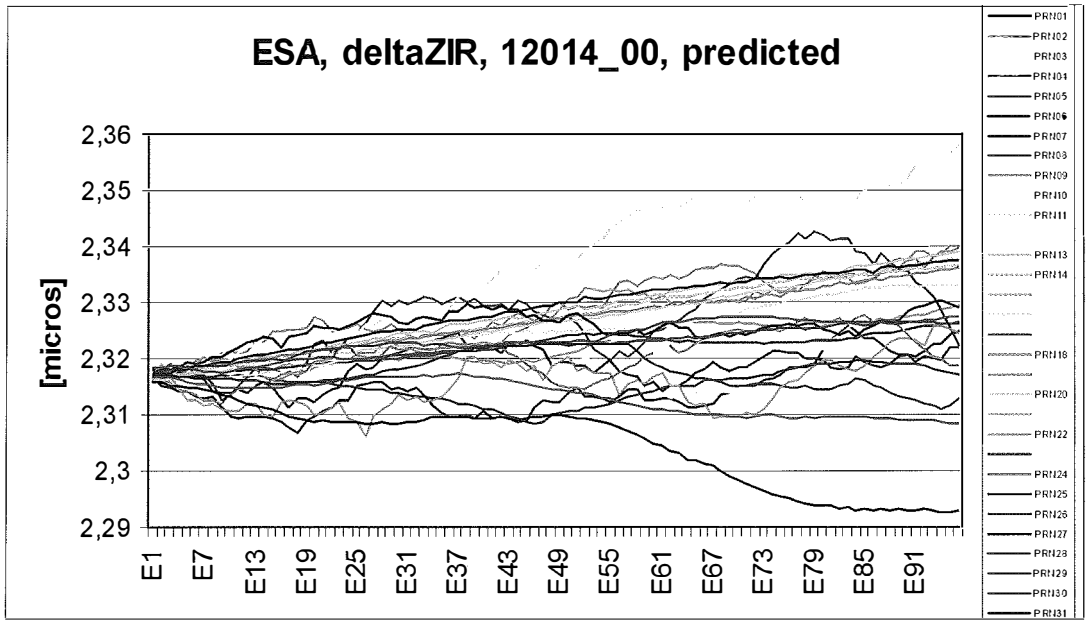


Figure 4b: ESA satellite clock prediction w.r.t. combined IGS Rapid clocks / GPS-week 1201, day 4.

rms for the predicted interval of 24 hours may reach 10 ns or more. As expected the rms-values increase in most cases with the length of the interval. A series of steady growing bars reflect a significant quadratic or periodic behaviour of the satellite clock (see figure 6a, PRN23). For comparison the AC-solutions presented in figures 6a,b coincide in time with figures 4a,b. Please note the different scale in figures 6a,b,c.

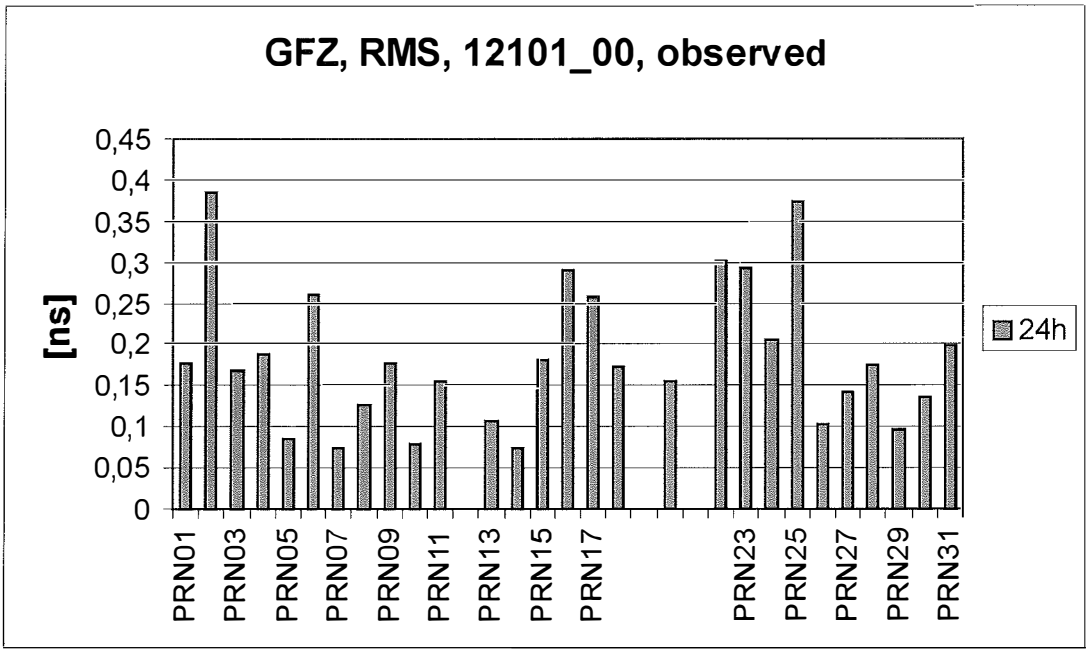


Figure 5: Satellite clock rms of GFZ observed Ultra-Rapid solution w.r.t. ZIR / GPS-week 1210, day 1.

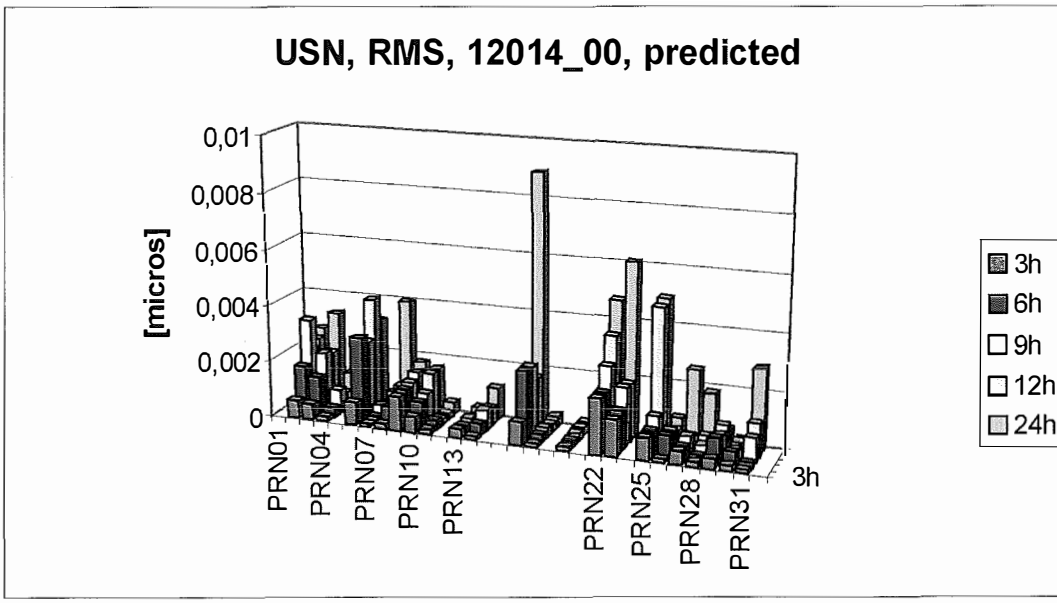


Figure 6a: Satellite clock rms of USN predicted Ultra-Rapid solution w.r.t. ZIR / GPS-week 1201, day 4.

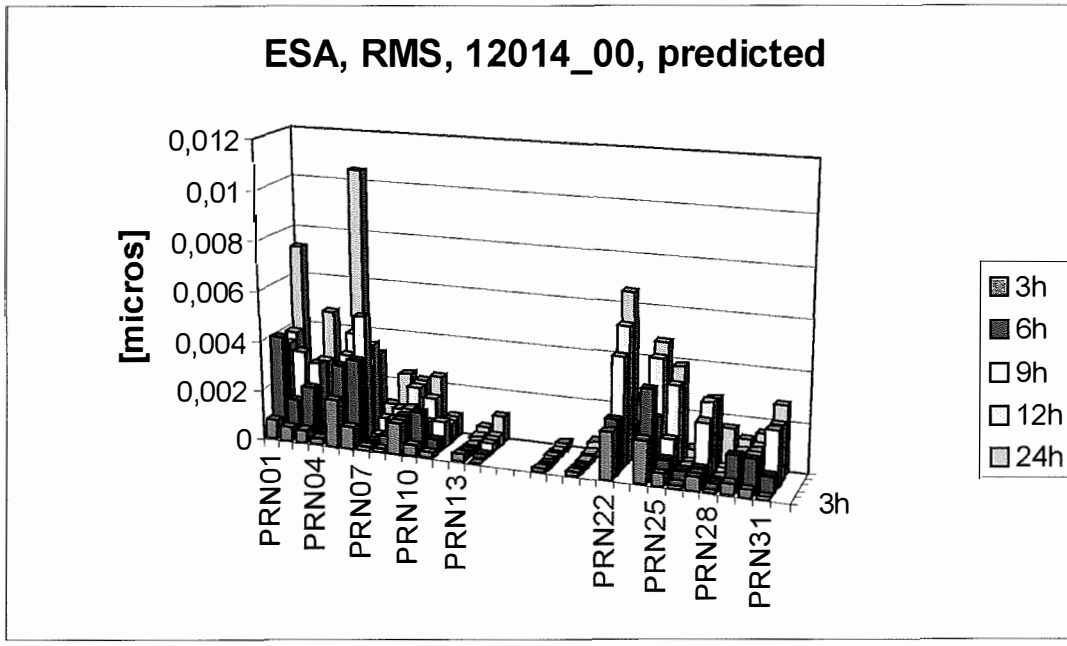


Figure 6b: Satellite clock rms of ESA predicted Ultra-Rapid solution w.r.t. ZIR / GPS-week 1201, day 4.

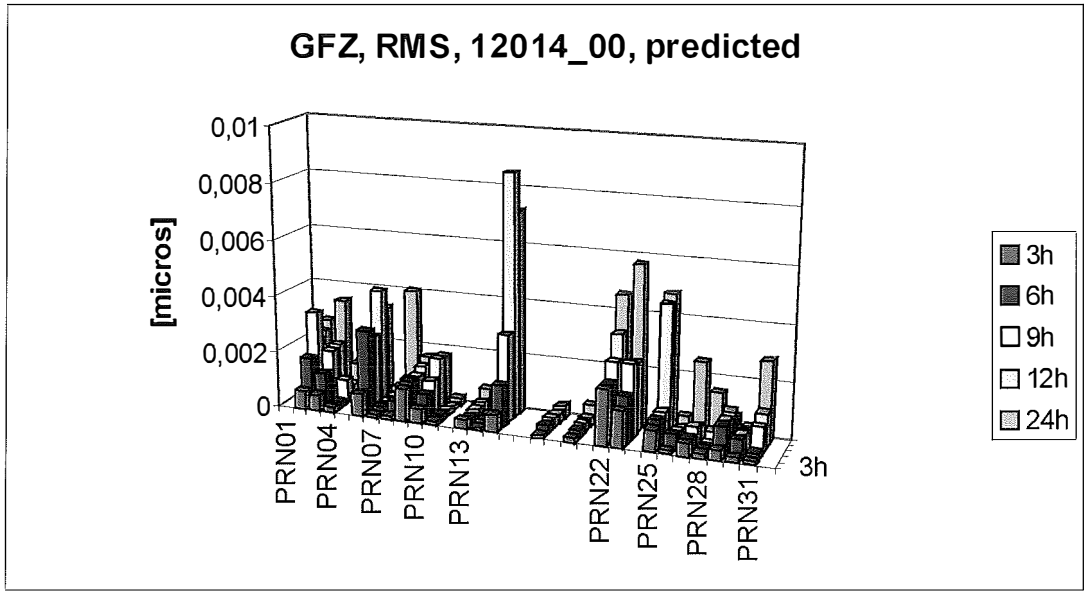


Figure 6c: Satellite clock rms of GFZ predicted Ultra-Rapid solution w.r.t. ZIR / GPS-week 1201, day 4.

3. Prediction of GPS satellite clocks

In order to explore reliable prediction models, valid over a span of 24 hours, we use the available observed part of the IGS Ultra-Rapid solutions of the past 48 hours. A least squares adjustment determines the coefficients of a polynomial of first or second order. Depending on clock type and behaviour (Cesium or Rubidium) we add cyclic terms.

To establish a continuous set of clock-offsets per satellite over the past 48 hours we have to deal with 'fictitious' clock jumps which show up in our sp3-files at the day boundaries. These jumps stem from the alignment of the IGS Rapid and Ultra-Rapid clock products to GPST. A very preliminary approach to bridge the gap in clock-offsets at the day boundaries is to shift all the data of the first 24 hours by a constant. Obviously a difference in drift still remains in the data. The resulting series serve as input for the determination of a polynomial to be used for clock prediction on the following day.

An interesting fact thereby is, that an expression valid for predicting a GPS satellite clock is obviously dependent on the clock-type of the individual satellite. The most recent GPS satellites (IIR) are equipped with Rubidium clocks, whereas others use Cesium clocks. The stability of a Cesium oscillator is about $1x10^{-14}$ over a day, whereas the stability for a Rubidium oscillator is limited to $1x10^{-12}$ over the same period, [9].

In table 1 the nature of the clock for each GPS satellite is shown, where "Rb" stands for Rubidium an "Cs" for Cesium:

PRN	CLOCK	PRN	CLOCK	PRN	CLOCK
01	Cs	11	Rb	23	Cs
02	Cs	13	Rb	24	Cs
03	Cs	14	Rb	25	Cs
04	Rb	15	Cs	26	Rb
05	Cs	16	Rb	27	Rb
06	Cs	17	Cs	28	Rb
07	Rb	18	Rb	29	Rb
08	Rb	20	Rb	30	Rb
09	Cs	21	Rb	31	Rb
10	Cs	22	Cs		

Table 1: Types of satellite clocks [8] as of April 2003

We analysed the drift reduced clock offsets to obtain reasonable apriori information about the behaviour of the individual GPS satellite clocks. From the remaining residuals we may deduce that the empirical formula describing the clock behaviour w.r.t. to a stable reference clock reads:

$$p(t) = a \cdot t^2 + b \cdot t + c + A_0 \cdot \sin(\omega \cdot t + \varphi) \quad (2)$$

with a, b, c ..... coefficients of the polynomial
A_0 ..... amplitude
ω ..... frequency
φ ..... phase shift

The argument t varies from 1–192 over the considered time span of 48 hours. We learn

from figure 7a that a typical Rubidium clock shows no significant cyclic terms and can be sufficiently described by a simple quadratic polynomial.

For satellites with Cesium clocks the situation looks slightly different. In case of Cesium clocks it is reasonable to set up a linear polynomial (instead of a quadratic one) with an additional sine

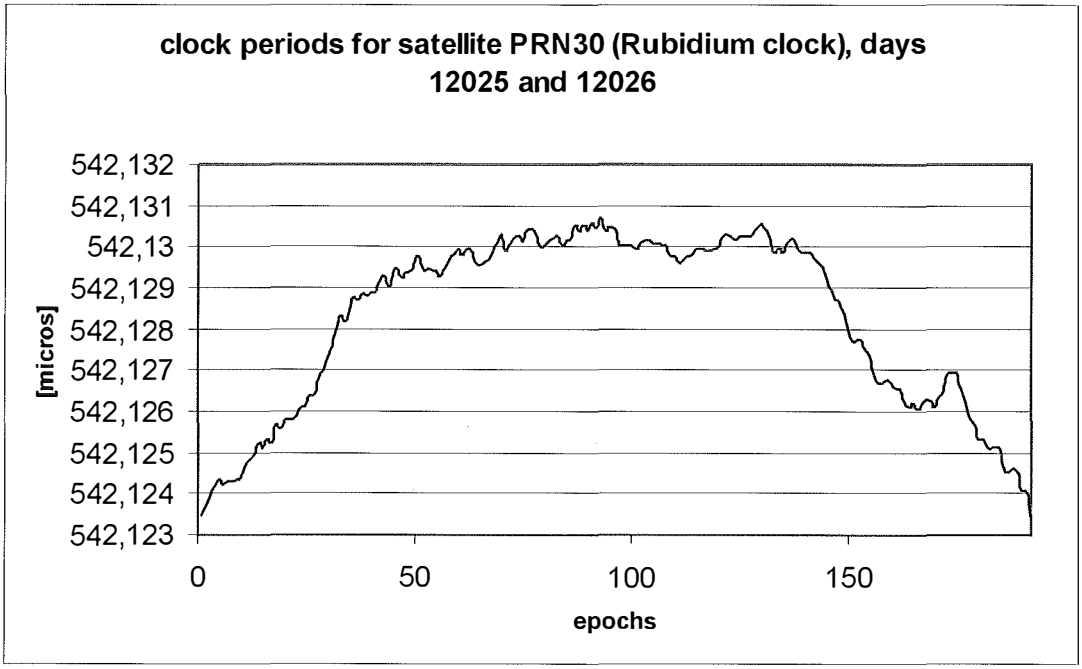


Figure 7a: Typical behaviour of a Rubidium clock over 48 hours / GPS-week 1202, days 5–6.

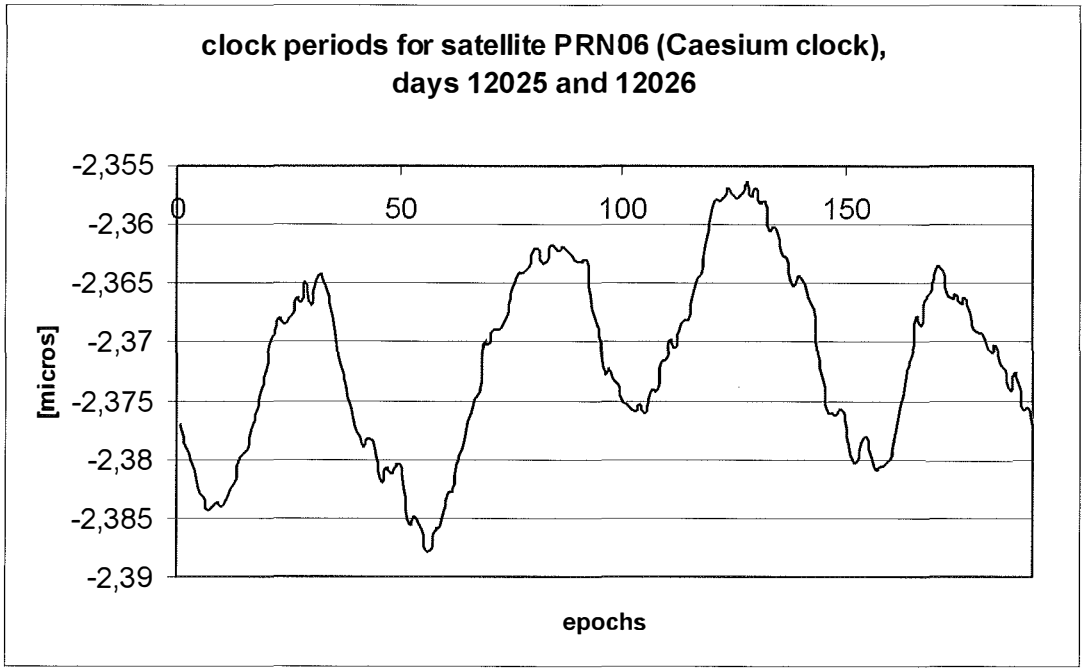


Figure 7b: Typical behaviour of a Cesium clock over 48 hours / GPS-week 1202, days 5–6.

term (see figure 7b). Within the determination of the sine term the amplitude and the phase shift to the input-data were fitted, whereas the main frequency was introduced as a predefined constant for each satellite. Most satellite clocks vary with a period of 12 hours which is close to the revolution period. These periods are not easy to de-correlate from orbit errors in the reference solution. We also found a couple of shorter, less significant, periods. A typical diagram of the variations of a Cesium satellite clock is shown in figure 7b. Both diagrams 7a,b cover 2 days.

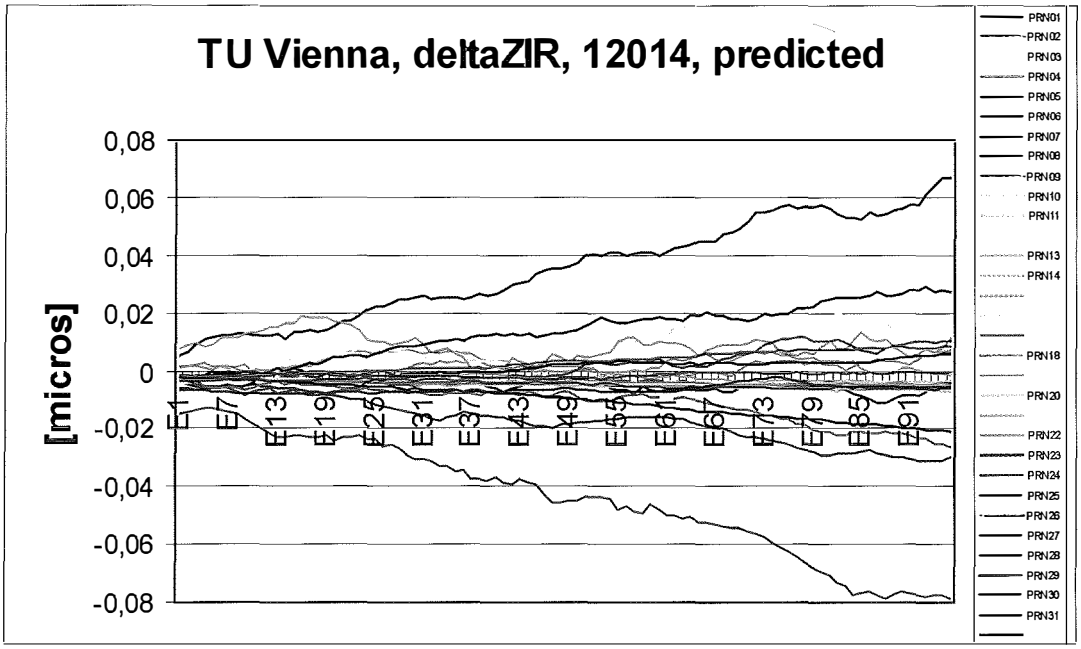


Figure 8: TU Vienna satellite clock prediction w.r.t. combined IGS Rapid clocks / GPS-week 1201, day 4.

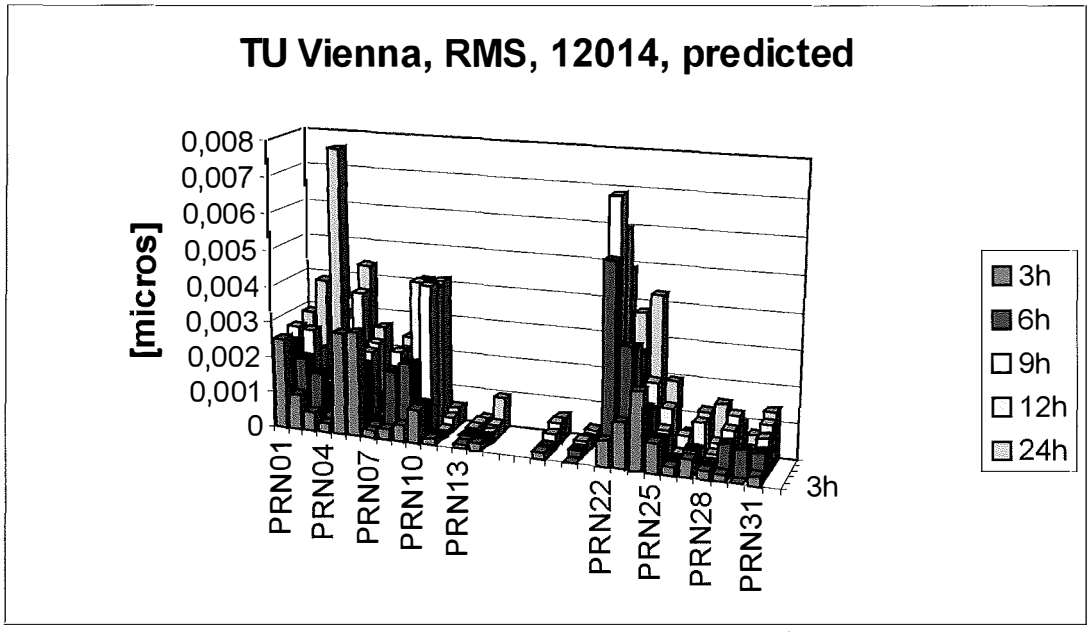


Figure 9: Satellite clock rms of TU Vienna predicted Ultra-Rapid solution w.r.t. ZIR / GPS-week 1201, day 4.

We may summarize that for satellites with no noticeable periods in the reduced clock-offsets solely the parameters of a quadratic function were estimated within a least squares adjustment. For those satellites which show a more complicated behaviour the parameters of a polynomial of first or second order and the amplitude and phase shift of an additional sine term were calculated. The coefficients for t^2 are in general rather small and in most cases significant. The amplitude of the cyclic term reaches up to a few nanoseconds. The phase shift is different for all satellites. The estimated functions are used to predict the satellite clocks on the following day. The results, labelled 'TU-Vienna', are compared with the IGS Rapid solutions for the same day (see figures 8 and 9). Satellite PRN15 is missing in our diagrams because no Ultra-Rapid solution was available for this satellite on days 2 and 3 of GPS-week 1201.

The results of our preliminary analysis are quite satisfying. However, figure 8 clearly indicates that further work has to be invested how to treat 'complicated' satellites like e.g. PRN01 and PRN05. Improvements are expected from dealing correctly with the different clock drifts in each 24 hours span of our input data. Also increasing the observation span for fitting reasonable forecast functions might be worth to investigate. Nevertheless the goal of providing clock predictions better than ± 1 ns for the subsequent 3–6 hours for real time point positioning will be extremely hard to achieve.

References:

[1] *Ray J.*, IGS/BIMP Time Transfer Pilot Project, IGS 1999 Technical Reports, November 2000

[2] *Ray J., Senior K.*, IGS/BIMP Pilot Project: GPS Carrier Phase for Time/Frequency Transfer and Time Scale Formation; Metrologia, 4th International Symposium on Time Scale Algorithms – 15 March 2002

[3] *Senior K., Koppang P., Matsakis D., Ray J.*, Developing an IGS Time Scale

[4] *Zumberge. J. F., M.B. Helfin, D.C. Jefferson, M.M Watkins and F.H. Webb, 1997:* Precise ponit positioning for the efficient and robust analysis of GPS data from large networks, Journal of Geophysical Research (JGR), Vol. 102, No. B3, pp. 5005–5017

[5] IGS, http://igscb.jpl.nasa.gov/

[6] AIUB, http://www.aiub.unibe.ch/

[7] TU Vienna, http://luna.tuwien.ac.at/forschung/satellitenverfahren/igs.htm

[8] USNO, GPS Timing Data & Information, http://tycho.usno.navy.mil/gps_datafiles.html

[9] NIST, Time & Frequency Division, http://www.boulder.nist.gov/timefreq/general/glossary.htm

Contact

Dipl.-Ing. Veronika Broederbauer, Dr. Robert Weber: Department of Advanced Geodesy, TU Vienna, Gusshausstr. 27–29/E1281, A-1040 Wien. email: veronika@luna.tuwien.ac.at, rweber@luna.tuwien.ac.at

Estimation of the Tropospheric Delay from GNSS Data over the Austrian Territory

Elisabeth Fragner and Robert Weber, Wien


Abstract

GPS has become an important tool both in navigation and in precise point positioning. One of the nuicance parameters limiting the accuracy of point determination is the water vapor content of the troposphere. On the other hand meteorologists are interested in the wet component of the troposphere as a valuable tool for Numerical Weather Prediction. Therefore GPS offers a low cost monitoring of water vapor with high temporal resolution.

We make use of continuous measurements of the GPS/GLONASS reference station network in Austria, which currently consists of about 30 sites with distances ranging from 50 km to 120 km. We calculate the zenith wet delays for a period of 2 months (February and March 2002). Subsequently the results are compared to contributions of different processing centers of the COST-716 project 'Exploitation of Ground Based GPS for Climate and NWP' [2] and with zenith path delay estimates provided by the IGS. As meteorologists need the water vapor within less than two hours, special attention is paid to the availability, reliability and especially to the quality of the satellite orbits used for the network calculations. For this reason we try to use rapid ephemeris instead of the IGS final orbits, whereby we perform a quality control of the rapid orbits. If one or more satellites show poor quality or if one satellite is missing at all we include the broadcast ephemeris information instead.

Zusammenfassung

In den vergangenen Jahren wurde in Österreich durch verschiedene Betreiber ein dichtes GPS-Permanentnetz mit einem Punktabstand zwischen 50 km und 120 km aufgebaut. Bisher war die Nutzung der Messdaten der GPS-Referenzstationen auf die geodätische Positionierung und diverse Navigationsaufgaben beschränkt. Die Brechung der GPS-Signale in der Troposphäre und der Ionosphäre wird dabei in der Datenauswertung üblicherweise als Störgröße behandelt und durch geeignete Algorithmen eliminiert oder zumindest reduziert.

Der gesamte Feuchtegehalt der Atmosphäre ist in den untersten Troposphärenschichten (bis zu einer Höhe von 10 km) in Form von Wasserdampf gespeichert. Die Verteilung des Wasserdampfes ist wesentlich für das Wettergeschehen verantwortlich und somit auch für Wettervorhersagen von großer Bedeutung. Seit wenigen Jahren versucht man deshalb das GPS-Positionierungsverfahren zu invertieren und das hohe Genauigkeitspotential der Messgrößen zur Beobachtung der Atmosphäre heranzuziehen. Man nutzt die Kenntnis der Stationskoordinaten und der GPS-Bahndaten um die troposphärische Verzögerung zu berechnen. Weiters erlauben genaue Messungen von Druck und Temperatur an der Bodenstation den hydrostatischen Anteil vom Feuchtanteil zu trennen, woraus der IWV (Integrated Water Vapour) berechnet werden kann. Als Integral über den gesamten Messstrahl liefert das Verfahren keine vertikale Auflösung der IWV Werte, was als Nachteil gegenüber den üblichen Ballonsondenmessungen gelten mag. Demgegenüber stehen aber sowohl die hohe zeitliche (30 min bzw. 1 Stunde) als auch räumliche Auflösung (horizontal alle 50 km) der Schätzwerte.

Unser Ziel ist es, aus den kontinuierlichen Messungen des österreichischen GPS-Permanentnetzes, welches zur Zeit aus beinahe 30 Stationen besteht, möglichst rasch nach Datenaufnahme meteorologische Parameter für numerische Wettervorhersagen abzuleiten. Dies verlangt einen schnellen Datenfluss, präzise Satellitenbahnen in beinahe Echtzeit (Internationales GPS Service - IGS) und eine Automatisierung der GPS-Auswertung mit Hilfe der Berner Software.

In dieser Arbeit werden für zwei Monate (Februar und März 2002) ZPDs (Zenith Path Delays) berechnet und sowohl mit Abgaben des IGS, als auch mit Ergebnissen von Auswertezentren, die im Rahmen des COST-716 Projekts „Exploitation of Ground Based GPS for Climate and NWP'' entstanden [2], verglichen. Da Meteorologen die Ergebnisse innerhalb von zwei Stunden benötigen, befassen wir uns mit der Verfügbarkeit, der Zuverlässigkeit und speziell mit der Genauigkeit der präzisen Satellitenbahnen. Es werden die Ergebnisse basierend auf den präzisen IGS ‚Final Orbits' den mit Hilfe von prädizierten ‚Ultra-Rapid' Bahnen (IGU) gewonnenen Resultaten gegenübergestellt. In verschiedenen Testszenarien werden in den IGU Bahnen gänzlich fehlende Satelliten oder Satelliten mit schlechter Bahnqualität durch die Broadcast-Information ersetzt.

1. Introduction

For weather forecasts meteorologists are in need of accurate and timely observations of water vapor. This is primarily because significant changes in the vertical and horizontal distribution of water vapor can occur rapidly (hours to minutes) during active weather processes. The majority of information about the vertical distribution of water vapor in the atmosphere comes from radiosondes that make in-situ measurements twice daily at widely spaced locations. The vertical resolution is good, but the spatial and temporal coverage is rather sparse. Alternatively the amount of water vapor can be measured by a water vapor radiometer. This instrument provides usually very accurate data, but its measurements are unreliable during precipitation and the device is also fairly expensive. GPS networks are capable of providing a rather dense and nearly continuously measured water vapor data.

Consider the GPS observation equation for code and phase measurements:

$$\text{Code: } \rho_k^s + \Delta\rho_{k,\,ion}^s + \Delta\rho_{k,\,trop}^s + c\Delta t_k + c\Delta t^s = P_k^s + \nu_k^s \quad (1a)$$

$$\text{Phase: } \rho_k^s - \Delta\rho_{k,\,ion}^s + \Delta\rho_{k,\,trop}^s + c\Delta t_k + c\Delta t^s + \lambda \cdot N_k^s = L_k^s + \nu_k^s \quad (1b)$$

where

L_k^s	is the carrier phase observation in units of length between satellite s and receiver k
ρ_k^s	is the geometric distance between receiver k and satellite s
c	is the speed of light in vacuum
Δt_k, Δt^s	are receiver and satellite clock offsets
$\Delta\rho_{k,\,ion}^s$	is the ionospheric path delay
$\Delta\rho_{k,\,trop}^s$	is the tropospheric path delay
N_k^s	is the initial phase bias
λ	is the wavelength of the carrier
ν_k^s	is the error or residual

The time delay imposed by the variable tropospheric refraction due to varying amounts of water vapor constitutes a limiting error source for this technique. Under the assumption that the ambiguities are successfully resolved, and the remaining error sources are well modelled, the tropospheric slant delay can be estimated from equation (1b). This slant delay $\Delta\rho(z)$ is usually mapped into the zenith $\Delta\rho(0)$ by means of an appropriate mapping function m(z):

$$\Delta\rho(0) = m(z) * \Delta\rho(z)$$

The total tropospheric delay can be separated into a hydrostatic and a wet delay, whereby it is preferable to use different mapping functions for the dry and wet part of the tropospheric delay:

$$\Delta\rho(0) = m_d(z) * \Delta\rho_d(z) + m_w(z) * \Delta\rho_w(z) \quad (2)$$

where

$\Delta\rho(0)$	is the total zenith path delay
$\Delta\rho_d$	is the dry tropospheric delay
$\Delta\rho_w$	is the wet tropospheric delay
$m_d(z)$	is the mapping function of the dry component
$m_w(z)$	is the mapping function of the wet component

As meteorologists are interested in the Integrated Water Vapor (IWV) or the Precitable Water Vapor (PW), we use the surface pressure and the station height to remove the hydrostatic zenith delay $\Delta\rho_d(0)$. This term may reach values up to 2.1 m at low altitudes and can be calculated with an accuracy of a few millimeters. The resulting zenith wet delay $\Delta\rho_w(0)$, which is mostly small (few cm up to 40 cm) but very variable, can be converted to IWV by the following expression

$$IWV = k * \rho_{H_2O} * ZWD \quad (3)$$

where

k	is an empirically determined factor (k~0.16)
ρ_{H_2O}	is the density of liquid water

To calculate ZPDs over the Austrian region we make use of the network of permanent GPS sites in Austria. The Austrian GPS network was established primarily for real-time positioning over the past years and consists at present (February 2003) of about 30 stations. The geographical distribution of stations is shown in figure1.

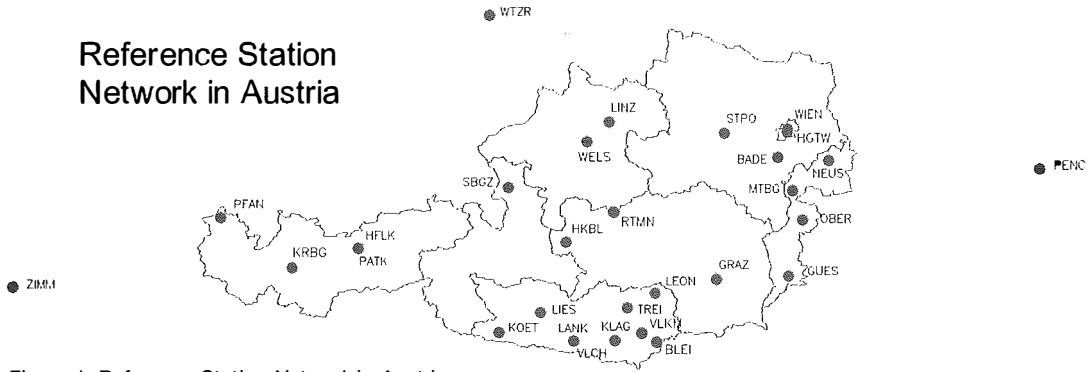


Figure 1: Reference Station Network in Austria

The receivers are operated by the Austrian Academy of Science, several local power supply companies (BEWAG, KELAG, WIENSTROM) in cooperation with the University of Technology Vienna and the Federal Office for Metrology and Surveying (BEV).

2. IGS Products

The International GPS Service, formed in 1992, has consistently improved and expanded the quality of its products. The IGS collects, archives, and distributes GPS observation data sets of sufficient accuracy to meet the objectives of a wide range of scientific and engineering applications and studies. The data sets are used to generate products like GPS satellite ephemeris, Earth rotation parameters, coordinates and velocities of the IGS tracking stations, GPS satellite and IGS tracking station clock correction information and last but not least tropospheric and ionospheric delay estimates.

Three types of GPS ephemeris are provided. While the final combinations are available within 13 days after observation, the Rapid orbit file is issued with approximately 17 hours latency. Both orbit products are of high quality as listed in table 1 but latencies are clearly inadequate for weather prediction applications, where data are assimilated in cycles of several hours duration. In order to be useful for weather predictions the precise orbits must be available in near real-time with an accuracy of at least ± 30 cm. The Ultra-Rapid combinations are generated twice each day (at 0300 and 1500 UT) and contain 48 hours worth of orbits; the first 24 hours are based on observations and the second 24 hours are predicted orbits. Currently this product, which combines estimates from seven IGS analysis centers, can achieve an orbit accuracy of better than ± 25 cm with respect to the highest quality final orbits (IGS).

Product	Final (IGS)	Rapid (IGR)	UltraRapid (IGU)
Delay	13 days	17 hours	3 hours
Ephemerides	< 5 cm	5 cm	< 25 cm
Sat. Clocks	0.1 ns	0.2 ns	5 ns

Table 1: IGS product accuracies

Figure 2 shows delivery times and the proportion between observed and predicted orbit part contained in the IGR and the IGU files.

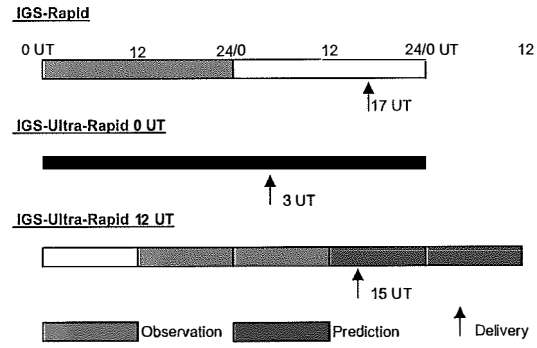


Figure 2: Latency of IGS products

3. Analyses

The main objective of this work was to investigate the impact of the quality of the satellite ephemerides on the estimated tropospheric delays. The near realtime product with less accurate satellite positions on the one side is opposed to the precise final orbits which were used as reference. We have calculated the total zenith delays (ZPD) of 27 permanent stations in Austria plus the neighboring stations PENC, WTZR and ZIMM over a period of 2 months (February and March 2002). The network was processed with the Bernese Software V 4.2 using the BerneseProcessingEngine (BPE), which allows a very automated data processing [5]. Data modeling was carried out under the following specifications:

- IGS final orbits/ IGS Ultra Rapid orbits
- 8 degrees cutoff elevation angle and height dependent observation weighting (cos z)
- no a priori troposphere model (full delay was estimated using Dry Niell Mapping Function)
- 2-hourly estimation of troposphere parameters
- ambiguity resolution: L5/L3

4. Results

The total ZPD at 3 stations (Graz, Hafelekar, Pfänder) is shown in figures 3 to 5 over a 10 days period (doy 032 - 041/2002) for both investigated orbit types, IGS Final (IGS) and IGS Ultra Rapid (IGU) orbits. For comparison, graphics 4 and 5 additionally display the ZPD estimates delivered by the IGS. These ZPD's are regularly generated on a weekly basis by IGS as a combined tropospheric product since 1997. They are based on the submissions of the individual IGS Analysis Centers. Furthermore our results were compared with those provided by four Pro-

cessing Centers delivering Near Real Time (NRT) data in association with the European COST-716 action (using UltraRapid orbits)[2], which are

GFZ (Potsdam, Germany)
GOP (Pency, Czech Republic)
ASI (Matera, Italy)
LPT (Wabern, Switzerland)

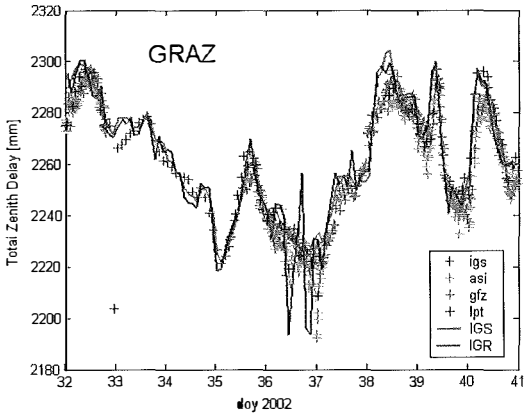


Figure 3: ZPD, Station GRAZ, doy 032 - 040

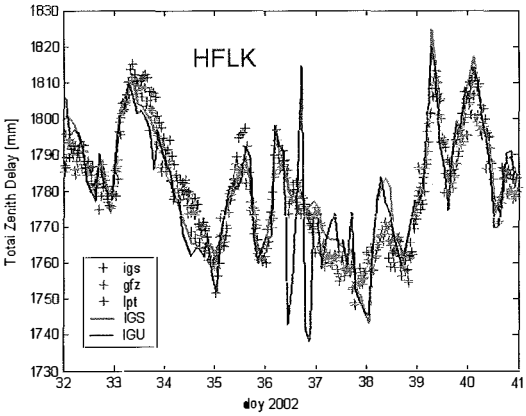


Figure 4: ZPD, Station HFLK, doy 032 - 040

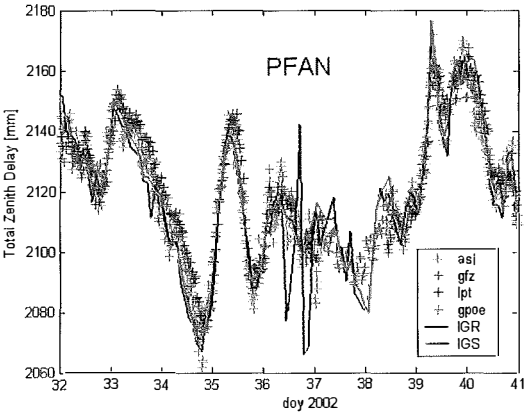


Figure 5: ZPD, Station PFAN, doy 032 - 040

As shown in the diagrams the two different orbit types result in small ZPD discrepancies with one remarkable exception during doy (day of year) 36. That day solutions show remarkable differences of up to 3.8 cm at all examined stations. The rms of the differences is ± 7.5 mm over the whole week but decreases to 14 mm when eliminating doy 36. A closer inspection of doy 36 points towards a poor orbit quality of GPS satellite PRN29 (flagged in the header by a high acuracy code number) which influences the results in a detrimental manner. The header of the standard GPS orbit format SP3 contains satellite specific accuracy information. The accuracy code has to be interpreted as the exponent of 2 (in millimetres), e.g. 5 implies an overall position accuracy of $\pm 2^5$ mm (~3.2 cm).

In a second step we solely focus on day 36. A number of test scenarios (SC1-SC4) with different modifications of the IGS Ultra-Rapid orbits were investigated.

SC1: We replaced all IGU satellite orbits with an accuracy code > 8 (PRN29, PRN 2) by the broadcast ephemeris information (Solution: IGU-B1).

SC2: We replaced all IGU satellite orbits with an accuracy code > 7 (PRN29,PRN 2, PRN 4, PRN24) by the broadcast ephemeris information (Solution: IGU-B2).

SC3: We replaced all IGU satellite orbits with an accuracy code > 7 AND included for all satellites missing in the IGUs the broadcast ephemeris information (Solution: IGU-B3).

SC4: We included for all satellites missing in the IGUs the broadcast ephemeris information (Solution: IGU-B4).

Considering these 4 scenarios the results for the same 3 stations and the computations made using the IGS and IGU orbits are shown in figures 6 to 8 and summarized in table 2.

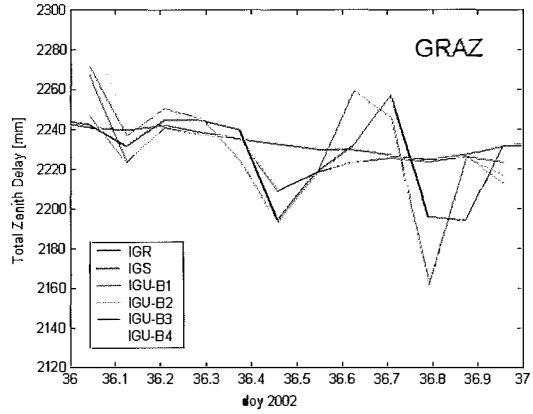


Figure 6: ZPD, Station GRAZ, doy 036

	all days but doy 36	doy 36	SC1	SC2	SC3	SC4
maximum deviation	1.1 cm	3.8 cm	3.0 cm	3.5 cm	6.4 cm	8.5 cm
rms	4 mm	20 mm	11 mm	13 mm	26 mm	34 mm

Table 2: Maximum deviation and rms error of ZPD differences between IGU test scenarios (initial solution excl. doy 36, doy 36, SC1-SC4) w.r.t. IGS solutions.

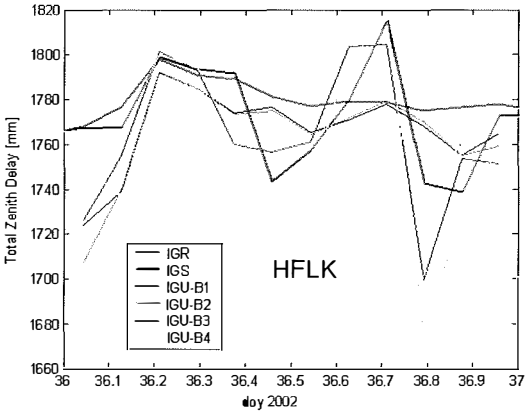


Figure 7: ZPD, Station HFLK, doy 036

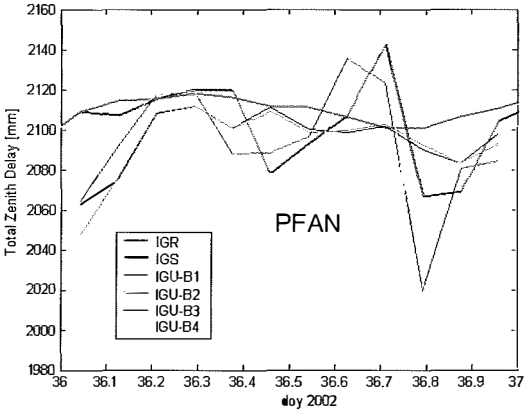


Figure 8: ZPD, Station PFAN, doy 036

We conclude that replacing satellites orbits of poor quality by the relevant broadcast information improved our estimated ZPD's. On the other hand, to add broadcast ephemeris information for missing satellites, has a negative effect. In most cases the included satellites degrade our results.

A closer look at the orbital representation of satellites PRN2,PRN4,PRN24 and PRN29 can clarify the situation. Under the assumption that the along track component is affected most by orbit mismodelling, we calculated the difference in along track of the IGU orbit (IGU) w.r.t. the IGS Final orbit. While PRN2,PRN4 and PRN24 show moderate deviations up to a few meters at the end of the prediction interval of 24 hours (see figure 9), the deviations of the PRN29 IGU ephemeris are quite huge (see figure 10). For comparison the accompanying broadcast information (BRD) of PRN2 and PRN29 is provided, too. Especially in case of satellites with accuracy codes > 8 the broadcast information can be a more relevant data source, especially in the second half of the considered prediction interval.

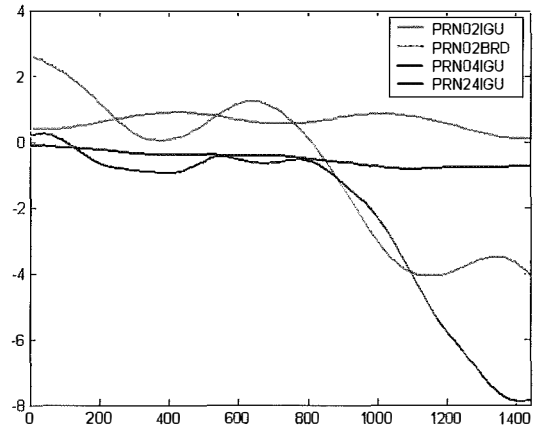


Figure 9: Orbital difference in along track of IGU orbit (IGU)/ Broadcast orbit(BRD) w.r.t. IGS Final orbit (PRN 2,4,24)

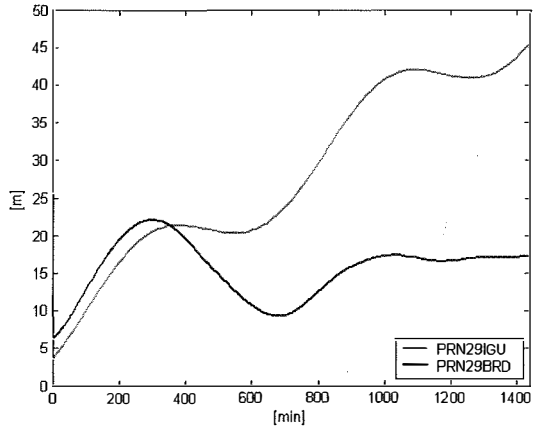


Figure10: Orbital difference in along track of IGU orbit (IGU)/ Broadcast orbit(BRD) w.r.t. IGS Final orbit (PRN 29)

5. Summary

As expected the time series show discrepancies between the zenith path delays estimated by using IGS and IGU orbits in the order of several mm up to 4 cm. These differences are mainly caused by the lower quality of the Ultra-Rapid orbits.

As a consequence we made an apriori quality check of the available IGS UltraRapid orbits and in case of bad orbit quality we replaced these satellites by those stemming from the broadcast ephemeris. In scenario 1 we clearly obtained better results. When including all missing satellites with their broadcast ephemeris information (SC3, SC4) no improvements in ZPD calculation can be reported. In some cases the added erroneous satellite orbit information propagated into huge deviations of the ZPD estimates.

Upcoming studies have to determine in more detail the accuracy code range (less than 8 or 9) which still provides satellite positions accurate enough for ZPD calculation at the 5mm level. Moreover they have to deal with questions like how reliable is such a simple one-digit characterisation of the extrapolated orbits and how useful are orbit extrapolations over periods longer than 12 hours? Hopefully useful in this context is the modified standard format for satellite orbits (SP3c), which should become effective very soon. The new format provides enhanced accuracy information, because it contains the calculated or predicted standard deviation of each position component (X, Y, Z) per epoch. Furthermore the update cycle of the IGU orbits will be reduced from 12 hours to 3 hours in the near future, which will support essentially the fast ZPD determination.

References

[1] *Beutler G., Schaer S., Rothacher M., 1999,* Wide Area Differential GPS, Astronomical Institute, University of Berne, 1999.

[2] COST-716 hompage: http://www.knmi.nl/samenw/cost716/index.html , Smith A., Exploitation of Ground Based GPS for Climate and NWP.

[3] *Fang P., Gendt G., Springer T., Mannucci T. (2001):* IGS Near Real-Time Products and Their Applications. GPS Solutions, Vol.4, No.4, pp 2–8.

[4] *Fragner E. (2003):* Estimation of the integrated water vapor from GNSS data over the Austrian territory. GPS Solutions, Vol.4, No.4, pp 2–8.

[5] *Hugentobler U., Schaer S., Fridez P. (Eds.), 2001,* Bernese GPS Software Version 4.2 documentation, Astronomical Institute, University of Berne, 2001.

[6] *Zhang J., Lachapelle G., 2001,* Precise estimation of residual tropospheric delay using a regional GPS network for real-time kinematic applications, Journal of Geodesy (2001) 75, pp 255–266.

Contact

Dipl.-Ing. Elisabth Fragner, Dr. Robert Weber: Institute of Geodesy and Geophysics, University of Technology, Gusshausstr. 27–29, A-1040 Vienna, Austria. email: efragner@luna.tuwien.ac.at, rweber@luna.tuwien.ac.at

Integration of GNSS and Loran-C

Johannes Vallant, Günther Abwerzger, Bernhard Hofmann-Wellenhof and Klaus Legat, Graz

Abstract

For many navigation applications, e.g. in urban or mountainous areas, insufficient satellite visibility of the Global Navigation Satellite System (GNSS) is an issue. This problem can be reduced by integrating GNSS with other dissimilar systems, where the drawbacks of the individual systems compensate each other.

An attractive option is the terrestrial radio navigation system Loran-C. Integrating GNSS and Loran-C improves the reliability and availability of the positioning information significantly. Within the GLORIA (GNSS and Loran-C in Road and Rail Applications) project, funded by the European Community (EC), the development and evaluation of a hybrid navigation receiver is demonstrated [1]. This innovative approach opens the door to new applications and to major improvements in existing application designs for land transport.

Zusammenfassung

Viele Landanwendungen der Navigation, vor allem im städtischen bebauten Bereich und im alpinen Raum, leiden unter der schlechten Satellitensichtbarkeit von Global Navigation Satellite Systems (GNSS). Diese Schwierigkeiten können durch eine Integration von GNSS mit anderen, verschiedenen Systemen kompensiert werden. Die Idee der Integration beruht darauf, dass ein Sensor die Nachteile der anderen Sensoren kompensiert. Neben der Verwendung von autonomen Navigationstechniken wie Koppelnavigation (dead reckoning) oder Inertialnavigation sind terrestrische Radionavigationssysteme wie z.B. Loran-C eine gute Ergänzung zu GNSS. Die Integration von GNSS und Loran-C erhöht die Zuverlässigkeit und Verfügbarkeit von Positionslösungen signifikant.

Dieser Bericht präsentiert die Entwicklung und Evaluierung einer neuen hybriden Empfängergeneration im Rahmen des von der EU finanzierten Projektes GLORIA (GNSS and Loran-C in Road and Rail Applications). Die Innovation beruht auf einer Integration von Rohdaten. Dieser neuartige Zugang verbessert bestehende Anwendungsmöglichkeiten und öffnet Türen für neue Betätigungsfelder auch außerhalb der Landnavigation.

1. Introduction

Envisaged are navigation solutions for land application focused on road and rail transport, but not exclusively limited to these domains. In the future, also pedestrian navigation will be favoured by reduced receiver size and weight.

In land navigation, specific situations with various environments are given: On the one hand, there are urban areas with high buildings and typical infrastructure of a city. On the other hand, no artificial infrastructure is present in rural areas, but we have to cope with limitations in signal reception due to difficult topography. Due to the required direct line-of-sight between satellites and receiver, today's satellite navigation systems most frequently used are not able to deliver position information every time in every environment.

The U.S. Global Positioning System (GPS), is the best-known GNSS system [2]. For navigation applications, primarily C/A-code measurements are used to derive the position information. Also carrier phase smoothing of code measurements is sometimes applied. For improving the accuracy and reliability of satellite-based positioning, Space Based Augmentation Systems (SBAS), like the American WAAS (Wide Area Augmentation System), the European EGNOS (European Geostationary Navigation Overlay Service), and the Japanese MSAS (Multi Satellite Augmentation System) can be used. Also terrestrial techniques (Eurofix in Europe) support high performance requirements in navigation by broadcasting augmentation information. Integrity information and enhanced accuracy of position solutions are the most important benefits resulting from these augmentation systems. However, it is not possible to overcome all insufficiencies of GPS. Beside the required direct line-of-sight to the satellites, the most critical problem is the accidental or deliberate jamming of satellite signals [3].

But GPS is not the only satellite-based navigation system, which is open for civil use: GLONASS (Global Navigation Satellite System) is the

Russian pendant to GPS. At present, only 7 of 24 necessary satellites are operational, but there exist plans of the Russian government to modernise the GLONASS system within the next years and to regain Full Operational Capability (FOC). However, GLONASS has the same problems and insufficiencies like GPS due the similar system properties.

Galileo is Europe's future satellite navigation system and its FOC is scheduled for 2008. This new civilian system belongs to the enhanced GNSS-2 level. [4]

There exist various approaches to overcome the shortcomings connected to satellite navigation. One promising candidate is to supplement satellite-based navigation by data of other sensors. The concept of this sensor fusion technique will be outlined in the subsequent sections.

2. Sensor fusion

Sensor fusion means a combination of different sensors to compensate drawbacks of one sensor by another. Some candidate sensors and systems are listed below.

2.1. Autonomous Navigation Techniques

Autonomous positioning techniques, i.e., techniques without support by an external system like terrestrial transmitters or satellites, are relative methods. This means that they determine positioning information relative with respect to a given reference station. To get the absolute position information, e.g. an initialisation of the autonomous navigation system has to be carried out. Due to propagation and accumulation of various measurement errors, autonomous techniques frequently suffer from accuracy degradation over time.

Electronic compasses for direction determination and odometers for relative distance measurements may be regarded as autonomous sensors. Differential odometers deliver distance and direction changes at once. Inertial navigation systems (INS) can also be used for measuring relative distance and direction variations. However, high quality INS are very expensive.

2.2. Radio navigation

Terrestrial Radio navigation uses on the one hand existing radio networks, like mobile phone or digital TV transmitter, and on the other hand dedicated navigation systems, like Loran-C. Position solutions derived from measurements in cellular networks and digital TV networks are characterised by a good availability, but currently cannot meet the performance requirements of land navigation. Loran-C, on the other hand, fulfils the minimal position accuracy in the test areas. Other features of the Loran-C signal are the high signal strength and a wavelength in the low frequency range. Due to these characteristics, a good signal penetration in all outdoor environments is given. In addition to the navigation signal, an augmentation signal (Eurofix) can simultaneously be transmitted via the Loran-C carrier. Main disadvantages of Loran-C are:

- Hardly predictable propagation effects of the carrier, if the signal travels over landmasses. This unknown signal propagation delay is referred to as ASF (Additional Secondary phase Factor) and can cause degradations of the absolute positioning accuracy down to some kilometres.
- Due to organisational uncertainties, the future of the system is unknown [6,7].
- Long signal integration times of measurements limit the use of current receivers for dynamic applications.

2.3. Mathematical methods and options of fusion

A very common mathematical method for combining different sensors and systems is to use a Kalman filter. The Kalman filter is a recursive and linear algorithm for the optimal estimation of various navigation parameters, which is based on a dynamic model of the vehicle motion. For more information about Kalman filtering in navigation applications see [6]. Also an epoch-per-epoch adjustment algorithm can be applied for the fusion of various sensors.

Generally, sensors can be loosely or tightly coupled. The integration of data can be done on the position data level (loose coupling) or on raw data level (tight coupling). The integration on the position data level offers the advantage that each system works independently and outputs an individually computed position. These individual positions are then integrated within the filter algorithm. When integrating the sensors on raw data level, all raw measurements are used together to deliver a common position solution. The advantage is that the integrated system provides useful navigation information even if one of the individual systems fails to compute an individual position.

When integrating different sensors and system, various aspects have to be discussed: Most sensors deliver measurements or position data based on different coordinate reference frames and time reference systems. GPS, e.g., uses the coordinate reference frame WGS-84 (World Geodetic System 1984). The position of the Northwest European Loran-C System (NELS) stations are also given in WGS-84 coordinates. If necessary, coordinates of different Cartesian reference frames can be harmonised by performing a seven-parameter coordinate transformation also called Helmert transformation. The time reference systems in use for the GPS/Loran-C integration are GPS system time and UTC (Universal Time Coordinated) as realized by the NELS control centre at Brest (France). The time synchronisation can be achieved by considering an additional unknown parameter to be solved within the estimation process. Note that there is a variable time offset between GPS time and UTC of 13 seconds (May, 2003).

Finally, it has to be mentioned that different systems deliver measurement or position data on different accuracy and availability levels. Another task in sensor fusion is the calibration of the integrated system which can be done in laboratory tests and setups.

3. Realization

The choice of sensors for an integrated navigation system depends on the characteristics of the various candidate systems. E.g., the terrestrial system Loran-C has a dissimilar signal characteristic compared to the space based GPS. Common vulnerabilities are almost not present. Another advantage for this combination is the good signal penetration of Loran-C in GPS hostile environments, like dense canopy in rural areas, or between high buildings in cities. Although Loran-C has low absolute accuracy, the relative (repeatable) accuracy of the system is very high and even comparable to GPS. The idea when combining the two systems is therefore to calibrate the absolute accuracy of Loran-C (i.e. to remove the influence of ASFs) during phases of good GPS availability and to use the calibrated Loran-C signal to continue the position computation during phases of limited GPS availability or also during complete GPS outages.

The combination is also suitable for additional sensors, although the mentioned combination can already fulfil the user requirements for many land transportation applications.

In the GLORIA project, GPS and Loran-C have been combined within one receiver type called DURAN (Dual Radio Navigation Receiver). This receiver integrates three major components, i.e. the LORADD prototype (innovative Loran-C receiver), a commercial GPS receiver, and a microprocessor for carrying out all computations.

3.1. DURAN Components

The LORADD is a fully digital, multi-chain, all-in-view Loran-C (and Chakya) receiver. It includes two 16-bit analog-to-digital (AD) converters, which operate at a sampling frequency of 400 kHz, and a high-end digital signal processor (DSP). The range-measurement loops track the incremental phase of all received and selected Loran-C signals. Since the phase of a signal may only be measured within one cycle, the receiver tracks each station from an unknown starting point that is "arbitrarily" fixed when turning on the receiver. The initial unknown number of full carrier cycles between the receiver and each transmitter station is obtained by tightly coupling the LORADD to the GPS component of the DURAN (this procedure is denoted as Loran-C calibration). Apart from its Loran-C navigation functionality, the LORADD also supports Eurofix, i.e., the GNSS augmentation service relying on Loran-C as a data link. Along with the receiver, Reelektronika has developed a prototype of an omni-directional magnetic field (H-field) antenna for the LORADD. According to earlier investigations, the H-field of the Loran-C signals provides a better penetration into urban areas than the electric field (E-field) [8].

The GPS part of the DURAN is realised by a commercial GPS receiver board developed by NovAtel (i.e., the GPS component is a common off-the-shelf (COTS) product) [9]. Although the OEM4 also provides carrier phase tracking on both GPS carrier frequencies (L1 and L2), only L1 C/A-code pseudorange measurements are processed within the DURAN.

The OEM4 was chosen since it is a high-quality instrument, which avoids introducing errors into the DURAN navigation solution that are typical of low-cost GPS receivers. The receiver is used together with its corresponding geodetic GPS antenna.

The heart of the DURAN is the integrated navigation software (IntNav) designed by TeleConsult-Austria (www.teleconsult-austria.at). In the

present version of the prototype, the IntNav software runs on a high-performance Pentium-type processor. For a future miniaturisation, this processor will be replaced by a more suitable DSP.

3.2. IntNav – The GPS and Loran-C integration software

As it becomes clear from the previous sections, the fusion of GPS and Loran-C should allow to compensate the main disadvantages of the individual systems. E.g., the low absolute positioning accuracy of Loran-C can be compensated by GPS, whereas Loran-C can bridge GPS outages caused by limited satellite visibility. Thus, performing a GPS-aided calibration of the Loran-C measurements seems to be the most convenient way to realise the integration. The calibration is carried out during phases of good GPS satellite visibility by a continuous computation of the theoretical Time of Arrival (TOA) of Loran-C signals and a comparison of that value with the corresponding measured TOA. The difference between computed and measured TOA yields a calibration value for the respective Loran-C station. During limited GPS availability, these calibration values are used to compensate the low absolute accuracy of Loran-C and, thus, to continue the position computation with near GPS-accuracy. The computation of Loran-C calibration values uses a simplified propagation model of Loran-C signals and bases on the current estimated receiver position. To ensure the reliability of the calibration, it is further essential to apply an automated decision-making procedure for deciding whether the GPS performance is qualified for calibrating Loran-C.

To realise these theoretical considerations, TeleConsult Austria has developed the GPS – Loran-C integration software IntNav. Although GPS and Loran-C snuggle together in this integration software, it has to be distinguished between individual pre-computational parts and a common integration part:

GPS pre-computations

Positions of GPS satellites: As GPS satellites move around the earth, the position of satellites has to be computed for each measurement epoch. The corresponding information can be derived from the broadcast ephemerides, which are transmitted by the GPS system itself.

Compensation of GPS signal propagation effects: The propagation of GPS signals is mainly affected by the ionosphere and the troposphere. Therefore, the integration software is able to apply models for compensating these propagation effects: The influence of the ionosphere is reduced by the broadcast 8-coefficients ionospheric model, whereas the influence of the troposphere is reduced by the modified Hopfield model. Applying these models to the measurements is optional.

Computation of measurement weights: To obtain individual weights for GPS measurements, mainly the elevation of the corresponding satellite is considered.

Loran-C pre-computations

Compensation of Loran-C propagation effects: The propagation of Loran-C signals over landmasses is mainly affected by the conductivity of the underground. The propagation delay, which is therefore introduced, is called Additional Secondary Factor (ASF). The pre-computation part of IntNav corrects Loran-C TOAs by applying calibration values, which have been derived during phases of good GPS availability.

Smoothing of Loran-C TOAs: The software has a built-in low-pass filter, which can optionally be applied to smooth Loran-C measurements.

Programming language	C/C++
Currently supported OS	Win 9x, Win 2k, Linux
Application environment	Console
Currently supported navigation systems	GPS, GLONASS, Loran-C
Measurement value	GNSS pseudoranges, Loran-C TOAs
Operation modus	Selectable: Real-time - Post-processing
Calibration of Loran-C	GPS-aided, real-time
Data pre-filtering	Elliptic low-pass filter
Quality check of solution	RAIM, DOP computation
Position computation algorithm	Selectable: Epoch-wise Adjustment - Kalman Filter
Overall structure	Modular, expandable
Currently supported devices	Ashtech GG24, NovAtel OEM4-3151R, Locus SatMate 1000 & 1020.

Table 1: Features of IntNav Software

Computation of measurements weights: The quality of Loran-C measurements is in a first approximation indirectly proportional to the distance between Loran-C transmitter and receiver. This is also the main input for individual Loran-C weight computations.

GPS and Loran-C integration part

The core of the IntNav software is an adjustment algorithm, which performs the integrated position computation. Besides the current position of the receiver, also some quality information, i.e., an estimation for the position accuracy, is obtained. The common integration part further consists of various consistency checks of the measurements, as well as a Receiver Autonomous Integrity Monitoring (RAIM) algorithm.

Concluding, Table 1 summarises the main features of the IntNav software.

The software is currently subject to further development and modification. Because of its modular structure, it can easily be adapted to various platforms and operating systems. Also the list of supported navigation devices can easily be expanded.

4. Receiver test and evaluation

The major aim of the tests was to compare the results of the DURAN with stand-alone GPS (reference case) and to investigate the potential benefits of the new receiver for road applications. Rail applications were also treated within the GLORIA project but will not be presented in this report. For all tests, the two chains of the Northwest European Loran-C System (NELS) with the master stations at Lessay (France) and Sylt (Germany) have been used. These Loran-C chains provide the best reception quality within the testing areas in the Netherlands, France, and Belgium for the major environment types rural, urban, highway. In these areas 14 static and kinematic tests were performed.

The test vehicles were not specifically modified for the tests. All measurement data were recorded in the internal memory of the receivers. The investigation of the DURAN performance was mainly done in post processing. The required offline version of the IntNav software used in this case operates completely analogous to the real-time version installed in the DURAN. The main advantage of post processing analysis is, that certain parameters of the algorithms can be modified to achieve better overall performance of the receiver. This optimization phase was carried out in parallel to the investigation of the receiver performance. Afterwards, the resulting final settings of the software were stored in the receiver for online position computing.

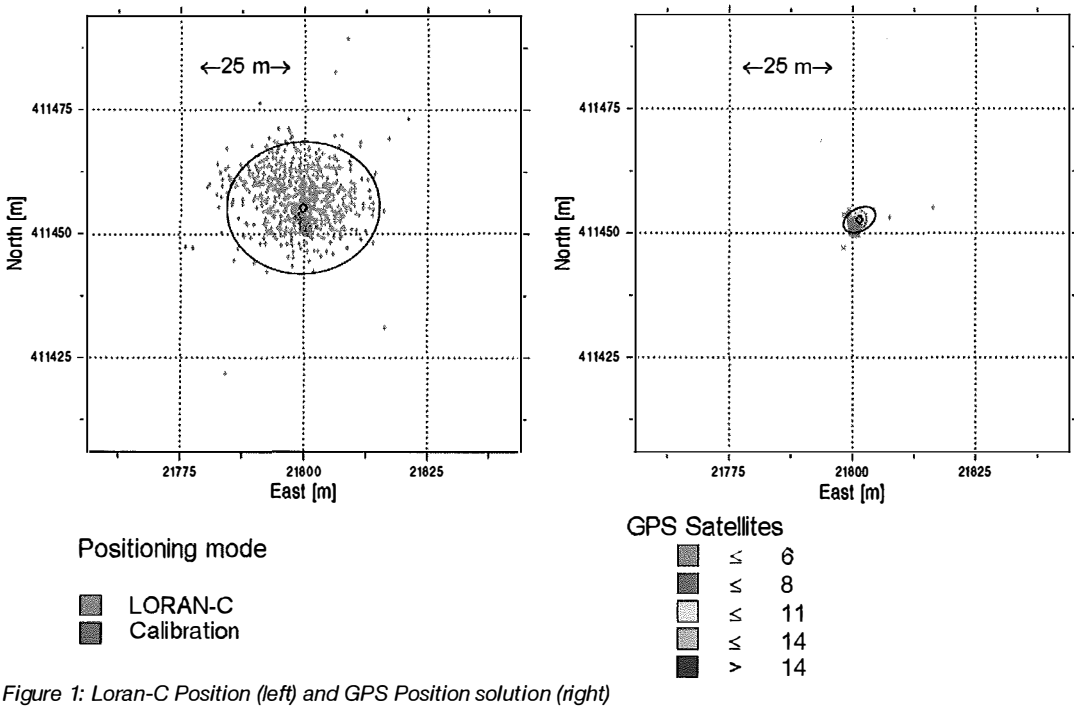


Figure 1: Loran-C Position (left) and GPS Position solution (right)

Method applied	Error ellipse (95%)		Bias [m]		Avail. w.r.t. time [%]	Max. out. [s]
	axes [m]	orient. [°]	north	east		
GPS stand-alone (reference case)	3.5 / 2.2	59	–	–	99.2	2
DURAN (fully integrated)	3.5 / 2.2	60	0.0	0.0	100	0
Calibrated LORAN-C	16 / 13	87	2.3	-1.3	93.2	15

Table 2: Numerical results of Static test Eiffel tower

TeleConsult employed some additional measurement equipment for determining alternative GPS reference trajectories. The additional equipment comprised two geodetic Ashtech GPS receivers (GG24, Z12) and the corresponding antennas. In case of GPS outages, no reference data are available. Still, the quality of the DURAN results may be derived from the tests: the resulting trajectories should continue smoothly during GPS outages. The data can be interpreted visually and the consistency of the DURAN results with the nominal trajectories can be verified using map information.

It was clear from the beginning, that even under optimised conditions the variance of the Loran-C position solution would be larger than for GNSS. The test near the Eiffel tower in Paris (F) was performed to investigate the static positioning quality of the Loran-C component of the DURAN and underlines our assumption [Figure 1, Table 2].

Kinematic tests are represented by a test on a motorway south east of Brussels (B). Figure 2 shows the details of the test track during a GPS outage. The DURAN is able to bridge the outage even though there are some short Loran-C outages as well [Figure 2, Table 3].

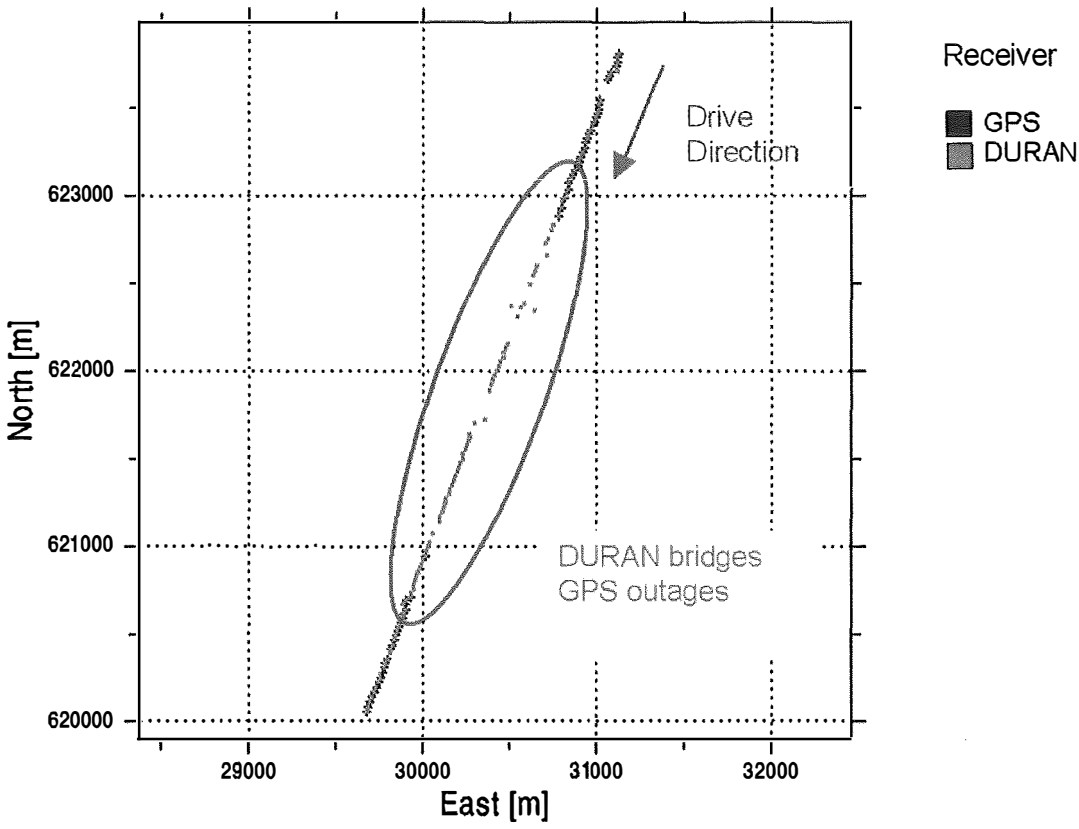


Figure 2: Kinematic test motorway

Receivers	Availability w.r.t. time [%]	Maximum outage duration [s]
GPS stand-alone (reference case)	91.0	76
DURAN (fully integrated solution)	99.7	37

Table 3: Numerical results of kinematic test motorway

5. Summary and outlook

The approach to integrate GNSS and Loran-C is feasible and is verified by good test results. The DURAN increases temporal and local availability of the positioning solution in case of partial GPS outages. Also the overall continuity of the integrated navigation system can be increased. Further, most of the requirements of navigation in road applications, like route guidance of private vehicles, tracing of vehicles for floating car data, and monitoring of dangerous goods on road can be met.

However, there are still some subjects – especially concerning Loran-C – to be investigated in more detail than they are known today. The major drawback of the current DURAN prototype is its limited resistance against electromagnetic disturbances. Furthermore, the improved reacquisition time of the receiver cannot be exploited at the moment because of the required recalibration of the Loran-C data after a Loran-C outage. If such an outage occurs simultaneously with a GPS outage, the new calibration of Loran-C can only be performed if GPS is "back at service". Further improvements are possible in the issue on when to use GPS to calibrate the Loran-C ranges. Especially for automotive applications, the miniaturisation of the receiver is crucial. DURAN has a large potential in this domain that needs to be exploited in the close future.

The currently limited accuracy of Loran-C in Europe is also caused by the sparse network of transmitter stations available. Further, it was emphasised that a significantly increased performance could be achieved if the network was extended only by a few additional transmitter stations.

References

[1] *Abwerzger G., Beyer J., Legat K., Maurer M., Meinhard D., Pfister J., van Willingen D. (2001):* GLORIA – Integrating GNSS and Loran-C for high requirement applications. Galileo's World, 2(3): 10–16.

[2] Department of Defense (2001): Global Positioning System standard positioning service performance standard. U.S. Assistant for GPS, Positioning, and Navigation, Defense Pentagon, Washington, DC 20301-6000.

[3] Department of Defense and Department of Transportation (2002a): 2001 Federal Radionavigation Plan. U.S. National Technical Information Service, Springfield, Virginia, DOT-VNTSC-RSPA-01-3/DOD-4650.5.

[4] Department of Defense and Department of Transportation (2002b): 2001 Federal Radionavigation Systems. U.S. National Technical Information Service, Springfield, Virginia, DOT-VNTSC-RSPA-01-3.1/DOD-4650.5.

[5] European Commission (2002): Galileo – The programme for global navigation services. ESA Publications Division, Noordwijk.

[6] *Hofmann-Wellenhof B, Legat K, Wieser M (2003):* Navigation – Principles of positioning and guidance. Springer, Wien New York. (In print.)

[7] NovAtel Inc. (2002): „NovAtel OEM4 Specification Sheet"; available under www.novatel.com

[8] *Pfister J., Kraft M., Legat K., Abwerzger G. (2002):* Assessment Report; GLORIA D10 WP5.

[9] *Reelektronika b.v. (2002):* „Receiver system prototype: DURAN"; GLORIA D8-B WP4.

[10] Volpe National Transportation Systems Center (2001): Vulnerability assessment of the transportation infrastructure relying on the Global Positioning System. Final Report, August 29.

Contact

Dipl.-Ing. Johannes Vallant, Dipl.-Ing. Günther Abwerzger: TeleConsult Austria Bernhard Hofmann-Wellenhof GmbH, Schwarzbauerweg 3, A-8043 Graz. email: jvallant@teleconsult-austria.at, abwerzge@teleconsult-austria.at

Univ. Prof. Dr. Bernhard Hofmann-Wellenhof, Univ.Ass. Dr. Klaus Legat: Technische Universität Graz, Abteilung für Positionierung und Navigation, Steyrergasse 30, A-8010 Graz. email: hofmann-wellenhof@tugraz.at, legat@tugraz.at

Means of Navigation for Automatic Level Crossing Control and the Concept of the ECORAIL Project

Elmar Wasle and Josef Ringert, Graz


Abstract

The European Train Control and Rail Traffic Management System (ETCS/ERTMS) shall strengthen the competitiveness and attractiveness of railway transportation. Additionally it intends to facilitate the inter-European passenger and freight traffic. The evolution from conventional systems to a standardised solution will be carried out by a stepwise approach. On the highest level, most of the track-side installations will be replaced by on-board equipment. One of the applications, which will benefit from this development, is the control of automatic level crossings. The key advantages of the new system for this application are cut-back costs for operators, increasing traffic flow, and reduction of emissions. The European Space Agency and the European Commission fund several projects which further develop the ETCS/ERTMS system. These projects investigate the feasibility and potential benefit of introducing the Global Navigation Satellite Systems (GNSS) into the railway domain. The project which deals with the combination of GNSS with automatic level crossing control is the „EGNOS Controlled Railway Equipment – ECORAIL".

Zusammenfassung

Die Einführung eines European Train Control and Rail Traffic Management System (ETCS/ERTMS) soll die Konkurrenzfähigkeit und Attraktivität des Schienenverkehrs verstärken. Das heißt, dass das neue System darauf abzielt, den innereuropäischen Personen und Güterverkehr zu erleichtern. Der Übergang von den derzeitigen nationalen Lösungen zu einem einheitlichen europäischen Ansatz erfolgt schrittweise, wobei verschiedene Ebenen der Implementierung unterschieden werden. In der vollen Ausbaustufe werden die streckenseitigen (track-side) Einrichtungen durch mobile Lösungen (on-board) in den Zugsgarnituren ersetzt. Eine jener Anwendungen, die davon betroffen sein wird, ist die Steuerung von automatischen Bahnübergängen. Die Vorteile, die sich aus der Implementierung des neuen Systems ergeben, liegen für diese Anwendung in der Kostenreduktion für den Eisenbahnbetreiber, in der Förderung des Verkehrsflusses und in der Reduktion von Fahrzeug-Schadstoffemissionen. Die Europäische Weltraumbehörde und die Europäische Kommission haben mehrere Projekte ins Leben gerufen, die das ETCS/ERTMS System noch einen Schritt weiter entwickeln, um die Unabhängigkeit von streckenseitigen Einrichtungen zu vergrößern. Dabei geht es um die Einführung des Global Navigation Satellite Systems (GNSS) in den Bereich des Eisenbahnwesens. Ein Projekt, das sich mit der Verbindung von GNSS mit der Steuerung von Eisenbahnübergängen beschäftigt, ist das „EGNOS Controlled Railway Equipment – ECORAIL".

1. Introduction – Fields of Transportation

The forth going competition between different means of transportation requires an increasing implementation of technological know how and strategies in management. The railway domain is a strong and attractive business partner presuming the railway operators will succeed in a cost-effective modernisation. The improvement of efficiency, the development of new applications and services will make the railway an interesting partner for inter- and multi-modal transport. Cooperation between the authorities, system operators, and system suppliers is needed to meet these goals.

Two main user groups can be identified in railway transportation: the passenger transport and freight transport. Regarding the necessary infrastructure, both groups require, apart from an acceptable cost-value-ratio, a high level of reliability, availability and safety – resulting in low door-to-door-times. These requirements have been met by introducing systems of train control and train management. Unfortunately, within Europe most of these systems are not interoperable. Several procedures of standardisation and certification are necessary to harmonise operational and technical specifications – national as well as cross-national. Safety relevant and interoperability issues play a major role in the process of standardisation and the consecutive certification.

To achieve this goal, the European Union initiated the specification and realization of a Eur-

opean Train Control System as part of the European Rail Traffic Management System (ETCS/ERTMS). Meanwhile these goals have been extended to account for new technologies and the new services provided. Before the new technologies and strategies are implemented into ETCS/ERTMS, their feasibility has to be proved and certified regarding their safety level. This article accounts for the means of navigation introduced into the railway domain and into the fields of ETCS/ERTMS. Therefore, a specific application is discussed and a project funded by the European Space Agency, which deals with the proof of feasibility, is presented.

2. ETCS/ERTMS

Train control is an important part of every railway operation management system. In the past, a number of different train control systems evolved in different countries at different times. Not all of these systems are compatible and interoperable with each other. Only a few of them are used in more than one country, and even in those cases differences in development arose. Therefore, the need of a standardised train control system is obvious.

The advantages of an internationally interoperable system are:

- cross border interoperability.
- improvement of safety.
- the possibility of an incremental introduction of new technologies.
- reducing bottle-necks by using high-tech systems.

The common literature about ETCS/ERTMS name several other advantages.

In 1990, the first initiatives started to establish a Train Control and Rail Traffic Management System. Taking into consideration the variety of train control systems and the varying needs of high-speed and/or conventional lines, a migration strategy from already existing systems to ETCS/ERTMS compatible systems was introduced. This migration strategy provides different ETCS/ERTMS application levels [1]:

Level 0 covers operation of ETCS equipped trains on lines not equipped with ETCS or national systems. Line side optical signals or other signalling techniques are used.

Level STM is used to run ETCS/ERTMS equipped trains on lines equipped with national train control and speed supervision systems. Trackside generated information is transmitted to the train via the communication channel of the underlying national system.

ETCS/ERTMS level 1 is a spot transmission based train control system to be used as an overlay for an underlying signalling system. Level 1 provides a continuous speed supervision system and is based on Eurobalises. An Eurobalise is a system for intermittent data transmission from the track to the train.

ETCS/ERTMS Level 2 is a radio based train control system which is used as an overlay of an underlying signalling system. Level 2 also provides a continuous speed supervision system and is based on Euroradio for track-to-train communication and on Eurobalises for train localization. Euroradio is a fail-safe, standardized transmission procedure via a GSM-R radio link. This system allows the transmission of fail-safe and non-secure data as well as speech.

Level 3 complements the previous levels in a way that no line-side signals are foreseen any more. Train localization and train integrity supervision are performed by a trackside radio block centre in co-operation with the train, which sends position information and integrity data to the centre.

The different levels require a demanding procedure of certification of the operational and technological solutions conducted to guarantee a high safety integrity level (SIL).

3. Automatic Level Crossing

One of the demanding applications which requires SIL 4 (cf. Section 8) is the control of automatic level crossings. Collision avoidance between the different means of transportation is done by an interaction of several technologies, which can be classified into a control, localisation, activation, monitoring, and signalling part. For a better understanding of the interrelations of these terms, the Automatic Level Crossings (ALX) are described in more detail.

A simple crossing is assumed as shown in Figure 1. Before the train is allowed to cross the road, the road traffic has to be stopped. Therefore the train activates the road signal. According to the latest Austrian legal regulations, the activation procedure begins with the yellow road signal which urges motorists and pedestrians to leave the endangered area. The red light phase follows the yellow light phase. At this point of time all road vehicles should have crossed the level crossing or stopped before it.

During the yellow and red light phase, the train approaches the level crossing with unaltered speed. To advise the train driver whether the level crossing was cleared, a monitoring signal is positioned at the monitoring distance. Having passed this signal, the driver either starts braking or proceeds with unaltered speed to pass the level crossing. The latest possible moment to activate the full braking and therefore the latest possible moment for the monitoring point is here denoted as point of no return. The monitoring signal must be clearly visible for the driver from a distance. That is the reason why the monitoring point and the point of no return do not necessarily coincide.

The point of no return is basically defined by the locally authorised train speed, the gradient of the track, and several other parameters. The distance between the activation point and the monitoring signal is denoted as striking-in distance. The distance between the monitoring signal and the level crossing refers to as monitoring distance.

After the train passed the level crossing the road signal has to be activated again to allow the road traffic to continue to pass the crossing – this is denoted as deactivation point. Furthermore, there is a control system which is the core element of an automatic level crossing. Refer to Figure 1 to get an impression of the situation and to identify the different parameters.

4. Track-side vs On-Board System

The conventional realisation of an automatic level crossing relies on so-called track-side technologies. Currently, different track-side based automatic level crossing control systems are used in different countries and at different lines. In a step of evolution the ALX control should be replaced by a radio based ALX control system. ETCS/ERTMS therefore introduces a combination of Eurobalises, train sensors (odometer,...) and GSM-R.

For the conventional track-side system wheel detectors, vehicle sensors, rail treadles, or switching contacts are used at the activation point to activate the closing of the level crossing. These instruments are connected by wire to the control system, normally placed right beside the level crossing (local). The control system activates the road signal and in return activates the monitoring signal in case that the road was closed. The monitoring signal is also connected to the control system by wires.

In case of the track-side system the striking-in distance is defined by the maximum speed (locally authorised speed) and activation delay. The activation delay is a function of maximum length of a road vehicle, duration of yellow light signalling, dimension of the level crossing, closing time of the barrier, etc. For the calculation of this distance the „worst-case-scenario" is assumed to guarantee maximum safety and security.

The system introduced by ETCS/ERTMS relies on Eurobalises for absolute position information and on an odometer for relative position information. Using Eurobalises, distributed along the line, and an odometer the actual position of the train is determined. In case that the train passes the virtual activation point the train transmits an

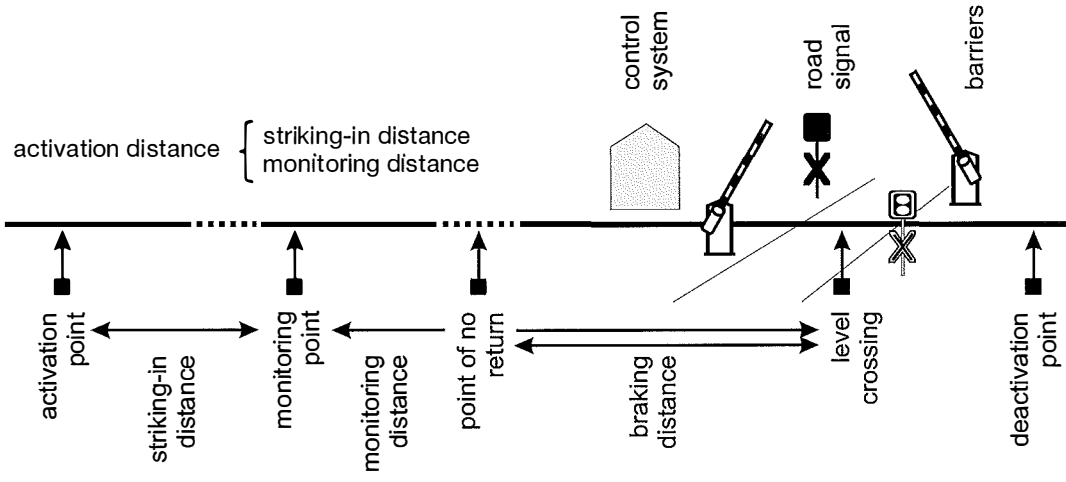


Figure 1: Schematic Figure of a Level Crossing

activation signal by GSM-R to the control system, which in contrast to the track-side system is placed either locally or remotely. The control system activates the road signal and in return sends a pass signal to the train. The certification of this system, where the main drivers are safety and interoperability, has been harmonised at the European level in ETCS/ERTMS pilot projects.

5. Advantage of an On-Board System

The main advantage of the on-board system is a speed-dependent activation of the level crossing. In case of the track-side system, the sum of striking-in distance and monitoring distance – further denoted as activation distance – is a function of a worst-case-scenario. In case of the on-board system, the activation distance as well as the monitoring distance can be adjusted to actual train parameters (speed, length, weight, etc.).

To give an example, a locally authorised speed of 100 km/h is assumed. The striking-in distance therefore amounts to about 1000 m, and another 1000 m have to be added for the braking distance. The overall distance of 2 km is covered by a train in about 72s. Taking into consideration that the position of the activating contact is fixed and some trains do not achieve the locally authorised speed, the time motorists have to wait behind red road signals and closed barriers can amount up to several minutes. Consequently, a train running at 50 km/h causes an increase in the waiting period for motorists of about 100 %.

Therefore an on-board / speed-dependent system guarantees an increase of traffic flow for motorist by means of optimised closing time, and a reduction of air polluting emissions due to shorter waiting periods. Reconsidering the example mentioned before and taking into account that 1 m of wiring would cost approximately € 70,– the presumed activation distance of 2 km amounts to € 140.000,–, for one activation direction. Therefore the railway operators would save considerable costs for the wiring by installing an on-board system.

Note that the replacement of the track-side system by an on-board system is only applicable in case that it provides a sufficient safety level.

6. GNSS On-Board System

ETCS Level 1 is a combination of a track-side and an on-board system. The system still relies on the Eurobalises. Although the location of a Eurobalise along the track is rather arbitrary, the number and distribution is a function of the accuracy of position information required. Since the odometer shows an accumulating error, the relative information of the odometer has to be updated by absolute information on a regular basis.

This leads to the idea to replace the track-side Eurobalises by virtual balises. Thereby new means of navigation are introduced into the field of railway transportation. In future systems, the Global Navigation Satellite Systems shall provide the absolute position information (cf. Figure 2).

The train determines its position autonomously. Approaching a level crossing the train transmits, dependent on its speed, an activation signal at the virtual activation point to the control

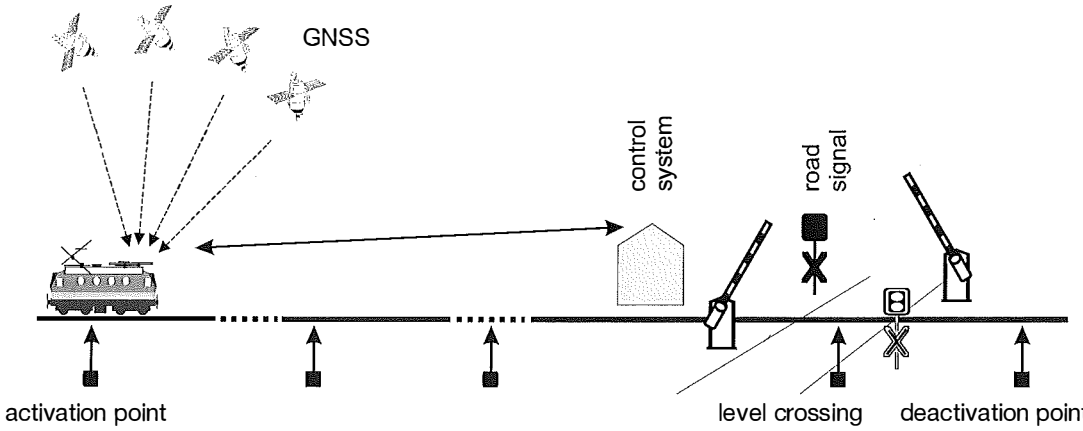


Figure 2: The train determines its position autonomously and transmits an activation request to the control system. The control system closes the level crossing and triggers the information to pass.

system via GSM-R. At the deactivation point, the train again sends a signal to the control system that the level crossing is cleared.

Assuming an ETCS Level 2 network the cost reduction by replacing the fixed Eurobalises by GNSS equipment could be considerable. Fixed balises for 2500 km of railway lines will cost less than M€ 10. The comparable technical equipment relying on GNSS, is estimated to cost only a fraction.

7. Database

The control of an automatic level crossing requires the on-board system to have information about the relative position of the train and the level crossing. In particular, the information about the distance between the ALX and the train is needed. The on-board unit uses this distance to identify, whether the point of activation has been reached. Since GNSS provides a three-dimensional position information, this position has to be matched to the railway line. Therefore, a database including the geometry of the railway line, all ALX and ALX-related points, also including all relevant information like authorised speed etc. have to be provided to the on-board unit.

The database stores the railway line in a hierarchic model, where the railway line itself represents the highest level, followed by a track segment within the line. The lowest level is represented by an edge, delimited by two vertices. Edges and vertices, as well as nodes, tracks, and lines are attributed with a number of information. In this context the database is denoted as digital route map or in a more common sense as Geo Information System (GIS).

By evaluating the GNSS signal, monitoring the speed of the train, and comparing this information with the stored track data, the activation of the level crossing is initiated automatically at the latest possible moment.

8. Integrity

As mentioned before, the on-board system has to guarantee the same security standards as the track-side system. Therefore, the on-board system has to pass a number of certification procedures. Extensive risk assessment and hazard analysis have to be applied to analyse the implemented architectures, methods, and techniques. This procedure shall provide a probability measure, that all functions are satisfactorily performed.

The risk analysis encompasses [2]

- the definition of the requirements of the railway system (independent of the technical realisation).
- the identification of hazards relevant to the system.
- the derivation of the tolerable hazard rates.
- the guarantee that the resulting risk is tolerable (with respect to the appropriate risk tolerability criteria).

The safety integrity of a function is defined by the probability of satisfactorily performing the required functions under all stated conditions within a stated period of time [3]. Five different levels (0 ... 4) of safety integrity have been defined, where level 4 is the highest. SIL 4 has to be met to guarantee an ETCS/ERTMS level 3 application.

9. Sensor Fusion

The GNSS as a sole means of navigation is not able to provide the high level of safety integrity. Although the US Global Positioning System (GPS) does not provide any information on integrity, the satellite based augmentation system (SBAS) EGNOS (European Geostationary Navigation Overlay Service) provides timely warnings and therefore integrity information within 6s. However, GNSS suffer too often from signal outages especially in topographic demanding areas due to shadowing.

A hybridised positioning system composed of GNSS equipment and train sensors would be able to overcome the data gaps in position determination and integrity information. However, the increasing complexity of the positioning system, key-word sensor fusion, follows a complex certification process caused by satellites geometry, hybridisation, and database.

10. ECORAIL

The European Space Agency (ESA) and the European Commission (EC) recognised the potential of introducing GNSS into the fields of ETCS/ERTMS. Both fund a number of projects, which deal with the implementation of satellite navigation into the railway domain to demonstrate the feasibility and benefits of GNSS in combination with ETCS/ERTMS.

One of these projects funded by ESA is the „EGNOS Controlled Railway Equipment – ECORAIL“ (cf. Figure 3). The project consortium led by Technicatome (Project Manager V. Thevenot)

combines experts of the fields of satellite navigation, interlocking and train control systems, geoinformation (ST Microelectronics, Systra, TeleConsult-Austria, Alcatel Austria) as well as the suppliers for and operators of railway networks (Technicatome, Alcatel Austria, Stern & Hafferl).

The project aims at the use of GNSS means of navigation to control automatic level crossing. ECORAIL therefore combines on-board and ground equipment:

- an on-board equipment which will be able to localise the train on the track. Beside the GNSS components GPS and EGNOS, also an odometer will be integrated into the system
- an on-board and a ground equipment which allow to exchange information (communication link)
- a ground equipment, which allows to record and compare the virtual actions coming from the on board equipment to the real command coming from the existing wired system

Figure 3: The ECORAIL project as seen by an artist's sketch combined with a real photo

Despite the ECORAIL system can not be allocated to a specific ETCS/ERTMS application level, it can be seen as a possible supplement. Relations to the different levels are obvious. The continuous positioning of the train fulfils level 3. Both, the ECORAIL system and ETCS/ERTMS – starting at level 2 – use a radio link for communication. Furthermore, the ECORAIL project and ETCS/ERTMS share common goals:

- improve the safety.
- save environmental resources.
- reduce costs by saving the expensive wiring.
- raise the ability to survive the commercial competition.
- integrate additional services based on satellite navigation.

11. Demonstration Phase

For the demonstration phase of the ECORAIL project, the on-board and the ground equipment will be installed on the Linzer Lokalbahn, a railway network of Stern & Hafferl in the federal province of Upper Austria. The demonstration has two main objectives. It should prove the feasibility of the satellite navigation positioning system for the chosen application. Apart from that, it should also be used to evaluate the performance of the navigation unit, both with respect to the particular application and to its general capabilities.

The primary aim is to prove that the ECORAIL equipment works in a full-size railway environment and that the specified functions are performed correctly (feasibility). It is also important to show that the system works reliably over a long period (reliability). Further it shall be shown that shorter closing times of the ALX barriers can be achieved by optimised activation (benefit potential).

The evaluation of the demonstration results will concentrate on the navigation system performance, on the robustness of ALX control and on possible reduction of ALX closing time. To evaluate and prove the performance of the ECORAIL system, three different modes of operation are considered:

- the simulated Operational Mode (OM) which sends the "close message" when the fixed activation point has been met.
- the Optimized Operational Mode (OOM) which sends the "close message" when the speed dependent virtual activation point has been met.
- the navigation test mode (NTM) which is similar to the OM but uses a constant confidence interval of 0.

Phase two of the ECORAIL project started with generating the digital route map (DRM). Therefore, the geometry of the railway line was acquired by a kinematic DGPS measurement campaign. In parallel to that, an EGNOS receiver was already installed on one of the trains to get a first impression about the EGNOS performance along the railway line. Figure 4 shows five different trajectories, offset by 1000 m in southward direction. Dark blue indicates the location where GPS and EGNOS are visible, the locations in red indicate where only GPS but no EGNOS is available.

EGNOS, as a reminder, provides the integrity information in the GNSS system. As shown in

Figure 4 there are some outages before and during approaching the activation distance. Further, the figure indicates that the EGNOS receiver performance varies with time. Therefore the integrity level has to be additionally guaranteed by the train sensors (odometer) in combination with the sensor fusion algorithm.

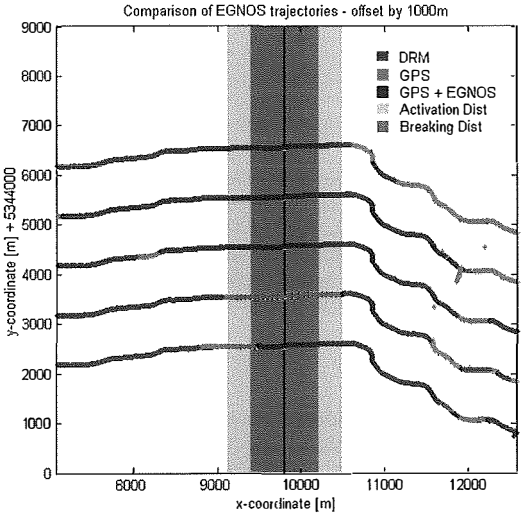


Figure 4: EGNOS availability during five survey runs at a railway network in Upper Austria (Stern & Hafferl) – the trajectories are offset by 1000m in southward direction

12. Outlook and Conclusions

The introduction of ETCS/ERTMS shall strengthen the competitiveness and attractiveness of the railway transportation. Additionally it intends to facilitate the inter-European passenger and freight traffic. ETCS/ERTMS train positioning is designed to use relative distances from known points along one-dimensional routes (no geo-referenced information).

The implementation of new means of navigation promises positioning functionality without Eurobalises. The introduction of GNSS as a global system increases the interoperability between national and regional systems. Within a demonstration phase, the feasibility of integration will be proved, and the performance of the system evaluated. A consecutive certification process is needed to guarantee the compatibility with the high safety requirements of the rail applications.

The high safety requirements however will require that GNSS operators provide a service guarantee, and as far as possible a risk analysis. Galileo as a future and enhanced European controlled GNSS will provide additional service and integrity information. This will further facilitate the implementation of the means of navigation into the railway domain.

Beside the safety critical application of train control, a position information system will also enhance non-safety critical applications like fleet management, service to the passenger, or optimisation of energy consumption.

References

[1] *ADTRANZ, ALCATEL, ALSTOM, ANSALDO SIGNAL, INVENSYS RAIL, SIEMENS (1999):* ERTSM/ETCS - Class 1 System Requirements Specification, Chapter 2, Basic System Description; Ref: SUBSET-026-2; Issue: 2.0.0; Date: 991222

[2] *CENELEC (1999):* Railway applications – Systematic Allocation of Safety Integrity Requirements (Ref. n°: prR009-004:1999 E) European Committee for Electrotechnical Standardization

[3] IEC 61508 – Functional Safety of Electrical / Electronic / Programmable Electronic Safety-Related Systems. International Electrotechnical Commission.

Contact

Dipl.-Ing. Elmar Wasle, Dipl.-Ing. Josef Ringert: TeleConsult-Austria, Schwarzbauerweg 43, A-8043 Graz. E-mail: ewasle@teleconsult-austria.at, jringert@teleconsult-austria.at

Development of a Vertical Comparator for System Calibration of Digital Levels

Helmut Woschitz and Fritz K. Brunner, Graz


Abstract

Today, digital levels are commonly used in precise levelling. Every level at the market has its specific error pattern, and knowledge about this is essential to obtain precise height readings. To identify and investigate the error pattern of digital levels, a vertical comparator was developed at the Graz University of Technology. System calibration is used to calibrate the level and the staff together. This paper reports about the design of and experiences with the vertical comparator. The standard uncertainty of this comparator is $\pm 3\mu m$ (computed in accordance with GUM, k = 2). The vertical comparator can be used for both, the quality control of digital levels and the routine system calibration which also yields the scale value of the system.

Zusammenfassung

Heutzutage werden zur Übertragung von Höhen hauptsächlich Digitalnivelliere verwendet. Bei Präzisionsanwendungen ist die Kenntnis über das Verhalten des verwendeten Nivelliersystems notwendig, um unverfälschte Höhenwerte zu erhalten. Um das Verhalten von Digitalnivellieren bestimmen und untersuchen zu können, wurde an der TU Graz ein Vertikalkomparator entwickelt. Die Methode der Systemkalibrierung wird angewendet, bei der im Kalibrierprozess das Nivellier und die Latte gemeinsam verwendet werden. In der Arbeit wird über die Entwicklung des Komparators und die Erfahrungen mit diesem berichtet. Die Messunsicherheit des Komparators beträgt $\pm 3\mu m$ (bestimmt nach GUM mit k = 2). Mit dieser hohen Genauigkeit eignet sich der Komparator für die Qualitätskontrolle von Digitalnivellieren, aber auch für die Routinekalibrierung, in der auch der Maßstab des Systems ableitbar ist.

1. Introduction

Currently, there are three different makes of digital levels available for precise levelling. They are manufactured by Leica, Topcon and Trimble (formerly Zeiss). All three makes have a resolution of 0.01mm and are commonly used with invar staffs of e.g., 3m length. The digital code and the associated technique to evaluate the pixel image are brand dependent. Algorithms used for the calculation of the staff reading are correlation, geometric averaging and Fourier analysis. A survey of the different measurement techniques was given by [1] and a detailed description by [2].

Extensive tests are carried out by the manufacturer before the release of a new digital level. However, every level at the market has its specific error pattern. So, independent tests are essential to establish appropriate measurement procedures and to define the attainable accuracy. It is thus essential to establish and operate a few independent calibration laboratories [3]. Here, university departments have an important role to play. Their investigations have already shown weaknesses of instruments and lead to improvements.

Digital levels calculate the staff reading by processing the image of the coded staff which propagated through the atmosphere and the optical elements of the level. To assess the influence of defective system components (equipment, software) on the measurement result, [4] suggested to use system calibration. For levels, the basic idea is to carry out a height reading with the level, move the staff by a known amount, carry out another height reading, and so on. The performance of the whole system can be derived from the differences of the height readings by the level and the true height changes.

At the Graz University of Technology (TUG) a calibration facility for digital levelling systems has been developed. Its original design was described by [5]. An assessment of the required accuracy showed that the comparator must perform at the micrometer level. Thus a complete redesign of the calibration facility became necessary to achieve this high precision. Special features of the TUG comparator are the mounting of the staff in its position of use (thus called

„vertical comparator") and the possibility to use sighting distances between 1.5m and 30m.

Using the vertical comparator at TUG, we have investigated the error pattern of the available digital levels [6]. In addition, we could show that system calibration is capable of determining the composite scale value of the staff and the level [7].

Currently, several institutions are considering to build a vertical comparator. Therefore here, we give a detailed report about the design of the vertical comparator at TUG. The hardware components are described in chapter 2, and the peripheral equipment for e.g., acquisition of meteorological data, in chapter 3. The vertical comparator system software is summarised in chapter 4 including a description of the calibration procedure. Finally, in chapter 5 the uncertainty of measurement using the comparator is estimated.

2. Design and Hardware

The performance of levelling systems depends on various factors, e.g., temperature, illumination, sighting distance. When testing an instrument, only one of these factors should be varied during the experiment to investigate the system's response. For system calibration, the height readings are varied by changing the staff's position. All other parameters should remain unchanged. This can be achieved in a laboratory.

At the TUG the Geodetic Metrology Laboratory (GML) was established during the last decade. The laboratory has a size of 33.2 x 6.3 x 3.5m^3 and is climatically controlled (temperature: 20.0°C $\pm$ 0.5°C, humidity: 50% $\pm$ 10%). The GML is situated on the ground floor of a building and its foundation is completely separated from the foundation of the building. Thus movements of the building induced by temperature, wind or traffic are reduced. Only artificial and therefore reproducible light is used in the GML.

The two photographs of the vertical comparator (fig. 1) provide an impression of the calibration facility.

The main parts of the comparator are: (1) a carriage for the level, (2) the frame of the comparator with a carriage moving the staff vertically, (3) the laser interferometer to measure the position of the staff, (4) the staff illumination assembly, and (5) the comparator system software,

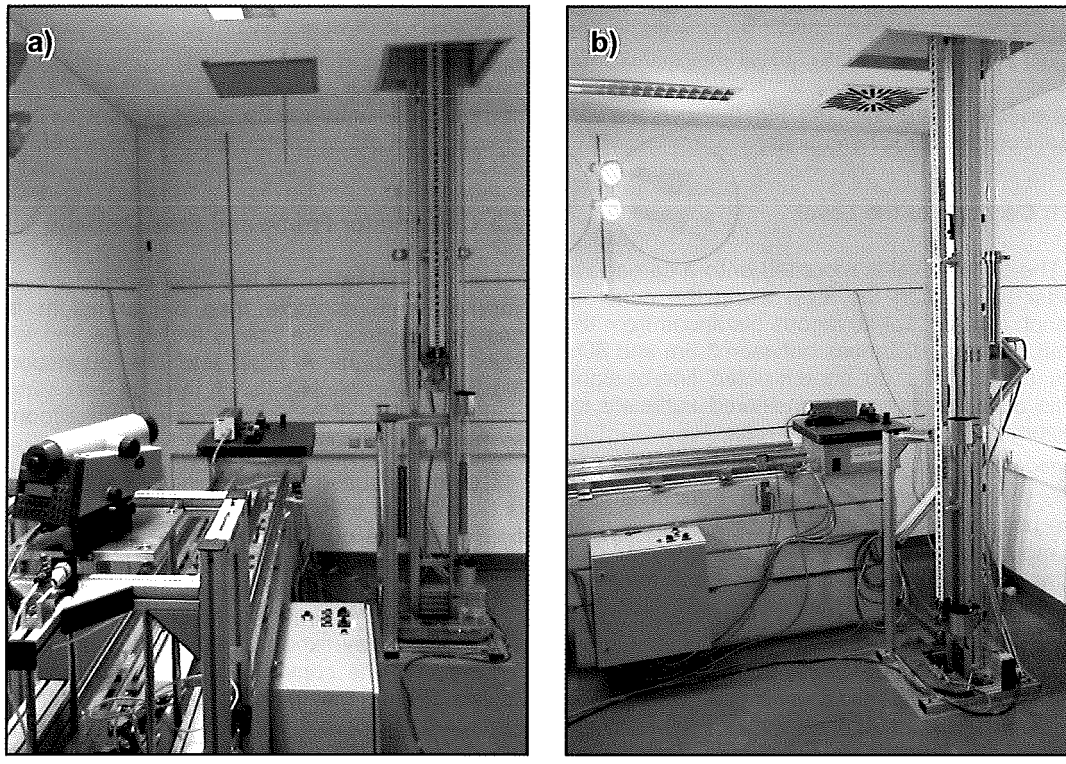


Figure 1: Overview of the vertical comparator showing (a) the level and (b) the staff illumination assembly.

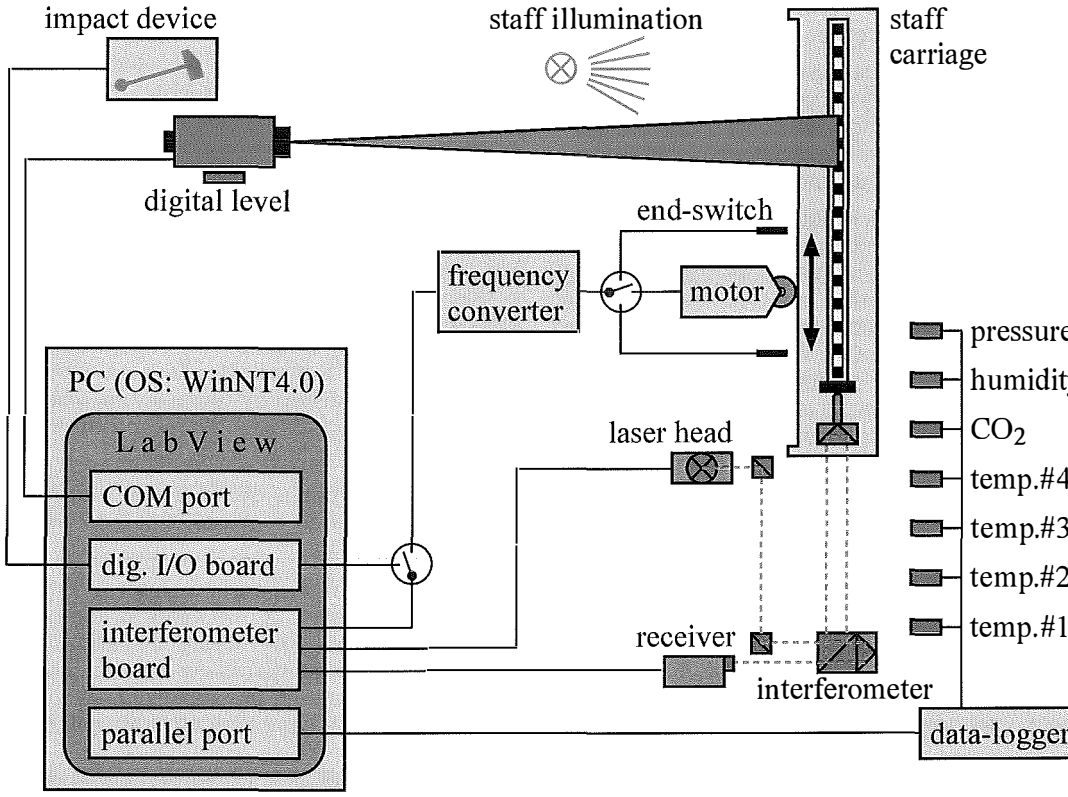


Figure 2: Components of the vertical comparator.

installed on a standard PC. Fig. 2 shows the vertical comparator schematically.

2.1. Carriage of the Level

The digital level is mounted onto a carriage in order to position the level along the concrete bench (see fig. 1a) at various distances from the staff. Sighting distances between 1.5m and 30m are possible for an unobstructed line-of-sight. This distance range is considered sufficient for calibrating digital precision levels.

The carriage consists of a wheel system and two separate frames (see fig. 3). Four invar rods are used for the inner frame on which the level is mounted. Invar is used to keep the level at a constant height independent of small temperature variations during the whole calibration process.

The second frame, made of robust aluminium profiles, surrounds the invar frame, see fig. 3. It is used to mount additional equipment, as for example, a pneumatic impact device. This impact device is optionally used to activate the level's compensator before each measurement for the investigation of the compensator's behaviour.

2.2. Comparator Frame and Staff Carriage

For the calibration of 3m long invar staffs, a vertical frame of more than 6m in height is needed, reaching 3m above and below the level's line-of-sight. A shaft into the foundation (fig. 4a) and an insulated shaft through the ceiling of the GML had to be built in order to make room for the 6.5m tall assembly.

The frame consists of aluminium profiles and is fixed to the foundation of the laboratory, see fig. 4a. As the comparator frame and the 30m concrete bench are on the same foundation, they cannot move differentially to each other. Consequently, the interferometer and the level's line-of-sight stay fixed in space which is a prerequisite for the construction of a comparator. At the ceiling, the frame is guided only – not mounted – to keep it free of tensions. The guiding device at the ceiling is used to adjust the comparator to its vertical position.

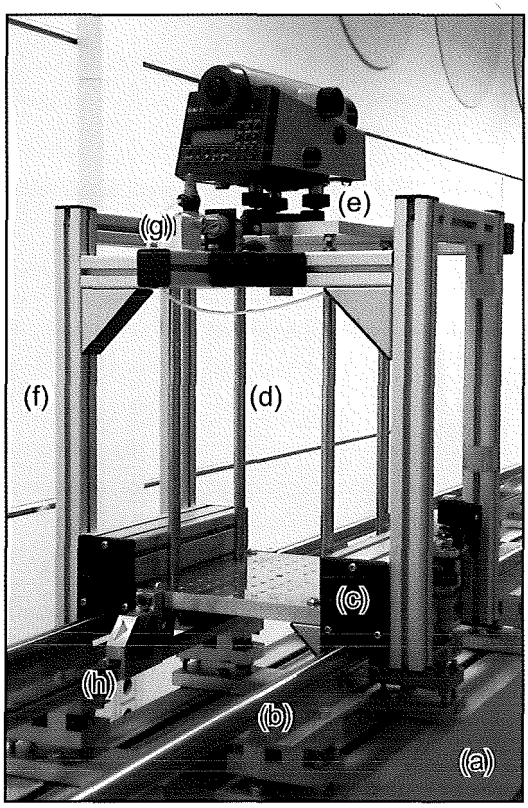


Figure 3: Level carriage: (a) 30m concrete bench, (b) rail system, (c) wheel system of carriage, (d) invar rods, (e) fastening plate of level, (f) aluminium frame, (g) impact device, (h) displacement protection.

The invar staff is mounted to a 3.4m long carriage. The carriage can be moved along two rails using a wheel assembly which is driven by an AC motor. The control signal for the motor is generated by a frequency converter coupled to the interferometer board.

The invar staff is set up on a bolt, see fig. 4b, and mounted to the carriage using two mounting brackets. These brackets allow the rotation of the staff by $\pm 90°$ which can be used to direct the staff towards the level, when needed. The rotation axis coincides with the plane of the staff's invar band as weil as the centre of the set-up bolt which is exactly below the invar band.

2.3. Interferometer Hardware

The staff carriage is monitored by a Hewlett-Packard interferometer consisting of a Zeemann stabilised laser head (HP5517B), a HP10702A linear interferometer, a remote receiver HP10780F with a fibre optic cable, an interferometer board (HP10889B PC Servo-Axis board), and additional optical accessories.

The resolution of the linear interferometer system is specified as $\lambda/128$, the nominal wavelength λ of the laser being 633nm (rounded) with a specified vacuum wavelength accuracy of ± 0.02ppm. To avoid a scale error, a calibrated laser head is used with a relative error of the laser frequency of 6.6×10^{-9}. For further details about particular interferometer measurements, reference is made to [8, p.86–122].

2.4. Interferometer Set Up

To adhere to Abbe's comparator principle (see e.g., [8, p.32]), the light path of the interferometer is adjusted to be in the same axis as the staff's invar band. The retroreflector is mounted at the lower end of the set-up bolt (see fig. 4b) which is also made of invar.

The interferometer is placed near the lower end of the bottom shaft, in the same axis as the staff's invar band. Due to the small diameter of the shaft and its inaccessibility, the interferometer had to be mounted on a platform that can be lowered into the shaft from the laboratory level. The main components of this structure are three invar rods of 1.8m length. The use of invar was necessary, because the temperature in the bottom shaft can be up to 6K lower than the air temperature of the laboratory (see fig. 5b).

All optical parts of the laser interferometer need to be properly aligned. The special design of the interferometer bearing unit and the arrangement of the components simplify this procedure. For the alignment of the laser beam also the beam benders outside the shaft may be used. These are mounted on a frame which is completely separated from the comparator's frame (see fig. 4a) to avoid any influence of a possible deformation of the comparator frame on the laser beam.

3. Peripheral Equipment

3.1. Staff Illumination

The current digital levels use CCD arrays which are sensitive in different regions of the spectrum. For the calibration of all types of digital levels, the illumination of the staff must cover the appropriate ranges of the spectrum. Four light bulbs (Phillips PAR38-EC) were chosen for this purpose. Two of them can be seen in fig. 1b.

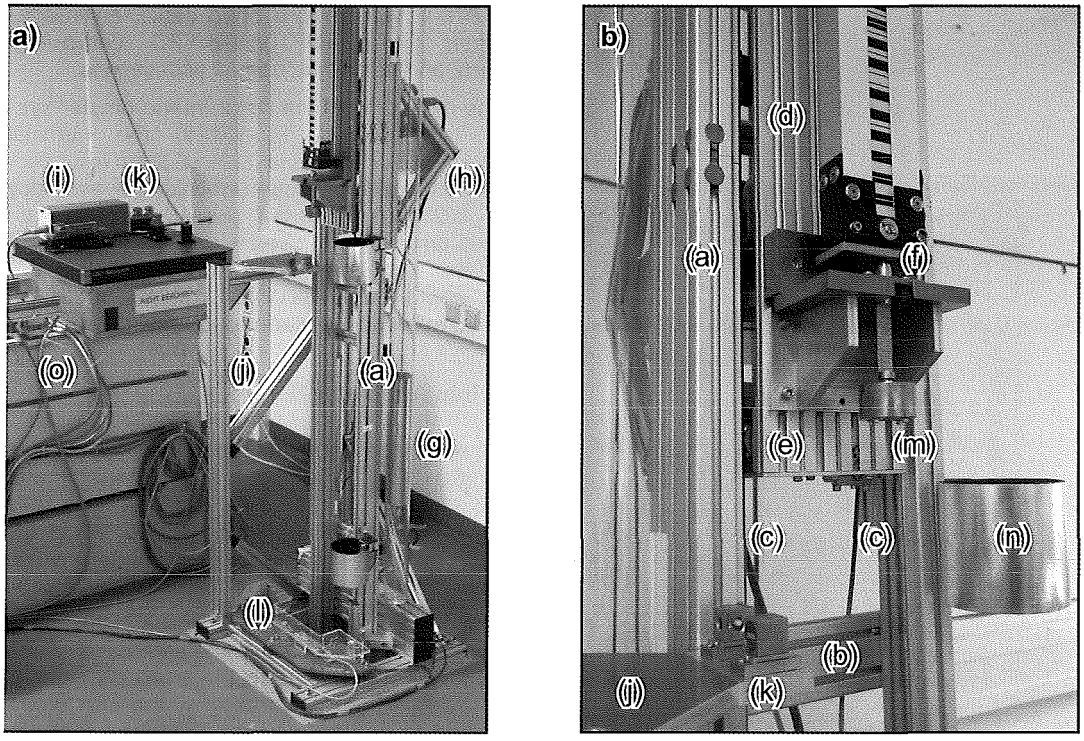


Figure 4: Vertical comparator's (a) lower part, (b) staff carriage.

Currently, the alternative use of a neon lamp (1.5m long and vertically mounted) is being investigated. It might be useful for special investigations, where a more homogeneous illumination is necessary.

3.2. Meteorological Equipment

The wavelength of the interferometer depends on the ambient air's refractive index which can be calculated using meteorological data. In the laboratory, the main influential parameters are temperature, air pressure, humidity and the carbon dioxide content of air.

For a distance accuracy of 0.1ppm the required accuracies of the meteorological equipment are: 0.1K for temperature, 0.37hPa for air pressure and 12% for relative humidity. The CO_2 content should be known to 680ppm. Details about the chosen meteorological sensors are described by [2].

Due to variability of the air temperature along the laser beam path four glass-covered Pt100 temperature sensors are used in different positions, see fig. 5a. The sensors are mounted to the frame of the comparator in a thermally isolated manner and protected from heat radiation (caused by e.g., the staff illumination) by a plastic cover, see fig. 4b. The accuracy of the temperature sensor is about 0.05K.

The other three meteorological parameters are measured at one position only (fig. 5a). The sensors have the following accuracies: 0.3hPa for air pressure, 3.5% for relative humidity, and 25ppm for CO_2 content. The temperature sensors were calibrated at 0°C and approx. 22°C, using a precision glass thermometer with a resolution of 0.01K. A laboratory mercury barometer with a resolution of 0.1hPa was used to determine the offset of the pressure sensor. For the humidity and CO_2 sensors the factory calibrations were used.

3.3. Representative Meteorological Parameters

Fig. 5a shows a cross section of the vertical comparator, with the interferometer being placed near the bottom of the shaft. Though the laboratory is climatically controlled, the temperature in the shaft is different. Compared to the temperature in the laboratory it is lower by up to 6K depending on the ground temperature. Fig. 5b

shows an example of the measured temperature distribution near the laser beam path.

The data of the four temperature sensors are used to approximate the vertical temperature distribution. Then, a single representative temperature value is computed for the actual laser beam path, depending on the position of the staff carriage. Similar to that, one representative value for the air pressure is computed, using the measured air pressure and the barometric height formula.

These values are used to calculate the refractive index of air and the proper atmospheric propagation correction for the laser path. The selection of the most accurate formula for the computation of the refractive index of air was investigated by [2]. Currently, the formula of [9] is used, however, it is planned to implement the resolution No.3 of IAG, 1999, see for example [10].

4. System Software

The vertical comparator system software (VCSS) is used for data acquisition and the control of the entire comparator. It was written using the graphical programming environment LabView5.0 and is installed on a standard PC running the Windows NT4.0 operating system. Drivers for the HP10889B interferometer board and for all current digital precision levels were developed. VCSS provides (a) an easy set up of a calibration run, (b) a fully automatic execution of the calibration, and (c) an output of a log-file which contains extensive information on the calibration run.

The initialisation of the system comprises also the input of the calibration parameters such as the type of level, the staff, and the positions of planned staff readings.

Before the calibration run can be started, a reference measurement with the level is needed to determine the distance between the interferometer and the staff at its initial position. Three modes are available for the reference measurement. Using the most accurate mode, the round-off error of the level is considered to yield a precision of the staff's position that is better than the resolution of the level.

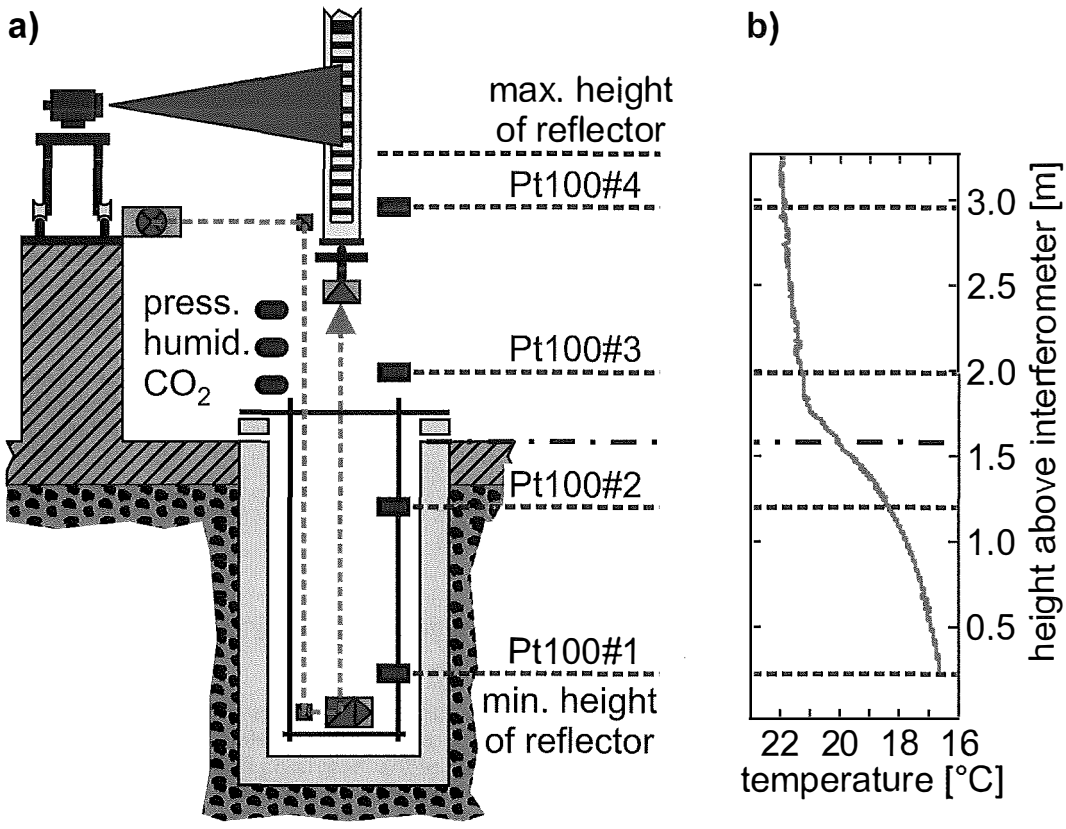


Figure 5: (a) Vertical comparator and distribution of sensors. (b) Measured temperature distribution.

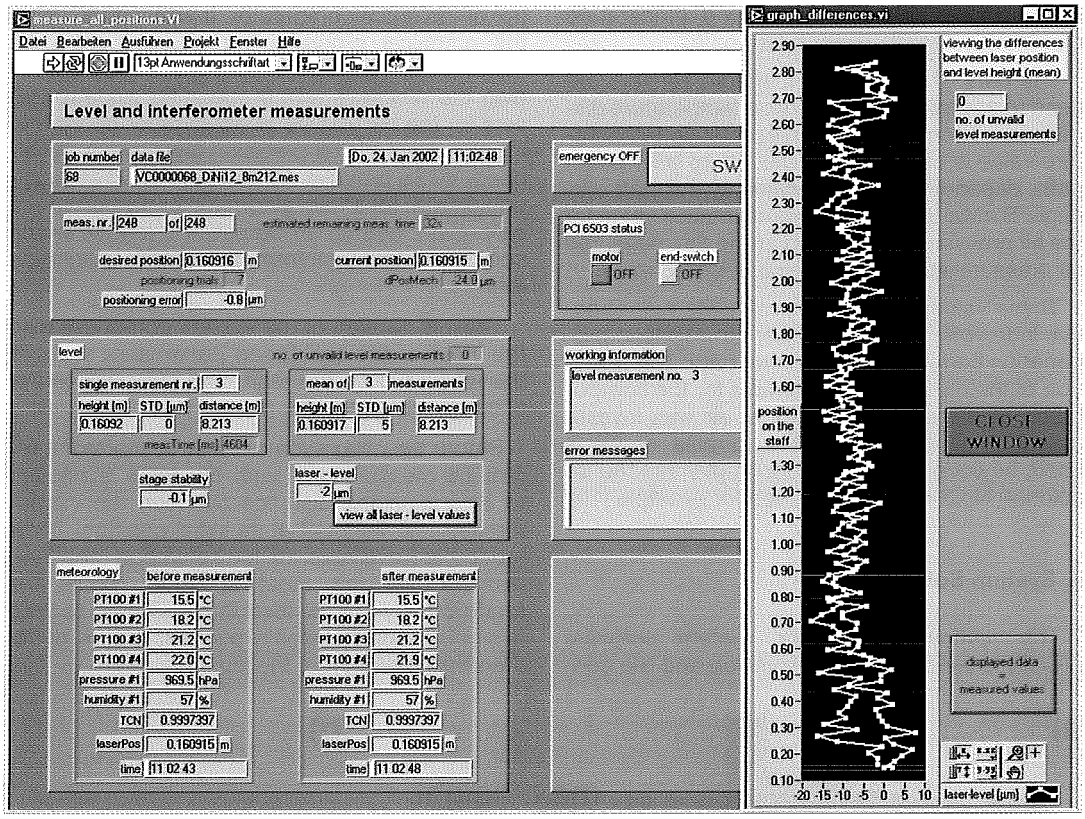


Figure 6: Snapshot of the vertical comparator system software's main window.

Once the calibration run has been started, a PC window provides graphical information about the position of the staff, the positioning status, actual meteorological data and the level measurements, see fig. 6.

The principal sequence of operation is as follows. Before the staff carriage is driven to a desired position, the refractive index of air is computed and the meteorological compensation factor of the HP10889B board is updated. Then the staff is moved to a specified position. Immediately afterwards, the HP10889B output signal is interrupted to ensure that the staff remains stable, whilst using the level. For the signal interruption a separate digital I/O board (National Instrument PCI6503) is used. The same board can be used to activate the impact device (see fig. 2). The program's execution is paused for half a second, before a position is read from the interferometer. This is done to avoid measurements possibly affected by an instability of the carriage due to oscillations. All the mechanical imperfections of the comparator's hardware cause a difference of 10 to 20*m between the carriage's settling position and the specified position. To obtain a positioning accuracy that is better than these values, several positioning trials are carried out until the positioning error is less than 2*m (i.e., current software setting), or a maximum number of trials is exceeded.

Whenever the level's impact device is activated, the program is paused to let the compensator settle down. Afterwards, the level measurement is started. An important quality control feature is the comparison of the interferometer readings before and after each measurement by the level.

5. Standard Uncertainty of the Vertical Comparator

The fundamental measuring unit of the comparator is the laser interferometer with the frequency of the laser head defining the „metre". A basic assumption is that the relative position of the interferometer and the level remains constant during a calibration run. However, for example

the thermal expansion of the interferometer bearing unit or of the level carriage, or a possible inclination of the laboratory's foundation might cause distortions which affect the measurements by the vertical comparator. The influence of some parameters may be eliminated by an adequate calibration procedure (e.g., [11]), however, a knowledge of the remaining influences is essential for quoting the comparator's uncertainty.

The ISO/BIPM „Guide to the Expression of Uncertainty in Measurement" [12] allows to estimate the uncertainty of the complex measurement system, taking into account also quantities that cannot be measured (e.g., [13]). First a model of the measuring process must be established. We start with the distance measurement L by the interferometer:

$$L = (C + \Delta C^{E} + \Delta C^{ON} + \Delta C^{OD}) \cdot \frac{\lambda}{R \cdot n} \cdot \cos\alpha - \Delta L^{IS} + \frac{D}{\Delta n} \quad (1)$$

Each term in eq. (1) is explained in tab. 1. To assess a vertical comparator measurement H, the external parameters of influence must be considered:

$$H = A - L - \Delta L^{S} + \Delta L^{LC} + \Delta L^{LOF} + \Delta L^{FC} \quad (2)$$

Also the terms of eq. (2) are listed in tab. 1. Additionally, the estimates of the standard uncertainties of the terms are given in tab. 1. They were determined using the results of dedicated experiments. Where experimental values were not available, the values were assessed using experience or were obtained from literature. Some of the standard uncertainties listed in tab. 1 had to be estimated using the GUM procedure, e.g., the combined standard uncertainty of *n* which was determined using the uncertainties of the meteorological sensors, of the measurement and the formula used.

The "law of propagation of uncertainty" [12] was applied to eqs. (1) and (2) to determine the combined standard uncertainty $u_c(H)$ for an interferometer distance of 3m. In this paper, the partial derivatives of eqs. (1) and (2) are not explicitly stated. To determine the expanded standard uncertainty U(H) of a comparator measurement H, a coverage factor of k = 2 was used, giving $U(H) = \pm 2.7\mu m$. With this factor the level of confidence is approx. 95%.

The value 2.7μm determined by the GUM procedure is in excellent agreement with a prior assessment based on repetitive system calibration runs. So for example, using the data presented by [7], an overall accuracy of the TUG vertical comparator of better than 3μm was estimated.

Acknowledgements

We want to thank the Österreichische Nationalbank for partly funding the reconstruction of the vertical comparator, Leo Gruber for elaborating many constructive plans, and Robert Presl and Rudolf Lummerstorfer for their work to upgrade the vertical comparator.

Symbol	Description	Standard Uncertainty
C	number of counts measured by the interferometer	27.7counts
ΔC^{E}	interferometer electronic error	0.3counts
ΔC^{ON}	interferometer optics non-linearity	0.6counts
ΔC^{OD}	interferometer optics thermal drift	10.1counts
λ	wavelength of the laser head	0.01ppm
R	resolution of the interferometer	–
n	refractive index of air	0.13ppm
α	cosine error	1mm/3m
ΔL^{IS}	move of the interferometer due to thermal expansion of the interferometer bearing unit	0.8μm
D	deadpath distance	10mm
Δn	change of the refractive index during the calibration run	1.3ppm
A	comparator constant; vertical spacing between the interferometer and the level	–
ΔL^{S}	thermal expansion of the staff's invar band	0.6μm
ΔL^{LC}	thermal expansion of the level carriage due to temperature changes in the laboratory; causes a vertical move of the level	0.1μm
ΔL^{LOF}	change of the level's line-of-sight during a calibration run	0μm
ΔL^{FC}	inclination of the laboratories foundation concrete during a calibration run	0μm

Table 1: Description of terms and uncertainties.

References

[1] *Ingensand H (1999)* The evolution of digital levelling techniques – limitations and new solutions. In Lilje M (ed.): The importance of heights. FIG, Gävle, Sweden: 59–68.

[2] *Woschitz H (2003)* System Calibration of Digital Levels: Calibration Facility, Procedures and Results. Shaker Verlag. In print.

[3] *Brunner FK, Woschitz H (2001)* Kalibrierung von Messsystemen: Grundlagen und Beispiele. In Heister H und Staiger R (ed.): Qualitätsmanagement in der Geodätischen Messtechnik. Konrad Wittwer Verlag, DVW Schriftenreihe 42: 70–90.

[4] *Heister H (1994)* Zur Überprüfung von Präzisions-Nivellierlatten mit digitalem Code. Schriftenreihe der Universität der Bundeswehr München 46: 95–101.

[5] *Reithofer A, Hochhauser B, Brunner FK (1996)* Calibration of Digital Levelling Systems. Österr. Z. f. Vermessung und Geoinformation 74: 284–289.

[6] *Woschitz H, Brunner FK (2002)* System Calibration of Digital Levels - Experimental Results of Systematic Effects. In Kopácik A and Kyrinovic P (eds.): Proc. INGEO2002 - 2nd Int. Conference of Engineering Surveying, Bratislava, November 2002: 165–172.

[7] *Woschitz H, Brunner FK, Heister H (2003)* Scale Determination of Digital Levelling Systems using a Vertical Comparator. Z. f. Geodäsie, Geoinformation und Landmanagement 128: 11–17.

[8] *Heister H (1988)* Zur automatisierten Kalibrierung geodätischer Längenmessinstrumente. Schriftenreihe der Universität der Bundeswehr München 27: 210 pages.

[9] *Bönsch G, Potulski E (1998)* Measurement of the refractive index of air and comparison with modified Edlén's formulae. Metrologia 35: 133–139.

[10] *Rüeger JM (2002)* Report 1993–1999 of the Ad-Hoc working Party of the IAG Special Commission SC3 (on Fundamental Constants, SCFC) on Refractive Indices of Light, Infrared and Radio Waves in the Atmosphere. In Rüeger JM (2002): Refractive Indices of Light, Infrared and Radio Waves in the Atmosphere. University of New South Wales, Unisurv Report S-68, 2002: 55–80.

[11] *Rüeger JM, Brunner FK (2000)* On System Calibration and Type Testing of Digital Levels. Z. f. Vermessungswesen 125: 120–130.

[12] *ISO/BIPM (1995)* Guide to the Expression of Uncertainty in Measurement. International Organisation of Standards, Switzerland.

[13] *Heister H (2001)* Zur Angabe der Messunsicherheit in der geodätischen Messtechnik. In Heister H und Staiger R (ed.): Qualitätsmanagement in der Geodätischen Messtechnik. Konrad Wittwer Verlag, DVW Schriftenreihe 42: 108–119.

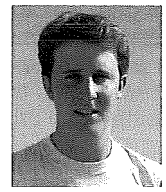
Contact

Dr. Helmut Woschitz, Univ.-Prof. Dr. Fritz K. Brunner: Engineering Geodesy and Measurement Systems, Graz University of Technology, Steyrergasse 30, A-8010 Graz, Austria. E-mail:
helmut.woschitz@tugraz.at,
fritz.brunner@tugraz.at

Evaluation of the Vibrational Spectrum of High Slim Towers with Wind Electrical Turbines

Johannes Fabiankowitsch and Heribert Kahmen, Wien; Phillip Matt, Natters





Abstract

To date, the movements of wind turbines during operational or idle periods have hardly been investigated at all. There is also little knowledge of the service lifespan of steel towers. Therefore extensive measurements were made, in cooperation with a power company, to arrive at new research results. The vibrational spectrum of high slim windmill towers were continuously and automatically recorded by a combination of accelerometer systems and point-positioning by satellite (RTK-GPS). Of particular interest were the movements of the tower during the rotor startup and braking phases and during full operation. An electronic measuring system, based on a PC, managed the signal conditioning, the measuring sequences and the evaluation.

Zusammenfassung

Bewegungen von Windkraftwerken außerhalb und während des Betriebes sind bisher noch wenig erforscht. Außerdem gibt es noch kaum Erfahrung für die Lebensdauer der Stahltürme. In Zusammenarbeit mit einem Energieunternehmen sind daher umfangreiche Messungen durchgeführt worden um hier zu neuen Forschungsergebnissen zu kommen. Die Bauwerksbewegungen an hohen schlanken Windgeneratortürmen werden durch eine Kombination von Beschleunigungssensoren und der Positionsbestimmung mittels Satellitenverfahren (RTK-GPS) kontinuierlich und automatisiert erfasst. Bei der Erfassung von Vibrationsspektren einer Windkraftanlage sind streng genommen verschiedene Betriebsphasen zu unterscheiden. Die Phase 1 ist gekennzeichnet durch den Stillstand der Rotorblätter, d.h., die Anlage ist abgeschaltet. In dieser Phase können vor allem die durch äußere Einflüsse (Sonne, Wind) hervorgerufenen Turmbewegungen, unabhängig von Anregungen durch die drehenden Rotorblätter, studiert werden. Es sind kurzperiodische Schwingungen auf Grund von Windeinflüssen und langperiodische Verformungen, die durch thermische Einflüsse (Stand der Sonne) hervorgerufen werden, zu unterscheiden. Von besonderem Interesse sind die Turmbewegungen während des Hochlaufes (Bremsens) des Rotors, welche als Phase 2 bezeichnet wird. Sie gilt als besonders kritisch, weil unter anderem dabei die sogenannte Eigenfrequenz des Turmes durchlaufen wird. Ein zusätzlicher Effekt kann eintreten, wenn vor dem Hochlaufen des Rotors oder während des Betriebes (Phase 3) die Gondel in eine neue Windrichtung nachgestellt werden muss. Dabei werden ja immerhin 63 Tonnen bewegt. Die Phase 3 betrifft den Arbeitsbetrieb. Durch die drehenden Rotorblätter ist eine Vielzahl von Vibrationen nachweisbar. Diese meist hochfrequenteren Schwingungen überlagern sich der Eigenschwingung des Turmes. Dank einer PC-Messelektronik, welche den Messablauf, die gesamte Signalkonditionierung und ein umfangreiches Auswerteprogramm bereitstellt, werden die Daten auf einem Laptop bearbeitet und ausgewertet.

1. Introduction

The use of natural sources of power has grown in importance significantly in recent years. Wherever landscape conditions combine to produce good, but more important, constant winds the temptation to use the prevailing winds as a source of energy is obvious. Consequently, by October 2002, one hundred boreal power stations had been erected in the federal state of Lower Austria alone.

As a consequence of their slim form, their height and their normally exposed locations, these towers are subjected to extreme influences, caused, either by the operation itself (rotation of the rotor) or by environmental conditions (wind, sun). To date, the cause and effect of induced tower movements have hardly been investigated at all.

Continuous recording of such tower movements put high demands on the measuring equipment. The circumstances of the tower to be studied, the environmental conditions and the wishes of the client had to be taken into account during the selection of the measuring-sensors.

The wind-generator-tower in question has a height of 60m, diameter at the base of 4.2m tapering to a diameter of 2.0m at the height of 60m. It consists of two hollow steel segments of

30m each, screwed together. The rotor-blades are 27m long. The rotor, including rotor-blades, weighs 25 tons. The rotatable gondola, including generator, weighs 38 tons. The rotor-blades are of the *"Active Stall"* type, meaning that they adjust their trim according to the prevailing wind velocity. This makes an even, steady rotation of the rotor possible during fluctuating winds. The gondola, which is detached from the tower, can follow the wind. This happens then, when the wind direction deviates from the direction of the rotor-axis more than ten degrees in the middle of a minute.

2. Assignment

The recording of the vibrational spectrum of a wind-generator-tower is to be strictly divided into different operational phases. With the help of suitable sensors, the goal is to measure these different operational phases.

Phase 1 is when the boreal power station is out of operation and the rotor-blades are still. This is the case when maintenance or repair work is carried out. During this phase, tower movements caused by environmental influences (sun, wind) can be studied independent of the influences of the rotating rotor-blades. Short-term vibrations caused by wind should be differentiated from the long-term deformations caused by thermic influences (position of the sun).

Phase 2 is characterised by the startup and acceleration or deceleration and shutdown of the rotors. This is considered to be especially critical because the so-called resonant frequency of the tower is passed through. An additional effect can occur when the gondola is adjusted to a new wind direction during startup or during operation (Phase 3). 63 tons are moved during this process.

Phase 3 is fully operational service. As a result of the rotating rotor-blades, a multitude of vibrations can be detected. These, mostly high frequency vibrations, are superimposed on the resonant frequency of the tower.

A combination of accelerometers and point positioning by satellite (RTK-GPS) were used to record the vibration spectrum of this tower. The RTK-GPS data was used as reference data.

3. Electronic control, sensor selection and synchronisation

The use of different sensors with varying physical operating principles make it neccessary to use a central controller. The data from the sensor and the positional data from the satellite receiver must also be synchronised to establish an unequivocal time definition. The measurement data was transferred from the central controller to a laptap where evaluation software, with a multitude of sub-programs ranging from data-recording through to the analysis of the data, make an online evaluation possible.

The electronic measuring system for PC *"Spider 8"* [1] was used as central controller. This system can be used to electronically measure mechanical dimensions such as movement, acceleration and speed. This system has four complete digital measurement-amplifiers using 4.8kHz carrier-freqency technology. It handles the complete signal-conditioning. This should be understood as the supply for passive reception, the amplification of the input-signal, the connector technology for maximal eight channels, a computer interface and the digitalisation. Each of the eight channels has ist own A/D converter allowing measurement rates of 1/second to 9600/second. These A/D converters are synchronised with each other, guaranteeing time-synchronised measuring through all channels.

With the selection of sensors, it is important to define whether movement or vibration is to be measured. A vibration measurement is concerned with the quadratic response of the object of investigation. A movement measurement is concerned with the speed or the shift of a static body (or part thereof). Three different sensors were used for this first projekt-study.

The accelerometer type B12 [2] works on the principle of induction, which, viewed mechanically is a highly-tuned spring-mass-system and viewed electrically is a passive supplier. Small dimensions (cylindrical form, 34 x 12.6mm) and minimal weight (17gr) are their characteristics. Two different types are to be used, the B12/200 and the B12/500. They differ in their assigned frequency and their measuring range. The B12/200 has a working frequency range of 0 Hz – 100 Hz, the B12/500 from 0 Hz – 250 Hz. As the name suggests, the B12/200 has a resonant frequency of 200 Herz, the B12/500 a resonant frequency of 500 Herz.

The accelerometer ISOTRON MODAL-63A-50 [3] is an active sensor that functions on the piezo-electric principle. It is characterised by low weight (20 gr), small exterior dimensions (cubes of 22.35mm), the integrated electonics but above all the capability to record accelerations in three, to each other, orthogonal axis simultaneously. From the specifications, we can take

the resonance frequency as being approx. 15000 Hz and that the sensor can be used in the 1 Hz – 2000 Hz range.

The GPS System 530 from the Leica company was used as the third sensor. Point-positioning will be carried out with a frequency of 10 Herz. For this purpose, a fix reference point was established in the vicinity of the wind-generator-tower and two receivers were mounted on the roof of the gondola. An important advantage is that extremely low frequencies such as the natural vibration of the building and the generally expected vibrations of the tower can be recorded without phase displacement.

The measurements from the accelerometers were synchronised with each other on the basis of the CPU-time of the laptop. However, the position data is related to GPS-time, which has the characteristic, to deviate from Atomic-time by a constant integer value in seconds. At the time of the measurements this value was 13 seconds.

The synchronisation between CPU-time and Atomic-time was achieved using the programme *"Atomsync 115"*. Standard differences between CPU-time and Atomic-time after the synchronisation ranged from 0.01 – 0.05 seconds. These differences arose due to delays during the internet transfer.

The synchronisation was carried out before the measuring began and again after the measuring was finished to calculate the time-drift, which was then linear interpolated.

4. Measuring Setup

The RTK-GPS setup is described in paragraph 3. The three accelerometers were mounted together on a platform. The internal alignment of the sensors on this platform was so that the measurement-axis of the B12/200 and the B12/500 were perpendicular to each other. The ISOTRON 63A-500 was then orientated so that its x-axis was in line with and its y-axis perpendicular to the B12/500. Its z-axis was aligned vertically perpendicular.

The complete measuring system – the platform with the sensors, the measurement-amplifier for the ISOTRON 63A-500, the central controller *"Spider 8"* and the laptop – was assembled in the last chamber of the tower under the rotatable gondola. A geodetic orientation of the platform was not possible as there was no "line-of-sight" connection to outside.

Unlimited access to the wind-generator-tower was garanted for three days. Coincidentally, maintenance work was also carried out during this time, which allowed measurements of Phase 1 (see paragraph 2) to be made. Measurements of Phase 2 could be carried out three times. Un-

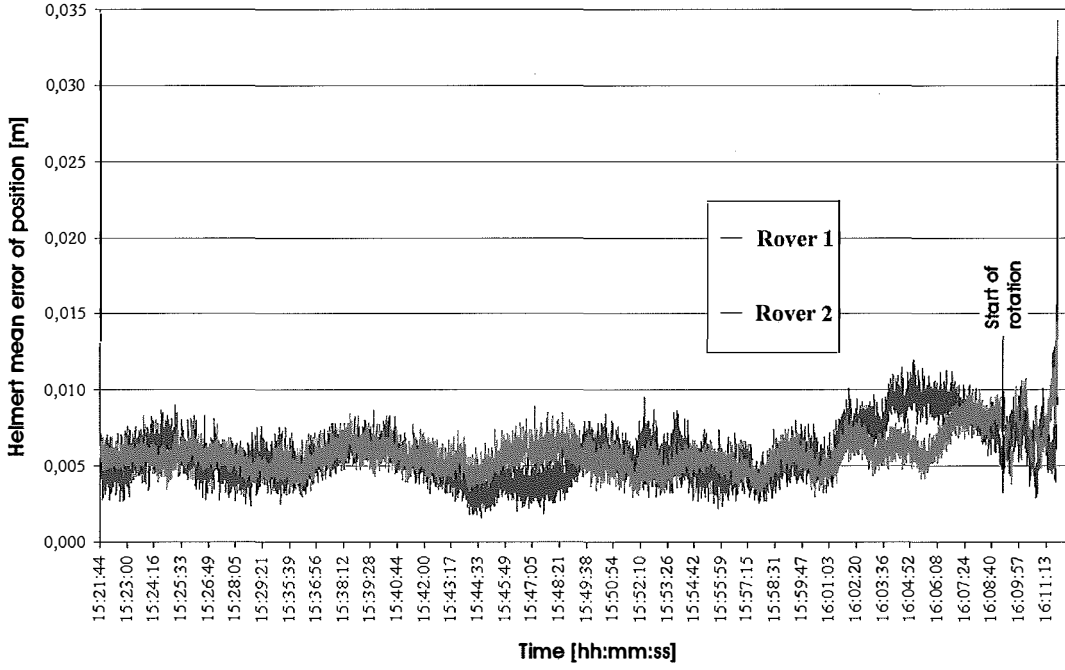


Figure 1: Helmert mean error of point position of both rovers (phase 1)

fortunately the wind was moderate during the three days and mostly from south-west.

5. Measurement results

In the calculation of tower movements in the frequency range, the resonant frequency of the tower takes a special place within the expected frequency spectrum. An estimate of the natural period of vibration T and the resonance frequency f under consideration of static wind energy [4] yields T = 2.73 seconds and f = 0.37 Herz.

5.1 Measurements with the RTK-GPS-system

Useful measurement data could only be obtained during Phase 1 (see paragraph 2), because only during Phase 1 were the obstructions and/or multipaths of the measuring signal negligible. Figure 1 shows the Helmert mean error of point position of both rovers.

It was shown that during machinery out of operation, the point-positioning determined with the RTK-GPS was accurate to approx. 5mm. There was a noticable reduction in positioning accuracy after 16:02. A reason could be the screening of the satellite signal for rover 2 by a stationary rotor-blade. The positioning accuracy of rover 1 is also reduced, but remains significantly better than rover 2. Similar causes can be assumed for the times 15:47 and 16:04 when the deviation of the positioning accuracy between the two rovers reach their maximum.

At about 16:08 the gondola begins to turn into the wind in preparation for operation (Phases 2 and 3). The accuracy of mean error of position of both rovers alternate quickly and deteriorate rapidly. As a result of the steady increase in rotation, the intervals between signal-reception and interruption become ever shorter. After about 16:11 the two rovers are no longer able to determine the ambiguities. With fewer than four satellites position determination is not possible. The amplitude spectrum of the RTK-GPS data shown in figure 1 is shown in the following figure 2.

Both rovers show a clear amplitude up to 3.6mm at a frequency of 0.413 Herz. A remarkable correlation to the resonant frequency that was estimated at the outset. Around 0 Herz amplitude values up to 3mm are seen. These values are a result of the mathematical approach and have nothing to do with the movements of the tower.

5.2 Measurements with the accelerometers

The analysis of the relevant data is not limited to Phase 1, but can be carried out from standstill

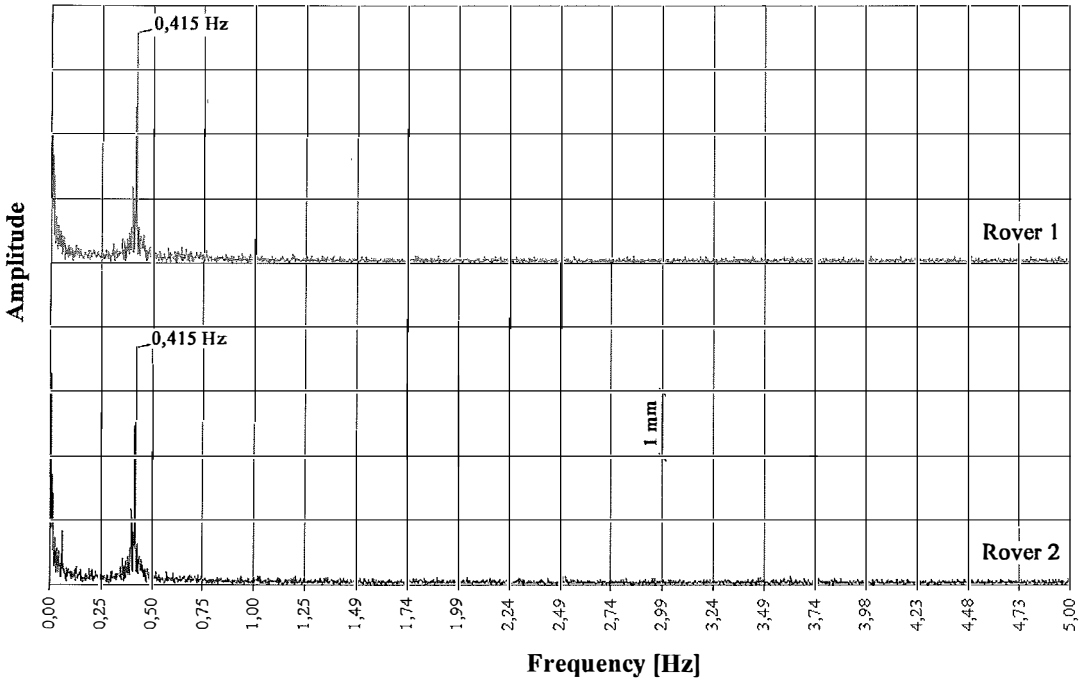


Figure 2: Amplitude spectrum of the RTK-GPS data (phase 1)

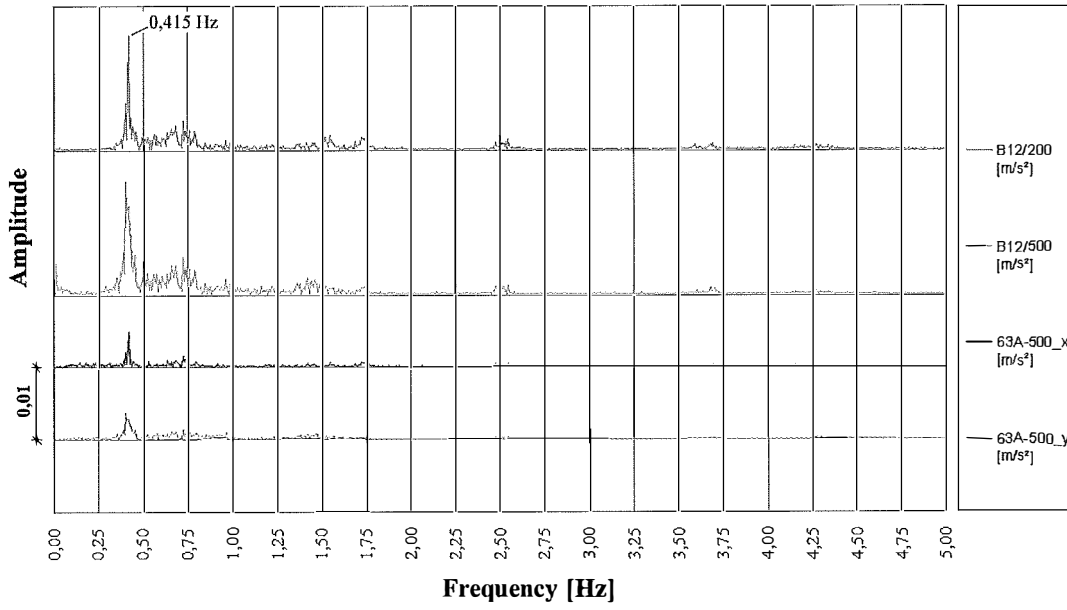


Figure 3: Amplitude spectrum of the accelerometers (phase 1)

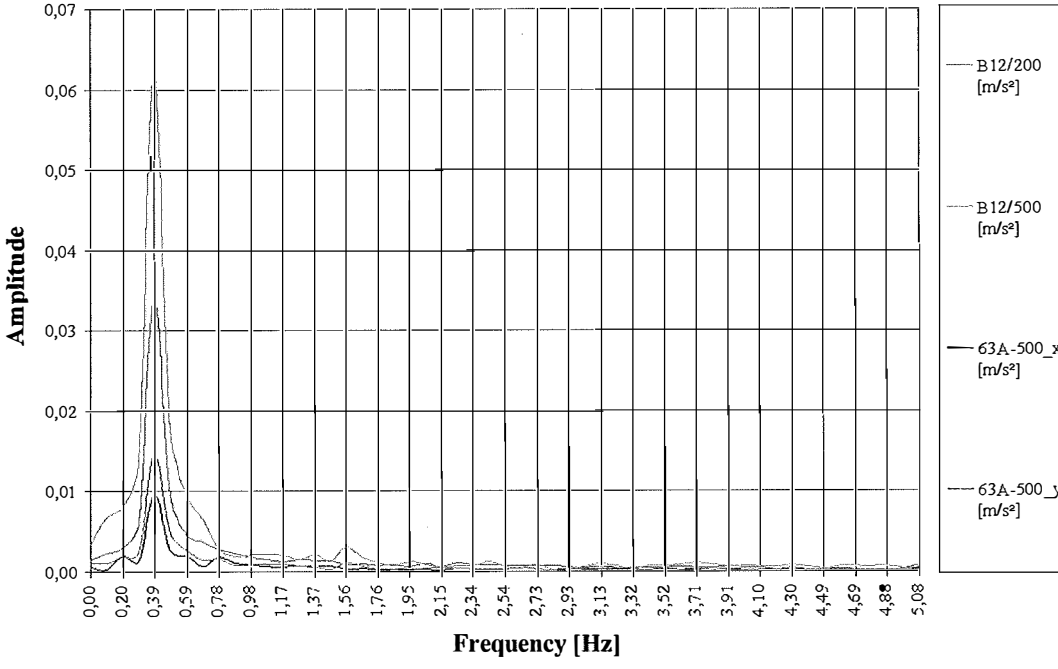


Figure 4: Amplitude spectrum of the accelerometers (gondola mouvement-braking)

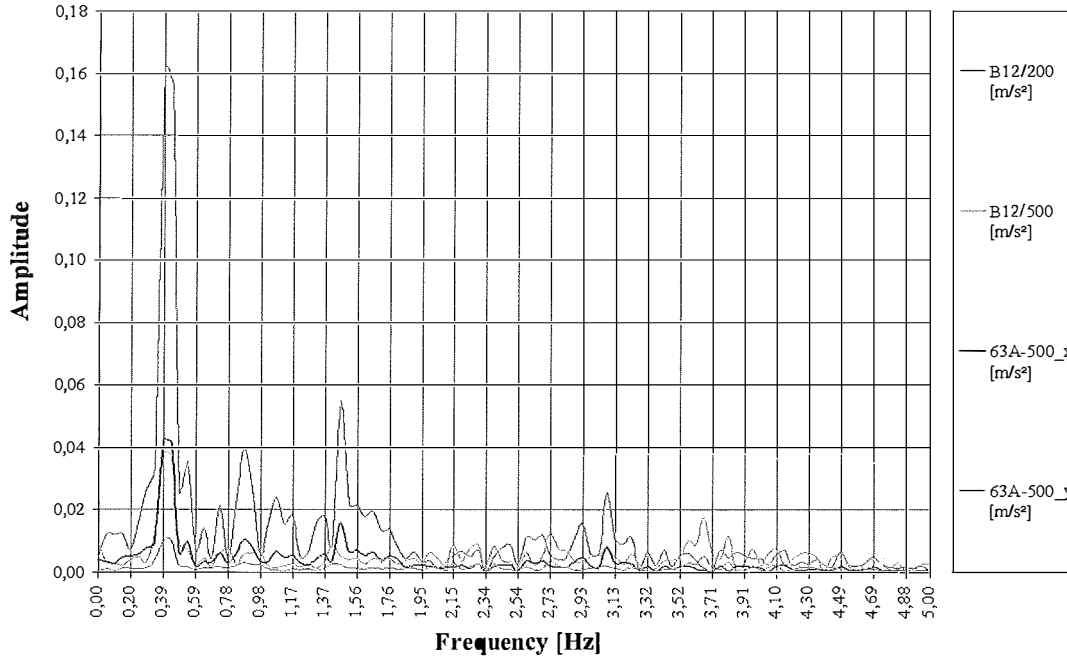


Figure 5: Amplitude spectrum of the accelerometers (startup phase)

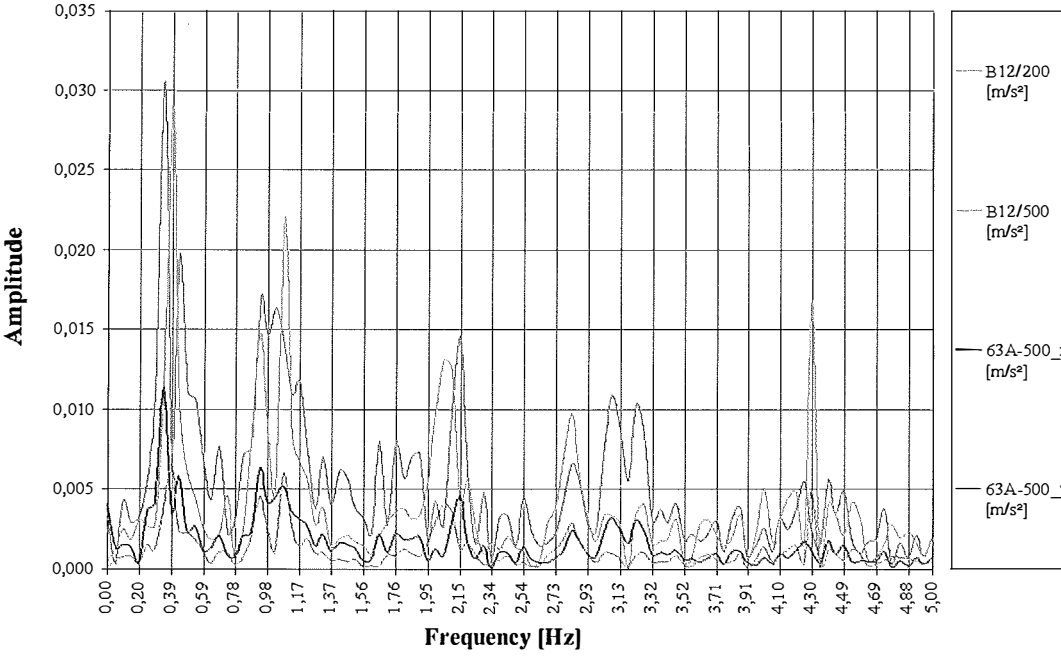


Figure 6: Amplitude spectrum of the accelerometers (phase 3)

through to fully operational service. The sampling rate is 200 Herz for all three sensors. The following figures 3, 4, 5 and 6 show the amplitude spectrums of the various service phases.

Figure 3 shows the same frequency range and time as Illustration 2. The largest amplitudes are supplied by the two acceleration-receivers B12. Generally speaking, the largest amplitudes are clearly at 0.415 Herz for all sensors. The GPS- and acceleration-measurements correlate very well concerning the resonant frequency. There are isolated higher frequencies, but with only very small amplitudes.

The turning of the gondola-axis into the wind was concluded with a very abrupt braking action. High acceleration values were recorded as a result of the large tower movements. The two sensors that were internally aligned to the y-axis measured higher acceleration values as those that were perpendicularly aligned.

Figure 4 shows that the largest amplitudes for all sensors are at 0.39 Herz and that the quickest acceleration was measured by sensor B12/500 at 0.062 m/s2.

The next diagramm shows an extract of the startup with the corresponding amplitude spectrums.

As would expected, there are more higher frequency moments recorded. Generally speaking, the largest amplitudes remain limited to the frequency range under 3 Herz. The amplitude maximum for all accelerometers is again to be found around 0.4 Herz. The quickest acceleration was measured by sensor B12/200 at 0.161 m/s2.

Several minutes pass from the beginning of the startup to fully operational service. At wind speeds around 5.5m/s, the active stall device allows the rotor to turn with approximately 14 rotations per minute.

Figure 6 shows that several clear amplitude-pikes appear across the frequency range. This is not really surprising, as the course of the measurement-signals in relation to time show fewer and fewer regularities. Further, the maxima of the accelerometers are clearly to be found under 1 Herz. The largest amplitudes for the B12/500 and 63A-500_y are at 0.39 Herz and for the B12/200 and 63A-500_x at 0.35 Herz. These amplitude values are partially considerably smaller than those recorded during startup (0.161 m/s2 against 0.031 m/s2 for the B12/200).

The vibration spectrum of the wind-generator-tower could be well recorded from standstill through to fully operational service. A general power-cut to the complete wind-generator-tower was the cause of the only data transmission failure.

6. Summary, Outlook

The problem, namely the recording and analysis of the tower vibrations as they occur by representation of the measurements of time and frequency range, that appeared to be easy at first glance, turned into a very complex one on closer inspection that also has an importance for further research.

As to the question which sensors can be used for such an assignment, it has to be said that the RTK-GPS-system can only be used within certain limitations. Obstructions and multipathing have shown clearly the limitations of this system. Especially disadvantageous was the fact that Phase 2, which was seen as being critical, could not be measured with the RTK-GPS-system. Figure 1 is an impressive example of these limitations.

The tower movements during standstill periods were very well recorded. With very few exceptions there were no obstruction problems. The long-term influences of the position of the sun could therefore be well measured. The time-frame that was available for these measurements was however insufficient. Besides, the main interest was in measuring the vibrations caused by the startup and during fully operational service.

The central controller proved to be highly suitable in all areas. Some limitations of the sensors had to be accepted as with the GPS.

The triaxiale accelerometer ISOTRON MODAL-63A-500 is a sensor that can be used in the higher frequency range (1 – 2000 Herz) and in the modal analysis. Nevertheless, it could also stand the test in the lower frequency range if one ignores the phase displacement in this frequency range. The measurement of the resonant frequency of the tower and the close correlation with the results of the other sensors verify this.

The best results came from the two B12 accelerometers. Thes highly-tuned spring-mass-systems (resonant frequency 200 and 500 Herz, attenuation constant D = 0.6, frequency ranges of 0 – 100 Herz and 0 – 250 Herz) are highly suitable in the lowest spectrum (close to 0 Herz). This was expected as the amplitude response, described as the relation of the amplitudes of the relative movements of the mass and their acceleration depending on the frequency, in the vicinity of 0 Herz is equal to 1.

References

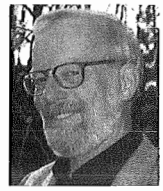
[1] HBM: Operating manual, PC measurement electronic Spider 8, Hottinger Baldwin Messtechnik (1999)
[2] HBM: Instrution manual accelerometer B12, Hottinger Baldwin Messtechnik (2000)
[3] Endeveco: Specification ISOTRON Accelerometer, Model 63A-100-500 Califomia (2001)
[4] *Krapfenbauer N., Stäussler E.,* Bautabellen-Studienausgabe, Verlag Jugend & Volk Ges.m.b.h., Wien, p.92f, p 193f (1996)

Contact

Dipl.-Ing. Dr. Johannes Fabiankowitsch: Vienna University of Technology, Institut of Geodesy and Geohysik, Departement of Applied and Engineering Geodesy, Gusshausstraße 27–29, A-1040 Vienna, Austria, email: jfabian@pop.tuwien.ac.at
Univ.-Prof. Dr.-Ing. Heribert Kahmen: Vienna University of Technology, Institut of Geodesy and Geohysik, Departement of Applied and Engineering Geodesy, Gusshausstraße 27–29, A-1040 Vienna, Austria, email: Heribert.Kahmen@pop.tuwien.ac.at
Phillip Matt: Troie 3, A-6161 Natters, Austria

Combined Evaluation of Geodetic and Geotechnical Data during Tunnel Excavation by Use of a Knowledge-Based System

Klaus Chmelina and Heribert Kahmen, Wien



Abstract

The paper presents how geotechnical knowledge can be used for the analysis of geodetic deformation data. With this contribution the application field of NATM-tunnelling (New Austrian Tunnelling Method) is addressed. The deformation analysis is based on heuristic rules which are implemented in a knowledge-based system.

Kurzfassung

In diesem Beitrag wird dargestellt, wie geotechnisches Wissen für die Analyse geodätisch hergeleiteter Deformationen genutzt werden kann. Der Beitrag behandelt eine wichtige Thematik der Neuen Österreichischen Tunnelbaumethode. Die Deformationsanalyse stützt sich auf heuristische Regeln, welche in ein Wissensbasiertes System implementiert wurden.

1. Introduction

The geodetic monitoring of 3-d displacements during tunnel excavation has become a standard procedure for tunnel projects. In NATM-projects (New Austrian Tunnelling Method, [3]), the immediate geotechnical interpretation of the periodically observed displacements must be seen as an integral part of the tunnelling method. As a consequence, qualified and experienced geotechnical experts are needed on site.

Currently, geotechnical interpretation is based mainly on the daily analysis of numerous types of sophisticated displacement diagrams together with many other listings and graphics showing further project data – a time consuming manual work. For realizing up to date analysis of large data volumes or online analysis of automatically monitored data, e.g. when robotic tacheometer systems are used, this manual procedure is obviously unpracticable. Until now, there exists no means by which these tasks are managed quickly, automatically and sufficiently reliable.

A knowledge-based approach is presented which shall assist interpretation work through an automatic analysis of the displacements by use of heuristic rules. These rules thus provide expert knowledge for the detection and interpretation of specific deformation phenomenas thus building the core of a knowledge-based system [1]. Three different examples using geotechnical knowledge to evaluate geodetic data, are presented.

2. Example ONE: How to Find an Explanation for Reactivated Deformations

In general any *unexplainable* deformation behaviour leads to an uncertainty which must be solved quickly (the causes of the deformations must be found) in order to ensure safe continuation of the project. Such a case can occur when deformations that have been consolidated in the past suddenly become reactivated. Whether such a reactivation is really *unexplainable* depends on the current situation on site. In case of NATM-tunnelling, deformations must be analysed together with the excavation progress. If e.g. an excavation face (e.g. bench, invert) approaches or passes the deformation area (situated in or above a tunnel) the reactivation of deformations usually is expected and explainable. On the other hand, any reactivation that can not be explained (e.g. by excavation activities) is alarming and perhaps critical for the construction progress. Fig. 1 characterizes the geotechnical interpretation steps which handle this problem.

Modelling these steps by conventional programming would lead to a complex code consisting of numerous interlocked IF - THEN statements, logical expressions (AND, OR, NOT) and loop commands. Therefore a rule-based system

was chosen here. With this system the interpretation process can be modeled in a modular, transparent and much more flexible way. With respect to a better understanding of rule based inferencing the use of WHENEVER instead of IF is more appropriate. The decisive difference and advantage is that by use of rules the interpretation must not necessarily be processed top-down. The developed rule-based system can ask all the questions shown in fig. 1 dynamically at the same time but does not have to ask them sequentially (as with IF-THEN). Further, the questions can remain permanently active in memory but do not have to be repeatedly asked (as with loops). To elucidate let's transfer the problem into heuristic rules:

Rule 1: WHENEVER deformation data match the pattern "Reactivation"
THEN produce a fact that marks this data as "reactivated"

The precondition of the rule consists of a representation of an abstract deformation pattern called „Reactivation". At implementation level this pattern is defined in a more or less empirical way as a time series of deformation data showing increasing deformation velocities during the last 3, 4 or 5 epochs. The action of the rule is to add (assert) a meaningful information (a fact) to the working memory describing which point, which epoch, etc. is concerned. This is completed for all such incidents thus the knowledge-based system becomes aware and mindful of them.

Rule 2: WHENEVER 1. there is a fact that marks data as "reactivated"
2. excavation data match the pattern "Close Enough"
THEN produce a fact that describes this link

The rule consists of two preconditions to be regarded as connected by the logical AND-statement. The first is fulfilled if there exists at least one of the previously mentioned facts pro-

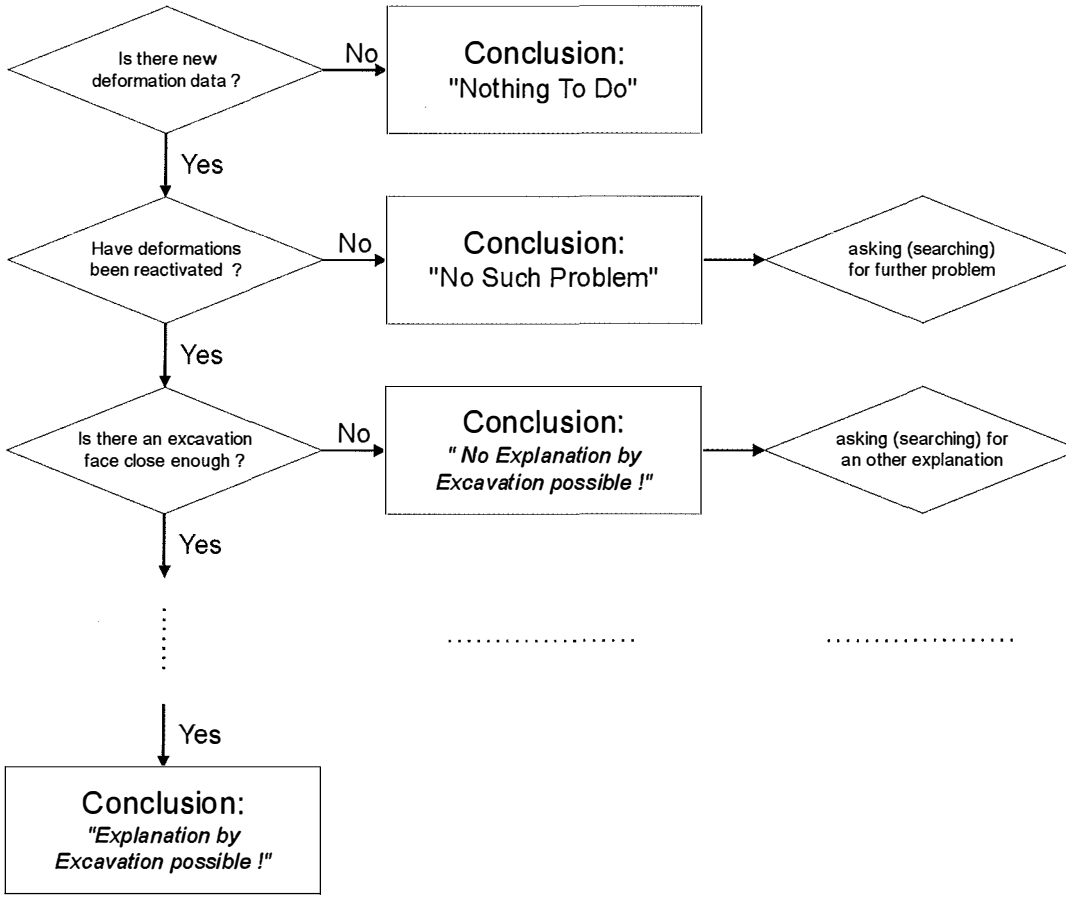


Fig. 1: Interpretation steps

duced by rule 1. The second consists of another abstract pattern *"Close Enough"* specifying that excavation data prove excavation activities to be a possible cause for a detected reactivation incident. For that certain distances in space as well as time constraints must be considered which are defined according to assumptions of the expert. If both conditions are fulfilled the rule adds a meaningful fact to the working memory expressing that there is (possibly) a connection between reactivation and excavation. The rule fires for each excavation activity meeting "close enough" and links it to the reactivation incident (e.g. a reactivation can be explained by more than one excavation activity). As can be seen, linking is performed whenever new excavation data or new deformation data is accessible. So rule 1 and rule 2 together keep track of both data.

For continuing interpretation (fig. 1) rule 3 could be specified by simply modifying the second precondition of rule 2 now assuming the opposite case that there is no excavation data matching the pattern "close enough". Thus the corresponding "No Explanation" – conclusion could be processed and stored. This somehow would mean that the normally large volume of excavation data has to be checked twice for each reactivation incident. To speed up runtime performance rule 3 is alternatively designed as follows:

Rule 3: WHENEVER 1. there is a fact that marks data as "reactivated"
2. there is no fact describing a link to excavation data

THEN conclude (display and store) that there is no explanation

Adopting this method, the system must know that it cannot begin applying rule 3 before rule 2 has been used which serves as an example that asking all questions at the same time is not always efficient. This meta-knowledge was implemented by an appropriate hierarchical structuring of rules. The majority of rules were kept at equal level.

By additional rules the knowledge-based system is able to continue forward-reasoning down to a theoretical final conclusion:

"The phenomena could have been caused by the excavation,"

which in case might be reached or not. At the same time further (either independent or depending) interpretational problems can be handled by other rules allowing for similar *horizontal* and *vertical* conclusions. If the user is not really interested in all problems he/she can preselect the desired ones in a configuration dialog. The selection is then transferred into facts triggering only the relevant heuristic rules.

The system can be called heuristic as the arrangement and composition of the rules was mainly driven by common sense, efficiency aspects and some good advices of experts. System tests proved that the software can process even large data volumes within acceptable times. What has to be observed more critically is the error budget. Wrong conclusions might be drawn especially when there are inaccuracies, errors and/or missing data. This problem was realized at an early stage and as a consequence, a knowledge-based error detection method also based on heuristic rules was implemented (see example 2 below).

With regard to the problem of wrong conclusions it is emphasized that the original intention was to develop a support system where the human expert still has the last comment. All final conclusions of the system are therefore reported in the form of insecure statements ("could").

3. Example TWO: How to Find and Analyse an Error in Deformation Data

As mentioned, the knowledge-based system must reason over actual measuring results. In order to avoid wrong conclusions due to bad quality of input data, error detection is needed in advance. (In practice the geotechnical expert does not consider error detection as a separate interpretational task – in other words: "It's not their job to find errors".) So as long as data quality problems are seldom he/she apriori trusts in the data provided. He/She is however able to recognize (possibly) incorrect data by specific characteristics or signs mostly driven by experience and common sense. Further, to become sure, he/she begins to investigate the problem in detail. A typical error investigation strategy is briefly described as follows:

1. Discovering of a characteristic error sign (e.g. an unexpected deformation) more or less by accident.
2. Determining of its relevance for the interpretation If the sign is rather small it might be considered irrelevant and neglected by the expert.
   If the sign is strong it still might be irrelevant

(e.g. if the corresponding data has no influence on the interpretation at all).

3. Deriving of a hypothesis (of a possible error cause, error type) from a known correlation between the sign and a cause.
   If various causes could be responsible the best one is chosen first (the most obvious, frequent or probable, the easiest to be checked, etc.). Now the sign becomes a symptom.
4. Searching for further symptoms supporting the hypothesis
   The search is driven by the distinct intention to prove the hypothesis.
5. Searching for signs, facts, hints, features, etc. contradicting the hypothesis
   Now the search is driven by the opposite intention to disprove the hypothesis. Which kind of search is started first depends on efficiency criterias. Investigations show that for psychological reasons experts usually try to verify (especially their own) assumptions first even if their falsification would mean less effort.
6. Evaluating of all gathered aspects for deriving a verification certainty.
   The question is: "Can the hypothesis be established as a diagnose ?"
   We may decide to stop and establish the hypothesis as a diagnose or to go back to point 3 deriving and evaluating an alternative hypothesis. If none of these can finally be verified sufficiently (better: satisfactory) either the initial error assumption is cancelled or the investigation result remains a list of possible but uncertain causes.

A knowledge-based system for handling this complex strategy needs an appropriate way to represent and propagate uncertainty – a problem which leads to the various concepts of probabilistic reasoning. Among those, the Certainty Factor Method [4] was chosen because it represents a pure heuristic approach especially developed for rule-based expert systems.

Certainty Factors – a brief introduction:

Each premise (condition) P_1, P_2,...,P_n of a rule (e.g. the statement: "there is an error sign") is attributed with a numeric parameter (a certainty factor) cf_1, cf_2,...,cf_n describing its degree of membership to the concept true. The parameter interval lasts from –1 (not true) to +1 (true). A certainty factor $cf_x = 0$ means that there is nothing known about the truth of P_x. Also, each rule R_1, R_2,...,R_n itself (e.g. the statement: "from the error sign can be concluded an error cause") has a certainty factor cfr_1, cfr_2,...,cfr_n expressing the strength (the truth) of the relation between the premises and the conclusion. When a rule R_x is executed, all involved certainty factors cf_1, cf_2,...,cf_n of its n premises and cfr_x are processed to obtain a certainty factor cfc_x for the conclusion C_x. In rule based expert systems a common way is to calculate:

$$cfc_x = \min(cf_1, cf_2, \ldots, cf_n) * cfr_x. \quad (1)$$

If by two different rules R_1, R_2 the same conclusion is drawn the now concurring different values of cfc_1 and cfc_2 must consequently be processed to a final cfc which can be completed as follows:

$$cfc = \begin{cases} cfc_1 + cfc_2 - cfc_1 * cfc_2 & \text{if : } cfc_1, cfc_2 \geq 0 \\ cfc_1 + cfc_2 + cfc_1 * cfc_2 & \text{if : } cfc_1, cfc_2 \leq 0 \\ \dfrac{cfc_1 + cfc_2}{1 - \min(|cfc_1|, cfc_2|)} & \text{if : } cfc_1 * cfc_2 < 0) \end{cases} \quad (2)$$

If further rules lead to the same conclusion the algorithm is applied sequentially. The mathematical operations for processing the certainty factors are thus chosen intuitively. They heuristically approximate the human way of propagating uncertainty.

An example is now given to imitate the above interpretation strategy by applying the Certainty Factor Method. It shall be analysed if there is a systematic height error of the instrument station in the measured vertical deformations. At first rule 4 is designed for detecting an abstract pattern *"Strong Settlement"*. Its conclusion (fact) is considered to be an initial and relevant error sign for a specific cause (steps 1, 2 and 3 of the strategy).

Rule 4: WHENEVER
1. deformation data match the pattern "Strong Settlement"
2. there is no fact so farthat marks the related epoch and station as suspicious

THEN produce a fact that marks the related epoch and station as suspicious

The rule fires exactly one time for all measuring epochs and instrument stations having at least one measurement showing the pattern. The system now knows all suspicious epochs and stations that have to be further investigated. Now (and only for them) there is applied the Certainty Factor Method. The following initial certainty factors for the following three statements (rules) are defined as:

$cfr_1 = 0.3\ e^{-(bew\_diff)^2}$: *From one strong settlement incident there can be concluded the error.*

If there is only one measurement available (which shows the sign) the error is concluded with a certainty of 0.3 thus staying rather uncertain. If there are more measurements showing the sign the certainty of concluding the error from one such sign becomes a function of the similarity of all signs. The function models the expectation that a height error of the station should be visible in all measurements in the same size. The less that one sign meets this expectation the less there can be concluded the error from it. This aspect is implemented by multiplying 0.3 with an exponential term considering the difference between the size of the sign to the mean size of all signs. The rule models step 4 of the strategy producing a list of positive certainty factors.

$cfr_2 = -0.1$: *From an inconspicuous measurement there can be concluded the error.*

If a measurement from the suspicious station appears as normal it is treated as if it would slightly contradict the hypothesis with a negative certainty of –0.1.

$cfr_3 = -0.5$: *From a measurement showing a strong heave there can be concluded the error.*

If a measurement shows a strong heave which indicates an opposite than initially expected direction of deformation, the case is regarded as a strong irritation. As a consequence, it contradicts the hypothesis with a high negative certainty of –0.5. The last two rules model step 5 of the strategy producing a list of negative certainty factors.

Finally, all certainty factors were processed by eq. 2 giving a final certainty of the hypothesis. Fig. 2 shows an example of the vertical deformations of 7 points measured at Oct. 27 for which the system states: "A systematic height error of the station could have happened. For this kind of error a certainty of 0.87 was derived." By studying fig. 2 the reader is invited to compare this result with his/her own assumptions.

By applying this method, the knowledge-based system is now able to offer an additional explanation for the interpretational problem discussed in chapter 2 (fig. 1) which says:

"The phenomena could be caused by an error."

4. Example THREE: How to Find Remarkable Deformations by use of Rating Knowledge

During tunnel design deformations are prognosed by use of diverse techniques (e.g. numerical simulations) providing the knowledge of how the planned tunnel (the rock) is expected to deform under certain conditions. Besides these official prognoses, the expert also has his/her own individual and much more extended expectations (usually derived from experience). During tunnel excavation the measured deformations are compared with the prognosed and expected ones. Unexplainable big differences between

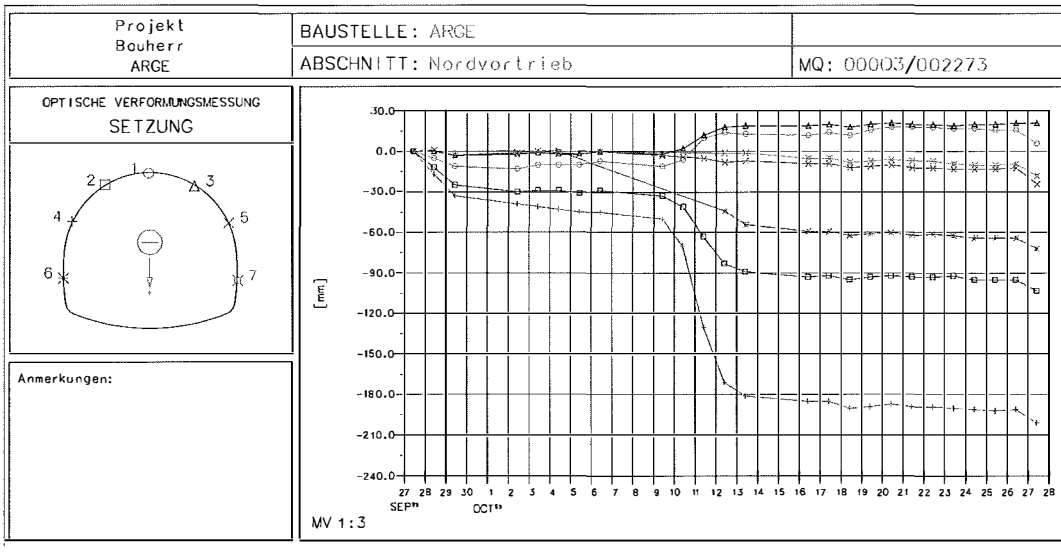


Fig. 2: Measured vertical deformations of 7 points

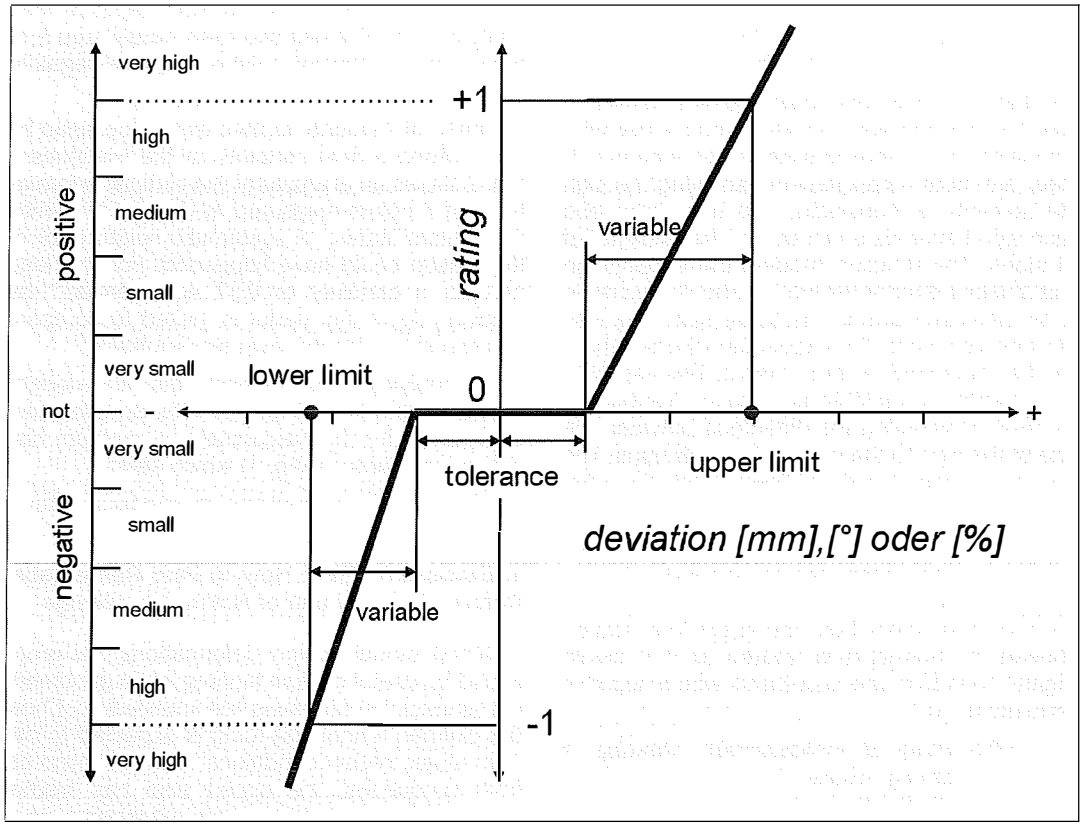


Fig. 3: Empirical rating model

them can be an alarming sign for the geotechnical expert on site. An automatic detection of such a situation makes necessary the representation of the prognoses and expectations as well as the expert knowledge of how to rate their deviation(s) from reality. In the developed knowledge-based system, prognoses and expectations can be specified by defining deformation curves, vectors, values, etc. and assigning them to distinct measuring objects (points, sections, areas). The rating knowledge is represented by empirical rating models (fig. 3) which can be selected, configured and assigned to a prognose or expectation.

The rating model allows us to transfer the difference between a real and prognosed (expected) measuring result (e.g. the difference in settlement [mm], the difference in vector orientation [°], the difference in deformation velocity [mm/d], etc.) into a numeric rating value and an associated linguistic variable. The transfer function might either be linear (as exemplified in fig. 3), exponential, logarithmic or determined by any other mathematic equation that best serves to approximate the users way (wish) of rating. In addition, a rating tolerance can be specified for maintaining the rating value equal to zero in case of small differences and an empirical measuring accuracy for putting the rating value to a worst-case value by simply adding the measuring accuracy to the difference before rating.

As soon as we input the knowledge into the system heuristic rules apply it to the incoming deformation data. The rules ensure that all deviations which exceed user-definable evidence limits (e.g. all deviations reaching the *very high* – level) are detected and reported automatically. Rule 5 below shows the principle structure:

Rule 5: WHENEVER
1. there is deformation data
2. there is a prognose for this data
3. there is a rating model for this prognose
4. the calculated rating value exceeds an evidence limit

THEN produce a fact

As can be seen, more than one prognose may exist (e.g. from different numerical simulations) as well as more than one rating model (e.g. from different experts). However the rule evaluates all of them thus becoming aware of all evident cases. Alternatively, by a simple modification of the rule the knowledge-based system becomes able to offer another answer to the interpretational problem discussed in chapter 2 which says:

"The phenomena was prognosed, maybe there is no problem at all."

In order to provide more information about this kind of answer the system additionally outputs context information to the prognose (prognosed by whom, when, based on what, etc.) which has to be input in advance (together with the prognose).

5. Conclusion

It is shown how interpretation of deformations, occurring during the construction of a tunnel, can be based on a knowledge-based system. The system is thought to support the geotechnical expert on site by reducing his/her interpretation time and effort which is necessary in case deformation monitoring is done frequently with automatic monitoring systems. Classical deformation analysis methods, only based on geometrical data, can not be used here, as in addition, information such as geological data, data about the excavation progress, data from simulation calculations, etc. have to be included in the evaluation process. This paper shows with three examples how heuristic rules can be developed and used in such a knowledge-based system. With the developed rules, deformations reactivated by the construction process, measurement errors and differences between prognosed and measured deformation data can be analysed.

References

[1] *CHMELINA, K.:* Wissensbasierte Analyse von Verschiebungsdaten im Tunnelbau, Dissertation, Inst. of Geodesy and Geophysics, TU-Vienna, 2002.

[2] *BONISONE, P., DECKER, K.:* Selecting uncertainty calculi and granularity: an experiment in trading-off precision and complexity. In: Proc. of Workshop on Uncertainty and Probability in Artificial Intelligence, S. 57–66, Los Angeles, CA, 1985.

[3] *MÜLLER, L., FECKER, E.:* Grundgedanken und Grundsätze der ‚Neuen Österreichischen Tunnelbauweise'. Clausthal, Trans Tech Publications, 1978

[4] *SHAPIRO, St.C.:* Encyclopedia of Artificial Intelligence. John Wiley & Sons, Inc. Publishers - New York, 2nd edition, 1992.

Contact

Dipl.-Ing. Dr. Klaus Chmelina: Geodata ZT GmbH Wien, Geyschlägergasse 14, A-1150 Vienna, Austria, email: chmelina@geodata.at

Univ.-Prof. Dr. Heribert Kahmen: Institute of Geodesy und Geophysics, TU-Vienna, Gusshausstrasse 27–29, A-1040 Vienna, Austria, email: hkahmen@pop.tuwien.ac.at

A Fuzzy System for the Assessment of Landslide Monitoring Data

Michaela Haberler, Wien

Abstract

In many cases landslide areas can be divided into several blocks, which are moving with different velocities in different directions. So, in case we are able to detect block boundaries, landslide monitoring can be performed more efficiently. The information about the relative movement of these blocks is a very important indicator for future movement behavior, if monitored very precisely at the block boundaries with continuously measuring geotechnical sensors.

To detect the boundaries of the blocks the following algorithm is used: the displacement vectors of the observed points (out of a geodetic deformation analysis) will be analysed by an affine coordinate transformation. The assignment of the observed points to the different blocks is done by an iterative algorithm; the thresholds for the several steps of the algorithm are calculated by a fuzzy system. The input parameters for this fuzzy system are e.g. the residuals of the transformations and strain parameters calculated from these transformation steps.

Finally, an example for application of this fuzzy system will be given.

Zusammenfassung

Meist können Hangrutschungsgebiete in einzelne Bereiche mit verschiedenen Bewegungsrichtungen und -geschwindigkeiten eingeteilt werden. Wenn mit Hilfe geodätischer Überwachungsmessungen diese einzelnen Teilblöcke des Rutschhanges ermittelt werden, kann über diese Blockgrenzen hinweg mit lokal messenden geotechnischen Sensoren permanent und hochgenau die Relativbewegung der Blöcke zueinander registriert werden. Daraus können wichtige Informationen über das zukünftige Bewegungsverhalten des gesamten Bereiches gewonnen werden.

Ein Teilbereich dieser Aufgabenstellung ist die Detektion der Grenzen zwischen den einzelnen Blöcken des Rutschhanges. Die Idee ist, dass geodätisch überwachte Punkte, die gemeinsam auf einem dieser Teilbereiche liegen, ähnliche Bewegungen ausführen. Mittels einer überbestimmten Affintransformation werden aus den Verschiebungsvektoren sowohl Starrkörperbewegungen (Translation und Rotation) als auch die Verzerrungen der verschiedenen möglichen Teilbereiche ermittelt. Anhand von Kenngrößen der Transformationen (z.B. Residuen, Strainparameter) können in einem Fuzzy System die einzelnen Teilblöcke des Rutschhanges bestimmt werden.

Anhand eines Beispiels werden die Komponenten des Fuzzy Systems vorgestellt.

1. Introduction

Landslides are one of the major types of natural hazards worldwide. Every year thousands of people are injured or even killed. Additionally, many buildings and infrastructure like railroads and traffic networks are destroyed. E.g., only in the US, 25 to 50 people are killed per year and the economic damage reaches US$ 2 billion. So, there is an urgent need for a suitable monitoring and alarming system. The project OASYS (Integrated Optimisation of Landslide Alert Systems) was started to fulfil these requirements by a multi-disciplinary approach [1].

OASYS consists of several steps:

1. At a regional scale, potential landslide areas have to be identified. This is done by satellite images (remote sensing data, INSAR), topographic and geological maps, historical reports,...
2. At a local scale, for each of these landslide areas a monitoring system is installed. First, a geodetic network is observed in several epochs and based on these measurements the block boundaries can be identified.
3. In a next step, high precision geotechnical sensors are installed across the block boundaries to permanently observe the relative movement of the blocks.
4. A knowledge-based system is used to analyse the collected data and to support the decision making authorities in case of danger.

This paper deals with one part of the project OASYS, the detection of the block boundaries based on geodetic deformation measurements.

2. Algorithm for block detection

We assume that the geodetic networks have been measured periodically and the coordinates and displacement vectors were calculated out of a geodetic deformation analysis. The assumption for our method is that displacement vectors for all observed points are available. The block detection algorithm is the necessary step to proceed from the measurements of the geodetic network to the following installation of the high precision geotechnical sensors in the area under investigation.

The block detection algorithm (see fig. 1) starts with the identification of a minimal block of four points (explanation see section 2.1); in the following iterative procedure other neighbouring points are added to the block until it is 'complete'. This assessment of completeness is done by a fuzzy system. Then the algorithm starts again, identifying another minimal block out of the remaining observation points. The practical implementation of the algorithm is described in section 4.

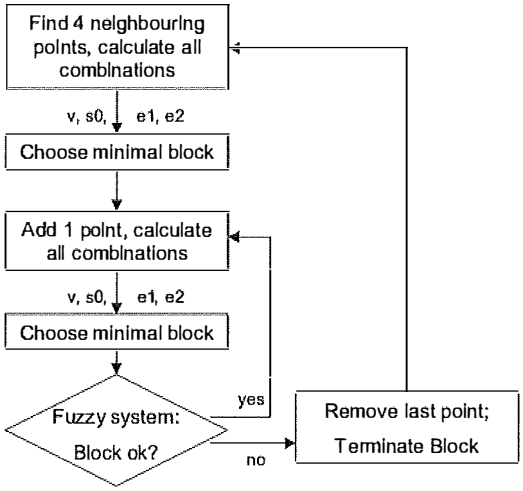


Fig. 1.: A coarse scheme of the analysis algorithm implemented in MATLAB.

2.1. Affine coordinate transformation

The block detection algorithm is based on the displacement vectors of the observed points. The idea is to use an over-determined affine coordinate transformation to assess the movement of the points between two subsequent epochs of measurements. This means that the coordinates of the points of epoch n are mapped onto the coordinates of the same points of epoch n+1. If some points move in the same direction (assuming that they are lying on one common block) then an over-determined affine transformation will give small residuals and a small standard deviation. In case points of different blocks were considered simultaneously the standard deviation will be significantly larger.

In two dimensions, the approach reads as follows:

$$\begin{aligned} y_{n+1} &= a \cdot y_n + b \cdot x_n + c \\ x_{n+1} &= d \cdot y_n + e \cdot x_n + f \end{aligned} \quad (1)$$

where y_n, x_n...coordinates of epoch n
y_{n+1}, x_{n+1}...coordinates of epoch n+1
a,..,f...transformation parameters

To solve this equation system, three identical points are necessary; for an over-determined solution at least 4 points have to be used. So the algorithm starts with a minimum block size of 4 points.

The six transformation parameters (a,..,f) can be interpreted as two translations (t_y, t_x), two rotations (w_y, w_x) and two scale parameters (m_y, m_x):

$$\begin{aligned} a &= m_y \cdot \cos w_y & b &= m_x \cdot \sin w_x & c &= t_y \\ d &= -m_y \cdot \sin w_y & e &= m_x \cdot \cos w_x & f &= t_x \end{aligned} \quad (2)$$

The six-parameter approach was chosen because of the special properties of landslides: The movement of a block can be described by a translational part and a rotation/distortion [2]. In most cases the distortion in the direction of the movement is larger than in other directions. So a second scale parameter is necessary to fully describe the anisotropic strain conditions.

2.2. Strain analysis

Due to the small displacements, the six transformation parameters usually are not meaningful (i.e. a clear distinction of the several blocks is not possible out of these parameters). On the other hand the affine transformation is analogous to a strain analysis assuming homogeneous and infinitesimal strain (e.g. [3], [4]). So there is a direct relation between the transformation parameters and the infinitesimal strain components e_{xx}, e_{yy} (rate of change of length per unit length in direction of x-axis resp. y-axis), e_{xy} (= e_{yx}, rate of shear strain) and the derived rotation angle ω. The strain ellipse represented by the semi-axes e_1, e_2 and the orientation θ of the maximum strain rate is calculated from these strain components analogous to the geodetic point error ellipse.

In contrast to Welsch [3], who recommends not to use translation parameters, here an integration of the translation parameters (t_y, t_x) seems useful to fully describe the movement pattern of a block

to avoid that the strain parameters are distorted due to translational movement.

2.3. Indicators for the block detection algorithm

The indicators used for the analysis in the fuzzy system can be determined by the results of the sequence of affine transformations. Some of the indicators are presented here:

- The standard deviation s0 is used for a first evaluation of the block properties. Investigations have shown that especially the change of the standard deviation from one step to the next one is a very important indicator for the assessment if the block is still 'correct'.
- The next parameters for the fuzzy system are the two semi-axes of the strain ellipse (TISSOT indicatrix): e_1, e_2. In case a point does not belong to the block investigated, e_1 and e_2 become significantly large. Here, the absolute values on the one hand and the change of the parameters between two subsequent steps of iteration on the other hand are used.
- Investigations have shown that the variation of the residuals is a good indicator to distinguish between a correct block (all points belong to one block) and an incorrect block (point of a neighbouring block is included). The interquartile range (used in the exploratory data analysis) gives an estimate of the variation of the residuals in every step of the transformation (see fig. 2). The interquartile range was analysed for 170 cases consisting of 4 points. The interquartile range of the 31 correct blocks (= 4:0) is significantly smaller than for the incorrect blocks (81, 58 cases resp.). '4:0' in fig. 2 denotes that all 4 points are lying on one block, '3:1': 3 points on one block, 1 point on the neighbouring block, '2:2': 2 points per block.

3. Fuzzy system

For the implementation of the algorithm MATLAB was used. It provides an initial fuzzy sys-

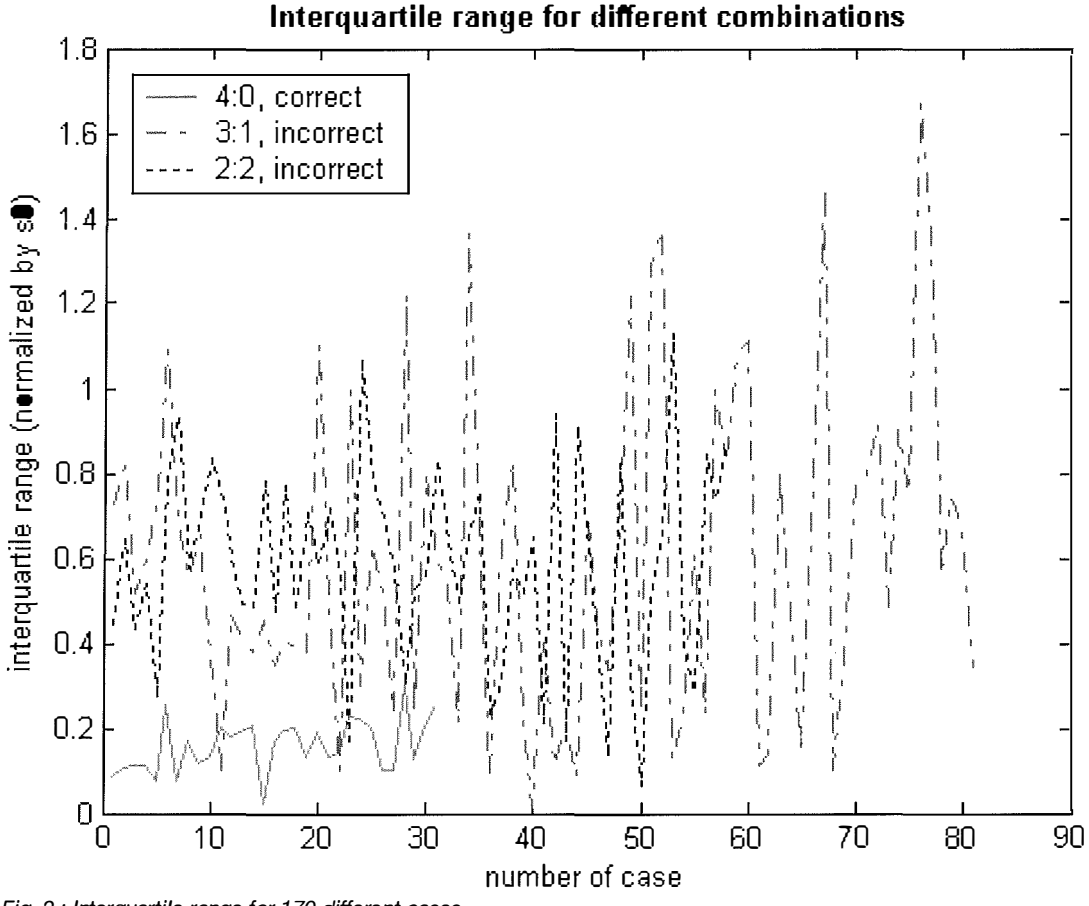


Fig. 2.: Interquartile range for 170 different cases.

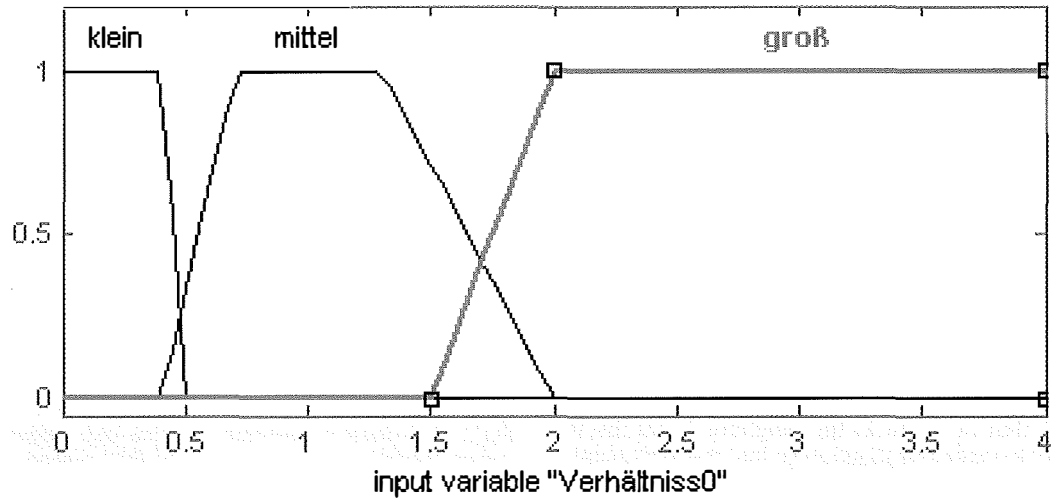


Fig. 3.: An example for the membership functions of an input parameter.

tem, which supports all necessary calculation methods (membership functions, methods for fuzzyfication, inference system, defuzzification). The developper of the fuzzy system has to choose the suitable methods, input and output parameters for his application.

3.1. Input variables and membership functions

The input parameters of the fuzzy system were already mentioned before, e.g. the change of the standard deviation of subsequent steps, strain ellipse parameters e_1 and e_2,... To use the input values in the fuzzy system, the 'sharp' values have to be fuzzified. Therefore, membership functions are used. MATLAB provides many membership functions, such as piecewise linear functions, the Gaussian distribution function,...

For most of the input parameters the membership function type 'trapezoidal' was chosen in a rather intuitive way. The inputs are split in 3 resp. 5 membership functions (per input). An example for the input parameter 'change of standard deviation s_0' is given in fig. 3. It is represented by three membership functions of type 'trapezoidal': small, medium and big.

3.2. Rules

The rules used in the inference system have been found empirically. At the moment, 25 simple rules have been implemented. But this topic is still under investigation.

The rules implemented must have a structure like: If (input is X) then (output is Y), followed by

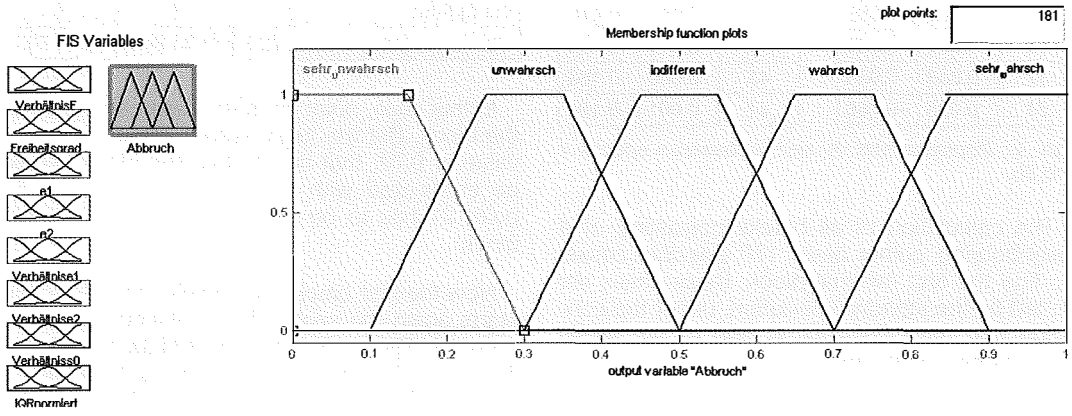


Fig. 4.: The output value (probability for termination) is represented by 5 membership functions: very unlikely, unlikely, indifferent, likely, very likely.

a weight for this rule, e.g. (1). Examples for some rules used in the inference system are:

If (e_1 is klein) then (Abbruch is unwahrscheinlich) (0.5)
If (e_1 is mittel) then (Abbruch is indifferent) (0.5)
If (e_1 is groß) then (Abbruch is wahrscheinlich) (0.5)

3.3. Output

After the evaluation of the actual values of the input parameters, a fuzzy set for each output variable exists that needs defuzzification to get a single, 'sharp' output value.

Here only one output is implemented. It represents the probability that the block is complete, i.e. that no point in the neighbourhood fits the block under investigation so that the search algorithm has to be stopped. The output value is represented by 5 membership functions, see fig. 4.

4. Example

In this section the algorithm is tested with the so-called Delft network. This simulation of a deformation network consists of several epochs with different movement patterns. The simulated measurements (distances and directions) can be found in [5]. In this testing scenario, epochs 1 and 3b were used to calculate a geodetic deformation analysis using the software PANDA [6]. The displacement vectors, which are the results of the deformation analysis, are the input for the block detection algorithm (see fig. 5 and tab. 1).

Point	dy [m]	dx [m]
3	0.232	0.158
5	0.197	0.180
11	−0.021	0.339
13	−0.072	−0.069
15	−0.087	0.032
17	−0.058	0.033
21	0.025	−0.166

Point	dy [m]	dx [m]
35	−0.054	−0.121
37	−0.035	0.004
39	0.113	0.248
41	0.042	0.270
43	−0.018	−0.181
45	0.002	−0.186
47	0.001	−0.124

Tab. 1.: Displacements dy and dx between epochs 1 and 3b.

The calculation of all blocks consisting of 4 neighbouring points gives 170 possible cases. To find the first minimal block, these cases are sorted by the standard deviation, under the condition that the values for e_1 and e_2 are within user-chosen limits (necessary to consider the actual material properties).

So the first block determined by the algorithm consists of the points 3, 5, 11, 41 with the standard deviation s_0 = 9.9 mm. In the next step all the neighbouring points are used to find the next point of the block:

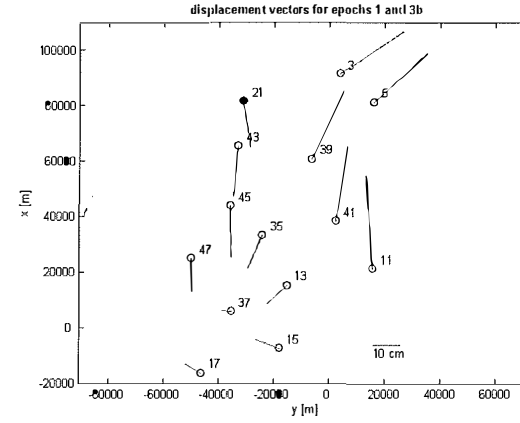


Fig. 5.: Graphical representation of the displacement vectors for the epochs 1 and 3b of the Delft network.

Possible candidates	13	15	21	35	**39**	43
std dev s_0 [mm]	78.3	59.5	52.6	66.2	**12.7**	55.8

The algorithm decides that the best candidate for a fifth point at the block would be point 39 because of the minimal standard deviation s_0 = 12.7 mm. The fuzzy system now calculates all the necessary input parameters and after evaluation of the implemented rules it concludes with the following output (fig. 6).

The defuzzified output value is 0.30, that means a probability of 30 % that the block should be terminated. So the next step of iteration is started to find the sixth point of the block:

Possible candidates	13	15	**21**	35	43	45
std dev s_0 [mm]	94.4	72.3	**71.3**	86.9	76.4	76.7

In this case, point 21 is the best candidate for the block. But the fuzzy system gives an output value of 0.72, that means that the block should be terminated with a probability of 72 %. So point 21 has to be removed from the selected block.

Now the first block is complete. The algorithm starts again the search for a minimum block of 4 points. The combination of points 13, 15, 17, 35 with a standard deviation s_0 = 12.0 mm is chosen. The results of the next steps can be found in tab. 2. After 6 iterations the algorithm stops because all points have been used.

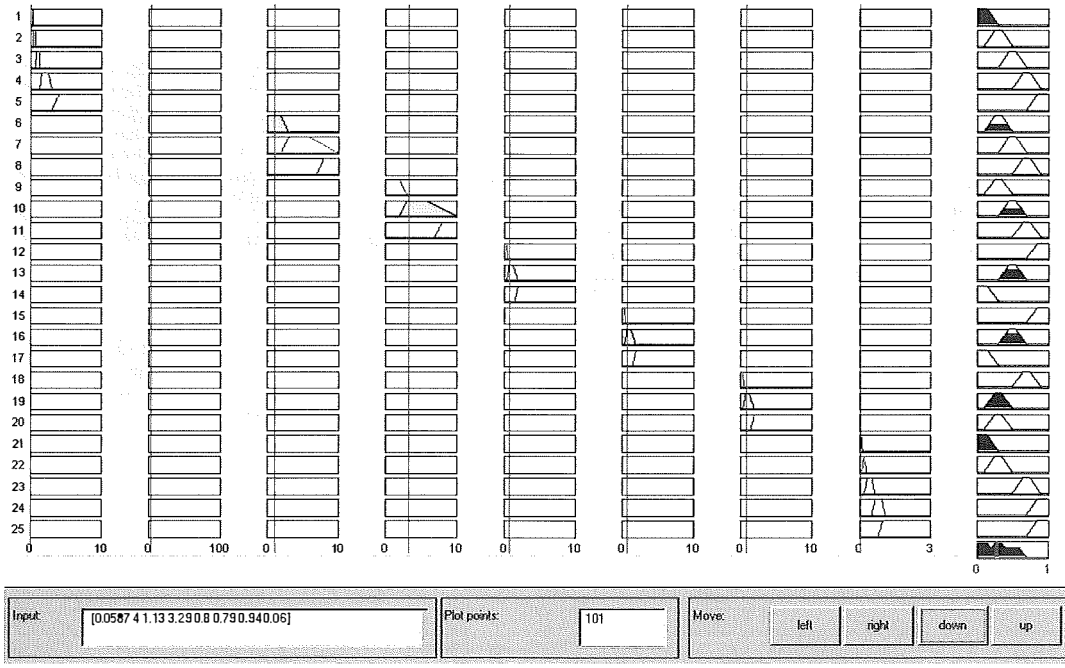


Fig. 6.: Screen-shot of the aggregation and defuzzyification process for the first run of the fuzzy system; output value: 0.30.

As a result, two blocks have been identified. This result corresponds to the pattern that people would intuitively find after viewing the graph of the displacement vectors (fig. 5 above).

The different block detection algorithms presented in [5] produce the same blocks as well.

5. Conclusion

In the past years, fuzzy methods were used more and more in geodesy (e.g. [7]). Fuzzy systems are a good method for the assessment of imprecise data or for the processing of linguistic variables because it is possible to reproduce the human way of decision making.

For the movement of a sliding area, it is hardly possible to build a suitable mathematical model fully describing all of the complex processes in the background (geology, hydrology,...). So this situation is well suited to be treated by fuzzy methods, where the knowledge of different sciences can be combined in a non-formal way. Here a possible assessment algorithm for landslide monitoring data is developed. The example of the Delft network shows that the problem given can be solved using a fuzzy system. But there is much more work to do; it is planned to include other input parameters like e.g. geological information (type of material of the area investigated). So the number of input parameters and rules will increase. Additionally, the fuzzy system will be tested on further examples.

Iteration number	Points included in the actual block									s_0 [mm]	Probability of termination
1	13	15	17	35						12.0	
2	13	15	17	35	47					11.9	0.30
3	13	15	17	35	47	45				11.0	0.30
4	13	15	17	35	47	45	37			17.1	0.39
5	13	15	17	35	47	45	37	43		24.3	0.48
6	13	15	17	35	47	45	37	43	21	31.6	0.50

Tab. 2.: Results of the subsequent steps of the algorithm. Point 21 is the last point, so the algorithm is terminated (neglecting the output value of the fuzzy system).

Acknowledgement

This work was partially supported by the European Commission, Research DG, Environment Programme, Global Change and Natural Disasters.

References

[1] *H. Kahmen, W. Niemeier:* OASYS: Integrated Optimisation of Landslide Alert Systems. This issue of VGI.

[2] *A. Antonopoulos, W. Niemeier:* Formulierung und Test impliziter linearer Hypothesen bei der geodätischen Deformationsanalyse. In: W. Welsch (Hrsg.): Deformationsanalysen '83. Beiträge zum Geodätischen Seminar 22. April 1983. Schriftenreihe Wissenschaftlicher Studiengang Vermessungswesen Hochschule der Bundeswehr München, Heft 9. München, 1983.

[3] *W. Welsch:* Description of homogeneous horizontal strains and some remarks to their analysis. In: Proceedings of the International Symposium on Geodetic Networks and Computations. DGK Reihe B, Heft Nr. 258/V. München, 1982.

[4] *F. K. Brunner:* On the analysis of geodetic networks for the determination of the incremental strain tensor. Survey Review Vol 25 , No. 192 (1979).

[5] *W. Welsch (Hrsg.):* Deformationsanalysen '83. Beiträge zum Geodätischen Seminar 22. April 1983. Schriftenreihe Wissenschaftlicher Studiengang Vermessungswesen Hochschule der Bundeswehr München, Heft 9. München, 1983.

[6] Handbuch zum Programm Panda (Programm zur Ausgleichung von geodätischen Netzen und zur Deformationsanalyse). GeoTec GmbH, Laatzen, 2002.

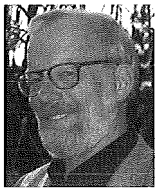
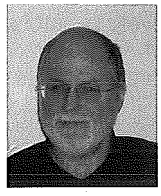
[7] *A. Wieser:* Robust and fuzzy techniques for parameter estimation and quality assessment in GPS. Dissertation, TU Graz. Shaker Verlag, Aachen, 2002.

Contact

Dipl.-Ing. Michaela Haberler: Institute of Geodesy and Geophysics, Department of Applied and Engineering Geodesy Vienna University of Technology, Gusshausstr. 27–29 /E1283, A-1040 Vienna.
email: mhaberle@pop.tuwien.ac.at

OASYS-Integrated Optimization of Landslide Alert Systems

Heribert Kahmen, Wien and Wolfgang Niemeier, Braunschweig

Abstract

In the past there has already been done a wide range of research work on landslides. Most of this research, however, had a bias towards one discipline, like GPS or geology. Therefore an international and interdisciplinary project was started. The proponents of the project, sponsored by the European Union, believe that a multidisciplinary integration of different methods has the greates potential for substantial progress in natural hazards management. The goal of the project is the development of methods that allow:
1. to detect potential landslides
2. an efficient and continuous observation of critical areas
3. the devivation of real time information about actual risks.

Kurzfassung

In der Vergangenheit hat es bereits viele Projekte gegeben, Hangrutschungen zu beobachten. In den meisten Fällen war dies jedoch mit dem Nachteil verknüpft, dass nur eine Disziplin beteiligt war, wie die Geodäsie mit GPS oder die Geologie. Es wurde daher ein internationales Projekt organisiert. Die Teilnehmer des Projektes, welches von der Europäischen Union finanziert wird, sind überzeugt, dass eine multidisziplinäre Integration verschiedener Methoden den größten Fortschritt für das Katastrophenmanagement bringt. Ziel des Projektes ist, Methoden zu entwickeln, die es erlauben:
1. regional Rutschungsgebiete zu detektieren
2. kritische Gebiete kontinuierlich mit hoher Genauigkeit zu beobachten
3. Echtzeitinformationen für das Abschätzen von Risiken zu gewinnen.

1. Introduction

Worldwide landslides are one of the major types of natural hazards killing or injuring a large number of individuals and creating very high costs every year.

Besides direct costs landslides are also reason for even higher indirect costs like interruption of important infrastructure facilities or losses for the tourist industry etc. In future it is very likely that the damages caused by landslides will even increase as the hilly areas, where the majority of the landslides occur, are used by a growing number of tourists and intersected by increasingly powerful transnational networks. In addition many global climate change scenarios predict an increase in the probability for heavy rain, which is a primary trigger for landslides (refer to Taiwan Sept. 2001, 128 death toll from landslides). This implies that there is urgent need for highly productive and reliable tools for landslide hazard management at an operational level.

In contrast to these prospects the national funds for hazard prevention and the number of personnel in the national authorities responsible for hazard management are stagnating or shrinking. Several natural disasters in the last years have also shown that the decision makers in hazard management are confronted with a justice and public that will evaluate their decisions in an increasingly critical manner and look very thoroughly for possible responsibilities.

There has already been done a wide range of research work on landslides [1] [2] [3]. Most of this work had a bias towards one discipline, like remote sensing or geology. The proponents of the project believe that for future work the multidisciplinary integration of different methods has the greatest potential for substantial progress in natural hazards management. The aim of the suggested research project is to set up an integrated workflow for landslide hazard management. This system should lead the practitioner from the data acquisition all the way to suggestions of risk management measures. The majority of fatalities occurs due to the fact that the affected persons are not aware of the danger and are surprised then by the actual landslide event.

The emphasis of the suggested project is the development of observation methods that allow

1. to detect potential landslides
2. an efficient and continuous observation of critical areas
3. the derivation of real time information about actual risks

2. Existing technology

Currently the investigation of landslides and unstable slopes is based on two groups of information sources:

2.1. Information sources on regional scale:

- Data on historical landslides:
  The relevant data should include information when and where landslides occurred in the critical areas.
- General conditions of the areas
  Map data of the general conditions, like digital elevation model, geology, tectonics, geomorphology, vegetation, climate, land use are indicators for active landslides [4].
- Remote sensing data
  Further sources for the potential landslide risk can be differential satellite image analysis, time series of airborne sensor data (photogrammetry, laser scanners, radar systems, geophysical measuring devices) and terrestrial mapping.

2.2. Information sources on local scale :

- Geodetical:
  GPS, precise levelling, tacheometers (measurement robots), multisensoral remote sensing techniques (using optical- and radar-sensors)
- Geophysical:
  geoelectric and electromagnetic field measurements
- Geodynamical:
  borehole tiltmeters, extensometers, hydrostatic tiltmeters
- Seismological:
  sensors for microseismic activity, seismic reflexion measurements
- Hydrological:
  sensors for groundwater level variations, water level variations, groundwater stream variations
- Meteorological:
  sensors for temperature, air pressure, precipitation

The information of the first source enables to detect the site of the landslide, those of the second source are used to describe the mechanism of the processes. Normally, however, only some of the parameters (deformation vectors, water power pressure) are used for the investigations and the measurement points are widely spaced divided across the landslide domain. The deformation measurements are frequently based only

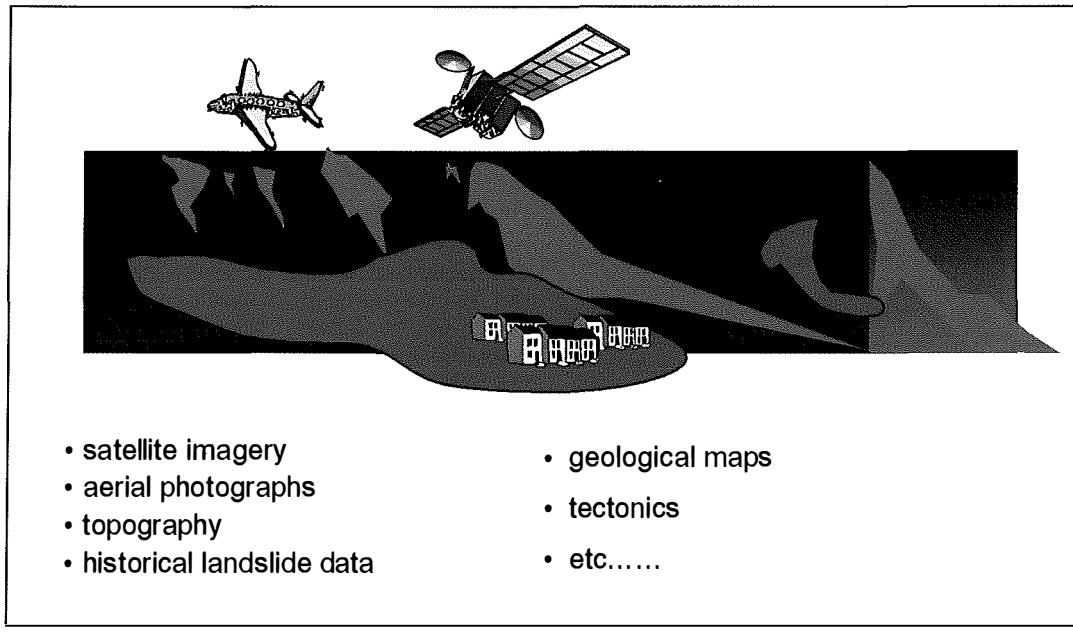


Fig 1: Large scale monitoring with airborne and satellite remote sensing

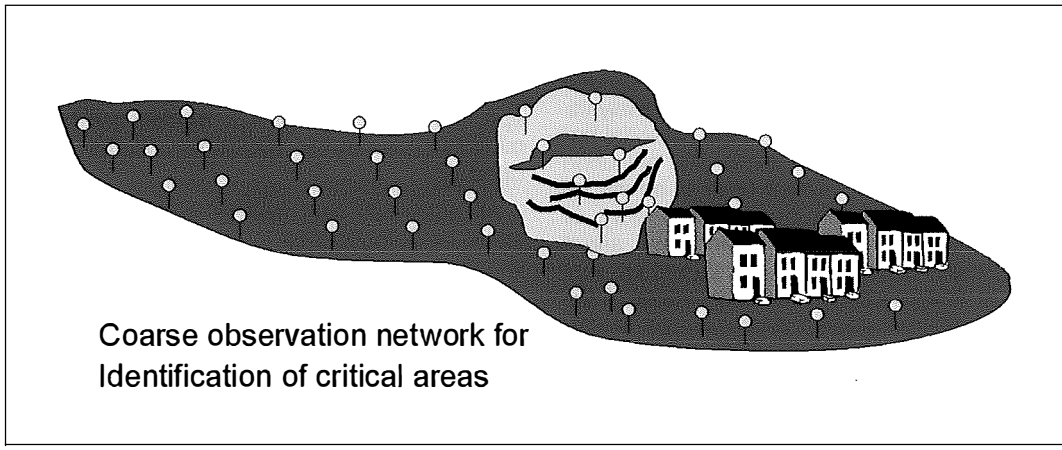


Fig 2: Large scale monitoring using GPS and total stations

on one measurement method (GPS, tacheometry). As these measurement methodes are relatively costly, normally only a limited number of observation points are used and the measurement systems are not run continuously. This methodology has made it possible to follow the evolution of the landslides precisely but yet it is difficult to predict the evolution.

By a better integration of this information sources a more reliable prediction should be possible in the future.

3. Integrated Optimization

An advanced model is now under investigation, based on large scale monitoring and a multi-component alert system. The method will consist of four different steps:

1st step:

Detection of potential landslides (Large scale monitoring)

To get information about the long-term geodynamical processes, a large scale evaluation has to be performed. This includes the historical data, all information describing the general conditions of the areas, as outlined in section 2.1, and remote sensing data, like aerial photographs, optical and radar images from satellites (Fig. 1).

More specific remote sensing techniques (InSAR), differential GPS (using phase observations) and tacheometric measurements shall be used to get additive information about the deformation process (Fig. 2). The measurements will be performed only three or four times a year and the results will be fields of vectors describing the displacements and velocities [10].

An advanced and generalised deformation analysis algorithm based as well on geometrical as on topographical, geological, hydrological and meteorological information has to be developed to improve the detection of taking-off domains, see [7] [11].

2nd step:

High precision permanent measurements in the taking off domains

High precision borehole tiltmeters, extensometers and hydrostatic levelling and further relative measurement systems shall be used in the area of the taking-off domains to get online information about the geodynamical process [6] [8].

This multi-sensor system will be running continuously and can therefore support a real time alert system, see Fig. 3 [5].

3rd step:

Impact and risk assessment / Development of stategies for alert systems

Risk assessment comprises the analysis of the empirical data and the development of an alert system deals with the management of an impending landslide hazard.

The analysis of empirical data includes the detection and interpretation of velocity fields in order to define zones of increased deformation. The final analysis will be supported by expert

systems, using methods of cluster analysis, neuronal networks, fuzzy techniques and others [7].

The process of risk analysis can be divided into hazard analysis and vulnerability analysis. Hazard analysis is the review of the potential hazardous processes. Scenarios for the evolution of a landslide area of interest have to be set up including an estimation of the probability of these scenarios. Different scenarios will affect different areas and therefore a different amount of people and property. The assessment of the impact of different hazard processes on the affected population and its property is called vulnerability analysis.

The integration of hazard and vulnerability analysis will lead to an estimation of the actual risk situation of the affected population. The risk management measures will depend heavily on the specific conditions and will include landuse planning, technical measures (e.g. drainage systems), biological measures (e.g. afforestations) and temporary measures (e.g. evacuations).

The information of the public about potential hazards and the fast and effective alerting and evacuation of the people in endangered areas are vital prerequisites for a functioning natural hazard management system. One of the main sources for fatalities caused by natural hazards is the lack of awareness of a certain hazard and the lack of knowledge about appropriate behavior in an emergency case. Therefore people are often surprised then by a landslide. This work package tries to tackle this deficit by the dissemination of background information to the public and by the development of new means of alerting and evacuation by the application of new communication technologies. This has to be performed by taking into account the cultural and socioeconomic boundary conditions in each specific case. Participative models for hazard mitigation/prevention measures will be developed [9].

An advanced alert system will consist of the following components:

- communication system which transforms the informations to a control station
- communication system which informs the persons who have to make decisions
- quality criteria to support the decisions
- communication system which informs the persons who have to start actions.

4. The project OASYS

The integrated workflow for landslide hazard management, as outlined in section 3, is depicted in Fig 4. The integration of the acquired data in a multidisciplinary approach in order to detect and classify landslides will be one of the most challenging tasks of the project. Therefore a multidisciplinary working group was formed to meet this task:

- Vienna Consulting Engineers (VCE), Austria
- Institut for Geodesy and Photogrammetry, Technical University Braunschweig, Germany
- Institut for Geodesy and Geophysics, Technical University Vienna, Austria
- Geodetical and Geophysical Research Institute of the Hungarian Academy of Sciences, Hungary

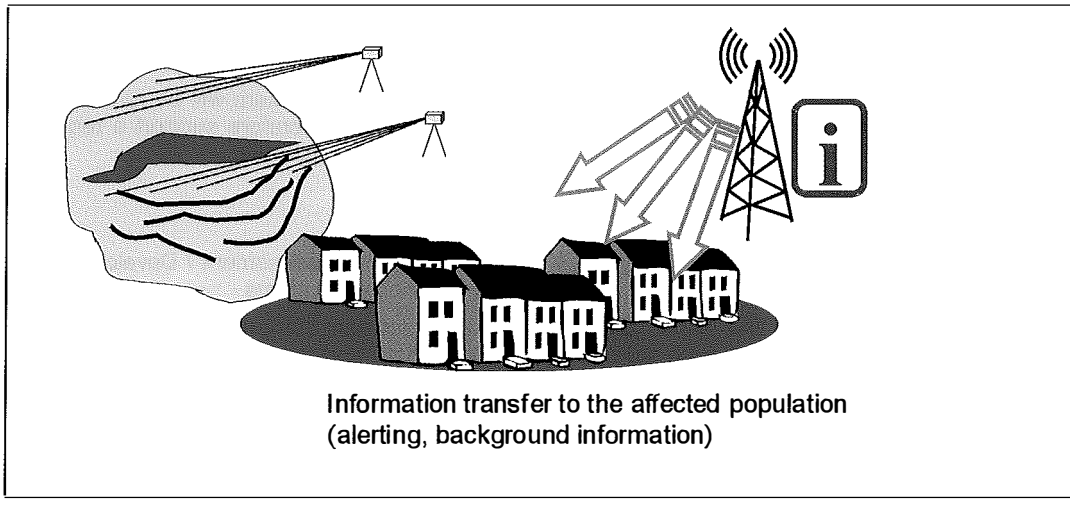


Fig. 3: Continuous local scale measurements and the information transfer

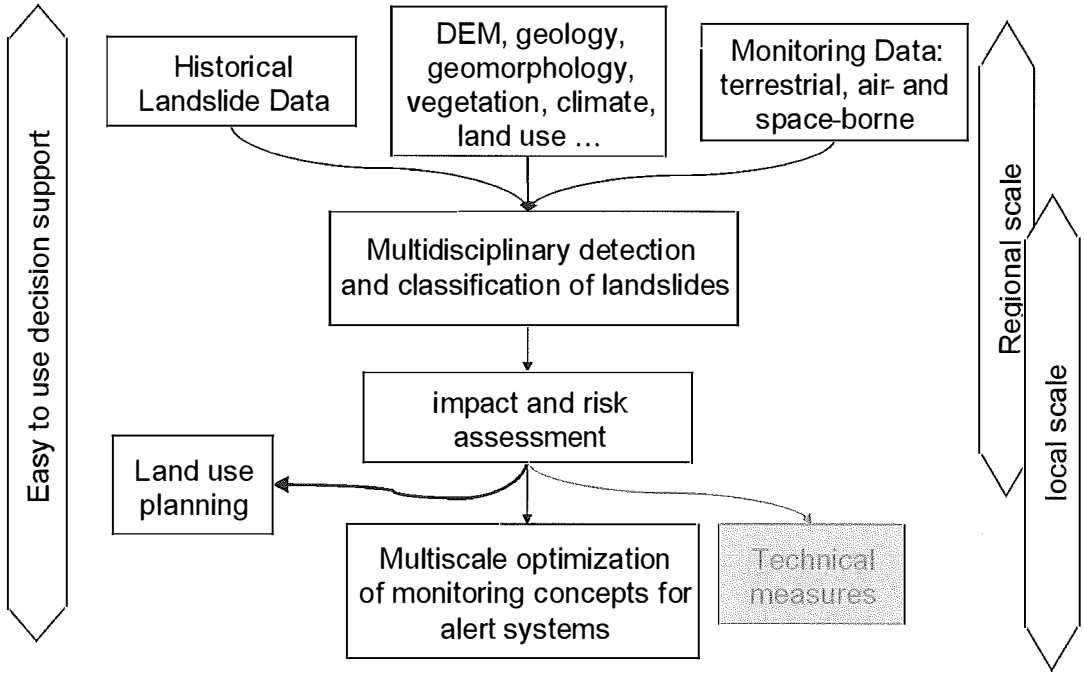


Fig. 4: Integrated work flow for integrated landslide hazard management

- Bureau for Applied Scientific Remote Sensing, Germany
- Geodata, Engineering Company, Austria
- Bureau of Investigations and Survey, China
- Hungarian Ministry of Interior, National Directorate General for Desaster Management, Hungary
- Egnatia Odos S.A., ConstructionCompany, Greece
- University of Modena, Italy
- School of Geodesy and Geomatics, Wuhan University, China
- Laboratory of Geodesy, Uni versity of Thessaloniki, Greece

The project OASYS is supported by the European Commission under the 5th Framework in the area "Energy, Environment and Sustainable Development Programme" under contract EVGI-2001-00061. The project started January 1, 2003.

References

[1] *ALMEIDA-TEIXEIRA M.E., et al.(Eds.), 1991:* Natural hazards and engineering geology. Prevention and control of landslides and other mass movements. Proceedings of the European School of Climatology and Natural Hazards course held in Lisbon from 28 March to 5 April 1990

[2] *BHANDARI R. K., 1984:* Simple and economical intrumentation and warning systems for landslides and other mass movements. State-of-the-art lecture, Proc. 4th Int. Symp. on Landslides, Toronto, vol. 1, 251–273.

[3] *CARRARA A., 1984:* Landslide hazard mapping: aims and methods, Mouvements de terrain. Association Francaise de Geographie Physique. Colloque de Caen, p. 141–151.

[4] *CUI Zhengquan, 1999:* Geology and Geomechanics of the Three Gorges Projects. Bureau of Investigations & Survey, Changjiang Water Resources Commission, MWR, May 30 1999

[5] *EINSTEIN H. H., 1988:* Landslide risk assessment procedure. Special lecture in Proceedings of the fifth international symposium on landslides, Lausanne, p. 1075–1090.

[6] „Geotechnical Instrumentation for Monitoring Field Measurements", Conference at Universityof Florida, March 2001

[7] *HABERLER M., 2003* A fuzzy system for the assessment of landslide monitoring data, This Volume

[8] *SAVVAIDIS P., LAKAKIS K., ZEIMPEKIS A., 2001:* Monitoring Ground Displacements at a National Highway Project: The Case of "Egnatia Odos" in Greece. Proceedings of IAG Workshop "Monitoring of Constructions and Local Geodynamic Process", Wuhan University, China, 2001

[9] *SPIKER, E. C., GORI P. L., 2000:* National Landslide Hazards Mitigation Strategy. Open-file Report 00-450, USGS.

[10] *THEILEN-WILLIGE,B., 1998:* Seismic Hazard Localization based on Lineament Analysis of ERS- and SIR-C- Radar-Data of the Lake Constance Area and on Field Check.- European Conference on Synthetic Aperture Radar, 25–27 May 1998 in Friedrichshafen,Proceedings, 447-550, VDE-Verlag, Berlin

[11] *ZHANG Zhenglu, HUANG Quanyi, CHMELINA Klaus, 2001:* Research on Geological Landslide Problems related to the Three Gorges Project. Proceedings of IAG Workshop on Monitoring of Constructions and Local Geodynamic Processes, 22.–24.5.2001, Wuhan, China

Contact

Univ.-Prof. Dr.-Ing. Heribert Kahmen: University of Technology Vienna, Institute of Geodesy and Geophysics, Gußhausstr. 27–29/E1283, A-1040 Vienna, email: Heribert.Kahmen@tuwien.ac.at

Prof. Dr.-Ing. Wolfgang Niemeier: University of Technology Braunschweig, Institut of Geodesy and Photogrammetry, Gaußstr. 22, D-38106 Braunschweig, email: W.Niemeier@tu-bs.de

Impressum

VGI Österreichische Zeitschrift für Vermessung & Geoinformation

91. Jahrgang 2003 / ISSN 0029-9650

Herausgeber und Medieninhaber: Österreichische Gesellschaft für Vermessung und Geoinformation (OVG), Austrian Society for Surveying and Geoinformation, Schiffamtsgasse 1-3, A-1025 Wien zur Gänze. Bankverbindung: Österreichische Postsparkasse BLZ 60000, Kontonummer PSK 1190933.

Präsident der Gesellschaft: Dipl.-Ing Gert Steinkellner, Schiffamtsgasse 1–3, A-1025 Wien, Tel. (01) 21176-4604, Fax (01) 2167550.

Sekretariat der Gesellschaft: Dipl.-Ing. Karl Haussteiner, Schiffamtsgasse 1–3, A-1025 Wien, Tel. (01) 21176-2311, Fax (01) 2167551.

Schriftleitung: Dipl.-Ing. Wolfgang Gold, Krotenthallergasse 3, A-1080 Wien, Tel. (01) 40146-212, Fax (01) 40146-333, Dipl.-Ing. Stefan Klotz, Schiffamtsgasse 1-3, A-1025 Wien, Tel. (01) 21176-3609, Fax (01) 2167551, Dipl.-Ing. Ernst Zahn, Schiffamtsgasse 1–3, A-1025 Wien, Tel. (01) 21176-3209, Fax (01) 2167551.

Manuskripte: Bitte direkt an die Schriftleitung senden. Es wird dringend ersucht, alle Beiträge in digitaler Form auf Diskette zu übersenden. Genaue Angaben über die Form der Abfassung des Textteiles sowie der Abbildungen (Autoren-Richtlinien) können bei der Schriftleitung angefordert werden. Beiträge können in Deutsch oder Englisch abgefaßt sein; Hauptartikel bitte mit einer deutschsprachigen Zusammenfassung und einem englischen Abstract einsenden. Namentlich gezeichnete Beiträge geben die Meinung des Autors wieder, die sich nicht mit der des Herausgebers decken muß. Die Verantwortung für den Inhalt des einzelnen Artikels liegt daher beim Autor. Mit der Annahme des Manuskriptes sowie der Veröffentlichung geht das alleinige Recht der Vervielfältigung und Wiedergabe auf den Herausgeber über.

Anzeigenbearbeitung und -beratung: Dipl.-Ing. Wolfgang Gold, Krotenthallergasse 3, A-1080 Wien, Tel. (01) 40146-212. Unterlagen über Preise und technische Details werden auf Anfrage gerne zugesendet.

Erscheinungsweise: Vierteljährlich in zwangloser Reihenfolge (1 Jahrgang = 4 Hefte). Auflage: 1500 Stück.

Abonnement: Nur jahrgangsweise möglich. Ein Abonnement gilt automatisch um ein Jahr verlängert, sofern nicht bis zum 1.12. des laufenden Jahres eine Kündigung erfolgt. Die Bearbeitung von Abonnementangelegenheiten erfolgt durch das Sekretariat. Adreßänderungen sind an das Sekretariat zu richten.

Verkaufspreise: Einzelheft: Inland 15 €, Ausland 18 €; Abonnement: Inland 50 €, Ausland 60 €; alle Preise exclusive Mehrwertsteuer.

Satz und Druck: Druckerei Berger, A-3580 Horn, Wiener Straße 80.

Offenlegung gem. § 25 Mediengesetz

Medieninhaber: Österreichische Gesellschaft für Vermessung und Geoinformation (OVG), Austrian Society for Surveying and Geoinformation, Schiffamtsgasse 1–3, A-1025 Wien zur Gänze.

Aufgabe der Gesellschaft: gem. § 1 Abs. 1 der Statuten (gen. mit Bescheid der Sicherheitsdirektion Wien vom 17.9.1996, Zl. IV-SD 1394/VVM/96): a) die Vertretung der fachlichen Belange der Vermessung und Geoinformation auf allen Gebieten der wissenschaftlichen Forschung und der praktischen Anwendung, b) die Vertretung aller Angehörigen des Berufsstandes, c) die Förderung der Zusammenarbeit zwischen den Kollegen der Wissenschaft, des öffentlichen Dienstes, der freien Berufe und der Wirtschaft, d) die Förderung des wissenschaftlichen Nachwuchses, e) die Herausgabe einer Zeitschrift mit dem Namen „Österreichische Zeitschrift für Vermessung und Geoinformation" (VGI).

Erklärung über die grundlegende Richtung der Zeitschrift: Wahrnehmung und Vertretung der fachlichen Belange aller Bereiche der Vermessung und Geoinformation, der Photogrammetrie und Fernerkundung, sowie Information und Weiterbildung der Mitglieder der Gesellschaft hinsichtlich dieser Fachgebiete.


Die Adresse ...
www.bev.gv.at
... mit dem neuen Service!
See you:
www.bev.gv.at
BEV
Bundesamt für Eich- und Vermessungswes